HILBERT SPACE LYAPUNOV EXPONENT STABILITY

GARY FROYLAND, CECILIA GONZÁLEZ-TOKMAN,
AND ANTHONY QUAS

ABSTRACT. We study cocycles of compact operators acting on a separable Hilbert space, and investigate the stability of the Lyapunov exponents and Oseledets spaces when the operators are subjected to additive Gaussian noise. We show that as the noise is shrunk to 0, the Lyapunov exponents of the perturbed cocycle converge to those of the unperturbed cocycle; and the Oseledets spaces converge in probability to those of the unperturbed cocycle. This is, to our knowledge, the first result of this type with cocycles taking values in operators on infinite-dimensional spaces. The infinite dimensionality gives rise to a number of substantial difficulties that are not present in the finite-dimensional case.

1. INTRODUCTION

A question of paramount importance in applied mathematics is: How to tell if the conclusions derived from a model indeed capture relevant features of an underlying system? Stability results address this question by giving conditions under which small changes in a model produce only small changes in the outcomes of the analysis.

In the last decade, multiplicative ergodic theory has been developed in the so-called semi-invertible setting (that is the setting in which the underlying base dynamics are assumed to be invertible, but no invertibility assumptions are made on the matrices) [11, 12, 14, 15]. This enables a fine analysis of time-dependent or driven dyanamics, where the driving is invertible, but the phase space dynamics need not be invertible. Driven linear dynamics, modelled by cocycles of (possibly non-invertible matrices or non-injective linear operators) fits into this framework, but one may also consider a variety of linearisations of driven nonlinear dynamics. An important linearisation is the replacement of a nonlinear dynamical system on a finite-dimensional phase space with its transfer operator describing the linear action of the dynamics on real- or complex-valued functions of the phase space. This

Date: September 6, 2017.

linearisation provides a useful mathematical tool to numerically analyse phase space transport in time-dependent nonlinear systems, such as models of (driven) geophysical flows [13].

The Lyapunov spectrum of the linear (or linearised) cocycle quantifies the magnitudes and timescales of growth and decay in the driven dynamics, and the corresponding so-called Oseledets spaces determine the modes in which this growth or decay occurs. From a modelling perspective, it is important to know that Lyapunov spectral analyses of models are robust to model errors and to numerical errors. That is, do the Lyapunov exponents and Oseledets spaces – obtained using either mathematical models or observational data, both of which contain errors – correspond to *real* features of the driven system? Further, are the Lyapunov exponents and Oseledets spaces insensitive to the inevitable numerical approximation errors in the numerical schemes required to extract these ergodic-theoretic objects from models or observational data?

The aim of this work is to provide an initial step in establishing conditions for the stability of Lyapunov exponents and Oseledets spaces, the essential components underlying multiplicative ergodic theory, in an *infinite-dimensional* context. The infinite dimensionality aspect is crucial for stability results for driven linear dynamics on infinite phase spaces and to eventually encompass the setting of transfer operators and other infinite-dimensional linearisations. In the infinite-dimensional context, and aside from works focusing exclusively on the i.i.d. perturbation (noise) setting, stability results have only been established either (i) under uniform hyperbolicity assumptions on the underlying cocycle, which for example cover the case of random perturbations of a fixed map [1, 6]; or (ii) for the top (first) component of the splitting, in the context of transfer operators [9], where the leading Lyapunov exponent is always 0, corresponding to a random fixed point.

Early results concerning stability of Lyapunov exponents for finite-dimensional (matrix) cocycles include [23, 18, 19, 16]. In the setting of invertible matrix cocycles, the closest results to this work are due to Ledrappier, Young and Ochs [24, 21, 22]. The difficulty of the stability problem at hand, even in the finite-dimensional setting, is highlighted by the existence of negative stability results for Lyapunov exponents of matrix cocycles [3, 4], which show that for non-uniformly hyperbolic cocycles, carefully chosen arbitrarily small perturbations may collapse the entire spectrum of Lyapunov exponents to a single exponent. In this finite-dimensional setting, the stability problem remains an active topic of research, and related recent results include [5, 2]. In the setting of semi-invertible matrix cocycles, the authors established stability results under stochastic perturbations in [10, 8].

In this paper, we study cocycles taking values in compact operators on a separable Hilbert space. The unperturbed cocycle is assumed to be strongly coercive, with exponentially-decaying transmission between higher order modes, so that the leading Oseledets spaces tend to be concentrated on low order modes. This issue of the cocycle sending an arbitrarily high-order mode to a low-order mode does not arise in the finite-dimensional setting. Additionally, unlike the finite-dimensional case, there is no natural Lebesgue-like measure on the infinite-dimensional space of perturbations. Hence as our model of noise we use additive Gaussian perturbations. The Gaussian nature of the perturbations allows for unbounded changes, and is also convenient for calculations. In order to maintain the noise as a *small* perturbation, the Gaussian perturbations are required to have stronger exponential decay than the unperturbed cocycle. We regard the model as a natural generalisation of the finite-dimensional Ledrappier-Young setting to infinite dimensions.

The main results of the paper, Theorems A and B, yield, respectively, convergence of Lyapunov exponents and Oseledets spaces of the randomly perturbed cocycles. The method of proof of stability of Lyapunov exponents builds on the work of Ledrappier and Young [21], which dealt with Lyapunov exponents in invertible matrix cocycles, as well as on our recent work [10], which had to handle the complications arising from non-invertibility of the matrices. Section 2.2 presents examples for which Theorems A and B apply.

2. The model and principal results

Throughout the paper $\sigma\colon \Omega \to \Omega$ is an invertible measurable transformation, $\mathbb{P}$ is an ergodic invariant probability measure, and H is a separable Hilbert space with basis $e_1, e_2, \ldots$.

The Hilbert-Schmidt norm is $\|A\|_{\mathsf{HS}}^2 = \sum_{i,j} \langle Ae_i, e_j\rangle^2$. Define a stronger norm: $\|A\|_{\mathsf{SHS}}^2 = \sum_{i,j} 2^{2(i+j)} \langle Ae_i, e_j\rangle^2$. We frequently think of operators with bounded HS norm as infinite matrices where the entries are square-summable. We write HS for the collection of Hilbert-Schmidt operators on H (those operators, A, satisfying $\|A\|_{\mathsf{HS}} < \infty$), and SHS for the collection of strong Hilbert-Schmidt operators (those operators satisfying $\|A\|_{\mathsf{SHS}} < \infty$).

We write $A_\omega^{(n)}$ for the unperturbed cocycle: $A_\omega^{(n)} = A_{\sigma^{n-1}\omega} \cdots A_\omega$, and call $A\colon \Omega \to \mathsf{SHS}$ the *generator* of the operator cocycle. Throughout the article, Δ will denote the random Hilbert-Schmidt operator with independent normal entries with mean 0 and where the (i,j) entry has standard deviation $3^{-(i+j)}$. Write γ for the measure on SHS corresponding to this distribution. We apply a sequence of independent perturbations $\mathbf{\Delta} = (\Delta_n)_{n\in\mathbb{Z}}$, where each Δ_n has the distribution above. For ω lying in the base space, we denote by $\bar\omega$ the pair $(\omega, \mathbf{\Delta})$ specifying the point of the base space and the sequence of perturbations. The space of such pairs is denoted by $\bar\Omega$, and is equipped with the transformation $\bar\sigma = \sigma \times s$, where s is the left shift on the sequence of perturbations and the ergodic invariant measure $\bar{\mathbb{P}} = \mathbb{P} \times \gamma^{\mathbb{Z}}$. The perturbed cocycle is parameterized by ϵ (a measure of the size of the perturbation) and defined by

$$A^{\epsilon(n)}_{\bar\omega} = (A_{\sigma^{n-1}\omega} + \epsilon\Delta_{n-1}) \cdots (A_\omega + \epsilon\Delta_0).$$

Theorem A. *Let $\sigma\colon \Omega \to \Omega$ be an invertible measurable transformation and let $\mathbb{P}$ be an ergodic invariant probability measure for σ. Let H be a separable Hilbert space and let $A\colon \Omega \to \mathsf{SHS}$ be the generator of an operator cocycle satisfying $\int \log \|A_\omega\|_{\mathsf{SHS}}\, d\mathbb{P}(\omega) < \infty$.*

Let $\bar\Omega$, $\bar\sigma$ and $\bar{\mathbb{P}}$ be as defined above. For each parameter $\epsilon > 0$, define a new cocycle $A^\epsilon\colon \bar\Omega \to \mathsf{SHS}$ over $\bar\sigma$ with generator $A^\epsilon(\bar\omega) = A(\omega)+\epsilon\Delta_0$. Then the Lyapunov exponents of A^ϵ (listed with multiplicity) converge to those of A as $\epsilon \to 0$.

Theorem B. *Assume the hypotheses and notation of Theorem A. Let the (at most countably many) distinct Lyapunov exponents of the cocycle A be $\lambda_1 > \lambda_2 > \ldots > -\infty$, with corresponding multiplicities $d_1, d_2, \ldots$. Let the corresponding Oseledets decomposition be $\mathsf{SHS} = Y_1(\omega) \oplus Y_2(\omega) \oplus \ldots$. Let $D_0 = 0$, $D_i = d_1 + \ldots + d_i$ and let the Lyapunov exponents (with multiplicity) be $\infty > \mu_1 \geq \mu_2 \geq \ldots > -\infty$, so that $\mu_j = \lambda_i$ if $D_{i-1} < j \leq D_i$.*

Let $\mathcal{U}_i = (\lambda_i - \alpha, \lambda_i + \alpha)$ be a neighbourhood of λ_i not containing any other exponent of the unperturbed cocycle. Let ϵ_0 be such that for each $\epsilon \leq \epsilon_0$ and each $D_{i-1} < j \leq D_i$, the j^{th} Lyapunov exponent μ_j^ϵ of the perturbed cocycle satisfies $\mu_j^\epsilon \in \mathcal{U}_i$. For $\epsilon < \epsilon_0$, let $Y_i^\epsilon(\bar\omega)$ denote the sum of the Oseledets subspaces of A^ϵ having exponents in $\mathcal{U}_i$. Then $Y_i^\epsilon(\bar\omega)$ converges in probability to $Y_i(\omega)$ as $\epsilon \to 0$.

For $\lambda > 1$, we let $\mathcal{D}_\lambda$ be the diagonal matrix whose (i,i) entry is λ^{-i}. Formally we can write the random operator Δ from Theorem A as $\mathcal{D}_3 N \mathcal{D}_3$, where N is a countably infinite square matrix of independent standard normal random variables.

Throughout the remainder of the paper there will be numerous constants. We will mostly just use the symbol C to indicate a constant, where C may refer to different constants at different places, even in the same proof. That is, whenever we write C, we refer to a quantity that may depend on k (the number of exponents that we aim to control), and on the underlying dynamical system, but not on ϵ, the size of the perturbations. The exception to this will be some of the principal propositions where estimates are collected for assembly in Section 9. In these propositions, constants will be numbered according to the proposition in which they are found, so that C_{34}, for example, is defined in Lemma 34.

2.1. **Discussion of proof strategy.** The strategies in all three papers, [21, 10] and this one, are similar in spirit: The idea is to split long sequences of matrices observed along the cocycle into *good* and *bad* blocks, depending on whether or not the long term behaviour of the cocycle corresponds to the observed behaviour within the block, and then handle carefully the concatenations. However, at the technical level, there are substantial complications in this new infinite-dimensional setting, arising from the need to handle *wild* perturbations occurring in possibly higher and higher modes.

As in the previous works [22, 10], the stability of Oseledets spaces is deduced from the stability of the Lyapunov exponents, but the strategy of the proof here is different. The approach of Ochs [22] applies only to invertible matrices, and the proof is essentially finite-dimensional. The core of the argument is: if the perturbed slowest Oseledets spaces were often far from its unperturbed counterpart, the contribution to the bottom exponent of the perturbed system on this part of the base space would be at least λ_{d-1}. Hence, convergence of the exponents implies the perturbed and unperturbed slow spaces are mostly nearby. This is basically an expectation argument. Subsequent Oseledets spaces are similarly controlled using exterior powers. The approach of [10] in the context of not necessarily invertible matrices relies on the use of Möbius transformations or graph transforms. The essence of the argument is one fixes all of the perturbations to the matrices other than the perturbation at time -1. Since there is exponential contraction in a cone around the unperturbed *fast space* (that is the span of the k-dimensional Oseledets spaces with largest Lyapunov exponents), all but a very small set of perturbations at time -1 cause the fast space to fall into the basin of attraction, and to end up near the unperturbed fast space. While this approach would still apply in the infinite-dimensional case, the new argument of this paper has the advantages that it is simpler and more general; in particular, it does not rely on any special structure for the perturbations, such as absolute continuity, which played a role in [10]. All that is required is that the perturbations are small with high probability. The approach in the current paper goes as follows: If the perturbed k-dimensional fast space is not close to the unperturbed fast space at time N (where N is the block size), then the minimum angle between the perturbed fast space at time 0 and the unperturbed slow space at time 0 has to be exponentially small. Whenever this happens, there is a growth drop of the k-dimensional volume of order $\exp(-(\lambda_k - \lambda_{k+1})N)$ over this block. An expectation argument ensures that this must happen rarely because otherwise the perturbed λ_k would be much less than the unperturbed λ_k.

Finally, we briefly describe in more detail the structure of the proof of Theorem A since there is considerable preparation before we start the proof. The bulk of the proof is concerned with giving a lower bound for the sum of the k leading perturbed exponents, that is the maximal logarithmic growth rate of k-volumes. Given ϵ, one defines a block length, $N \sim |\log \epsilon|$. For a large n, we estimate the top exponents of the product $A^{\epsilon}{}_{\bar\omega}^{(nN)}$, a perturbed block of length nN. First, we replace the (sub-additive) logarithmic k-volume growth, $\Xi_k(\cdot)$ by a related approximately super-additive quantity, $\tilde{\Xi}_k(\cdot)$ (Sections 7 and 8). We use this super-additivity to split $A^{\epsilon}{}_{\bar\omega}^{(nN)}$ into good super-blocks (of length a multiple of N) and bad blocks (of length $N-2$), that is $\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(nN)}) \gtrsim \tilde{\Xi}_k(A^{\epsilon}{}_{\bar\omega}^{(nN)}) \gtrsim \sum \tilde{\Xi}_k(\text{blocks})$. In section 4, ingredients for the estimate $\Xi_k(G^\epsilon) \gtrsim \Xi_k(G)$ are established, where G represents a good super-block and G^ϵ its perturbed version. In sections 5 and 6, ingredients for $\tilde{\Xi}_k(B^\epsilon) \gtrsim \tilde{\Xi}_k(B)$ are established (where B is a bad block and B^ϵ is its perturbed version). The estimates $\tilde{\Xi}_k(B) \gtrsim \Xi_k(B)$ and $\tilde{\Xi}_k(G^\epsilon) \gtrsim \Xi_k(G^\epsilon)$ are based on ingredients in Section 8. Re-assembling the pieces using sub-additivity of Ξ_k and accounting for the errors gives the result.

2.2. **Examples.** Here we present a simple class of cocycles for which Theorems A and B apply. Let $\sigma\colon \Omega \to \Omega$ be an invertible measurable transformation of $(\Omega, \mathbb{P})$, $\mathbb{P}$ an ergodic invariant probability measure for σ and $H = \ell^2$. Consider a sequence of log-integrable functions $a_i : \Omega \to \mathbb{R}$, each inducing a one-dimensional cocycle over σ, with generator $\omega \mapsto a_i(\omega)$ and associated Lyapunov $\nu_i = \int \log |a_i| \, d\mathbb{P}$. For each $\omega \in \Omega$, consider the (infinite) diagonal matrix $A(\omega)$, with diagonal entries $a_i(\omega)$. Notice that $\|A(\omega)\|^2_{\mathsf{SHS}} = \sum_{i=1}^{\infty} 2^{4i} a_i(\omega)^2$, which may be translated into explicit necessary and sufficient conditions for the hypotheses (i) $A : \Omega \to \mathsf{SHS}$ and (ii) $\int \log \|A\|_{\mathsf{SHS}} \, d\mathbb{P} < \infty$ to hold, and thus for Theorems A and B to apply.

Notice that, in a similar manner, one can construct examples of block diagonal and triangular cocycles with explicit conditions prescribed to ensure $A(\omega) \in \mathsf{SHS}$ and $\int \log \|A\|_{\mathsf{SHS}}\, d\mathbb{P} < \infty$, so that the theorems apply.

3. Notation and the quantity $\tilde{\Xi}_k$

Recall that the Grassmannian of a Banach space is the space of closed complemented subspaces. In a Hilbert space, every closed subspace is complemented (by its orthogonal complement). We define $\mathcal{G}_k(H)$ to be the space of (necessarily closed) k-dimensional subspaces of H and $\mathcal{G}^k(H)$ to be the space of closed k-codimensional subspaces of H. The collection of all closed subspaces of H will be written $\mathcal{G}(H)$. We will reserve the symbol S for the unit sphere of H throughout the article.

We define a metric on $\mathcal{G}(H)$ by

$$\angle(U,V) = \max\left(\max_{u\in U\cap S}\min_{v\in V\cap S}\|u-v\|, \max_{v\in V\cap S}\min_{u\in U\cap S}\|u-v\|\right),$$

that is the Hausdorff distance between the intersections of the two subspaces with the unit sphere. We remark that this differs by at most a bounded factor from another metric, the 'gap' between closed subspaces defined in Kato [17]. This is a complete metric on $\mathcal{G}(H)$.

We also make use of a measure of transversality between two subspaces of complementary dimensions: if $U \in \mathcal{G}^k(H)$ and $V \in \mathcal{G}_k(H)$, then

$$\perp(U,V) = \frac{1}{\sqrt{2}} \min_{u\in U\cap S, v\in V\cap S} \|u-v\|.$$

The normalization is chosen so that if U and V have a common vector, then $\perp(U,V) = 0$, while if they are orthogonal complements, then $\perp(U,V) = 1$. We have the reverse triangle inequality: if $U', U \in \mathcal{G}^k(H)$, $V, V' \in \mathcal{G}_k(H)$, then $\perp(U',V') \geq \perp(U,V) - \angle(V,V') - \angle(U,U')$.

We already introduced the classes of linear operators HS and SHS on H with their associated norms, so that we have $\mathsf{SHS} \subset \mathsf{HS} \subset K(H)$, where $K(H)$ stands for the compact linear operators on H. We write $\|\cdot\|_{\mathsf{op}}$ for the operator norm, so that $\|\cdot\|_{\mathsf{SHS}} \geq \|\cdot\|_{\mathsf{HS}}$ for elements of SHS and $\|\cdot\|_{\mathsf{HS}} \geq \|\cdot\|_{\mathsf{op}}$ for elements of HS.

For compact operators on H, the notions of singular vectors and singular values pass directly from the finite-dimensional case. If $A \in K(H)$, we write $s_1(A) \geq s_2(A) \geq \dots$ for the singular values (with multiplicity in decreasing order). The maximal logarithmic rate of k-dimensional volume growth is given by $\Xi_k(A) := \log(s_1(A)\cdots s_k(A))$.

Define

$$\tilde{\Xi}_k(A) = \mathbb{E}\,\Xi_k(\Pi_k \Delta A \Delta' \Pi_k),$$

where Π_k denotes orthogonal projection onto the subspace of H spanned by $e_1, \ldots, e_k$ and Δ and Δ' are independent copies of the random Hilbert-Schmidt operator. The key reason for the introduction of $\tilde{\Xi}_k$ is that it satisfies an approximate *super-additivity* property (see Proposition 24) that complements the sub-additivity of Ξ_k.

We denote by $\bar{\Omega}$, the space $\Omega \times \mathsf{SHS}^{\mathbb{Z}}$ and act on $\bar{\Omega}$ with the transformation $\sigma \times s$, where s is the left-shift map on $\mathsf{SHS}^{\mathbb{Z}}$. The space $\bar{\Omega}$ is equipped with the measure $\mathbb{P} \times \gamma^{\mathbb{Z}}$, where γ is the multi-variate normal distribution on SHS described above in which distinct elements of Δ are independent and the (i, j) element is normal with mean 0 and variance $3^{-2(i+j)}$. We write $\bar{\omega}$ for a typical element of $\bar{\Omega}$, that is a pair $(\omega, \mathbf{\Delta})$, where $\mathbf{\Delta} = (\Delta_n)_{n\in\mathbb{Z}}$.

Informally, we expect an inequality like $\tilde{\Xi}_k(A) \geq \Xi_k(A) - \mathcal{E}^{\mathrm{L}}{}_k(A) - \mathcal{E}^{\mathrm{R}}{}_k(A)$. By $\mathcal{E}^{\mathrm{L}}{}_k(A)$ (which stands for 'left energy'), we mean a measure of the modes on which the top k left singular vectors are distributed, while $\mathcal{E}^{\mathrm{R}}{}_k(A)$ measures the modes where the right singular vectors are supported. For example, if the top left singular vectors are e_7, e_8, e_{11} and e_{13}, we expect $\mathcal{E}^{\mathrm{L}}{}_4(A)$ to be approximately $39 \log 3$.

Lemma 1. *Let V be a k-dimensional subspace of H. Let D be a bounded operator on H. There exists an orthonormal basis $v_1, \ldots, v_k$ for V with the property that $Dv_1, \ldots, Dv_k$ are mutually orthogonal.*

This follows from the singular value decomposition of finite-dimensional operators.

Lemma 2. *Let U and V be k-dimensional subspaces of H. Then the two quantities appearing in the definition of $\angle(U, V)$ are equal:*

$$\max_{u\in U\cap S} \min_{v\in V\cap S} \|u - v\| = \max_{v\in V\cap S} \min_{u\in U\cap S} \|u - v\|.$$

Proof. Let Π_U be the orthogonal projection onto U and Π_V be the orthogonal projection onto V. Then the singular vectors of $\Pi_V \circ \Pi_U$ give an orthogonal basis of U, $u_1, \ldots, u_n$ with images $s_1v_1, \ldots, s_nv_n$, where $v_1, \ldots, v_n$ form an orthogonal basis of V (if $\Pi_V\Pi_U u_i = 0$, then v_i can be chosen to be an arbitrary unit vector of V satisfying the orthogonality condition). Write $u_i = s_iv_i + w_i$ with $w_i \in V^\perp$. One can then check that $\langle u_i, v_j\rangle = 0$ if $i \neq j$. Notice that u_i and s_iv_i are either equal or non-collinear. It follows from the above that $U + V$ may be expressed as the orthogonal direct sum $\mathrm{lin}\{u_1, v_1\} \oplus \ldots \oplus \mathrm{lin}\{u_n, v_n\}$. One can now check that the linear map R from $U+V$ to itself mapping u_i to v_i and vice versa is an isometry interchanging U and V. Applying this map yields the desired equality. $\square$

Let V be a k-dimensional subspace of H, and Π be the orthogonal projection onto V. We define the *energy* of Π (also the 'energy of V') to be

$$\mathcal{E}_k(\Pi) = -\Xi_k(\Pi \circ \mathcal{D}_3) = -\sum_{i=1}^{k} \log \|\mathcal{D}_3 v_i\|,$$

where the (v_i) are as guaranteed by the Lemma 1 with the operator D taken to be $\mathcal{D}_3$.

Lemma 3. *For any $k \in \mathbb{N}$, there exists a $C > 0$ such that if Π and Π' are orthogonal projections onto k-dimensional subspaces and $Q \subset \mathsf{HS}$ satisfies $\gamma(Q) \geq \frac{1}{2}$, then*

$$\left| \mathbb{E}\left(\Xi_k(\Pi\Delta\Pi') \middle| \Delta \in Q\right) + (\mathcal{E}_k(\Pi) + \mathcal{E}_k(\Pi')) \right| \leq C.$$

Proof. Let $u_1, \ldots, u_k$ be the basis guaranteed by Lemma 1 (applied with $D = \mathcal{D}_3$) for the range of Π and $v_1, \ldots, v_k$ be the corresponding basis for Π'.

Now $\Xi_k(\Pi\Delta\Pi') = \log\det|M|$, where M is a random matrix whose (i,j) entry is $\langle u_i, \Delta v_j\rangle$. The entries of M therefore have a multi-variate normal distribution. Each has mean 0, so the unconditioned distribution of M is determined by the covariance of the pairs of entries of the matrix.

Using the fact that the coordinates of the u's and v's are bounded and the entries of Δ decay exponentially, we calculate

$$\begin{aligned}
\mathrm{Cov}(M_{ij}, M_{i'j'}) &= \mathbb{E} \sum_{l,m,l',m'} (u_i)_l \Delta_{lm} (v_j)_m (u_{i'})_{l'} \Delta_{l'm'} (v_{j'})_{m'} \\
&= \sum_{l,m} 3^{-2(l+m)} (u_i)_l (u_{i'})_l (v_j)_m (v_{j'})_m \\
&= \langle \mathcal{D}_3 u_i, \mathcal{D}_3 u_{i'}\rangle \langle \mathcal{D}_3 v_j, \mathcal{D}_3 v_{j'}\rangle,
\end{aligned}$$

where for the second line, we used the fact that distinct entries of Δ are independent, and so have 0 covariance. We see then, by the choice of u's and v's, that distinct entries of M have 0 covariance, and so are independent. The variance of the (i,j) entry of the matrix is $\|\mathcal{D}_3 u_i\|^2 \|\mathcal{D}_3 v_j\|^2$, hence the unconditioned distribution of the (i,j) entry of the matrix is $\|\mathcal{D}_3 u_i\| \|\mathcal{D}_3 v_j\|$ times a standard normal.

Notice that the entire i row has a multiplicative factor of $\|\mathcal{D}_3 u_i\|$ and the entire j column has a multiplicative factor of $\|\mathcal{D}_3 v_j\|$, so that the determinant is $\prod_i \|\mathcal{D}_3 u_i\| \prod_j \|\mathcal{D}_3 v_j\| \det(N_k)$, where N_k is a $k \times k$ random matrix with independent standard normal entries, so that taking logarithms, we see $\Xi_k(\Pi\Delta\Pi') = -\mathcal{E}_k(\Pi) - \mathcal{E}_k(\Pi') + \log|\det N_k|$.

Replacing Δ with a conditioned version has the effect of multiplying the density of N_k by a factor in the range $[0,2]$. Since $\log|\det(N_k)|$ is an integrable function, there are uniform upper and lower bounds for $\int \log|\det(N_k)|\rho(N_k)$ over all functions ρ taking values in $[0,2]$, so that

$$\left|\mathbb{E}\left(\Xi_k(\Pi\Delta\Pi')\middle|\Delta\in Q\right)+\mathcal{E}_k(\Pi)+\mathcal{E}_k(\Pi')\right|\le C,$$

as required. ☐

Corollary 4. *There exists $K>0$ such that if Π' and Π'' are two orthogonal projections and $Q\subset \mathsf{HS}$ satisfies $\gamma(Q)>\frac{1}{2}$, then*

$$\left|\mathbb{E}\left(\Xi_k(\Pi\Delta\Pi')|\Delta\in Q\right)-\left(\mathbb{E}\,\Xi_k(\Pi\Delta\Pi_k)+\mathbb{E}\,\Xi_k(\Pi_k\Delta\Pi')\right)\right|<C$$

Proof. By Lemma 3, we have the following

$$\begin{aligned}&\left|\mathbb{E}\left(\Xi_k(\Pi\Delta\Pi')\middle|\Delta\in Q\right)+\mathcal{E}_k(\Pi)+\mathcal{E}_k(\Pi')\right|\le C;\\&\left|\mathbb{E}\,\Xi_k(\Pi\Delta\Pi_k)+\mathcal{E}_k(\Pi)+\mathcal{E}_k(\Pi_k)\right|\le C;\\&\left|\mathbb{E}\,\Xi_k(\Pi_k\Delta\Pi')+\mathcal{E}_k(\Pi)+\mathcal{E}_k(\Pi_k)\right|\le C,\end{aligned}$$

where C is the constant from Lemma 3.

We calculate that $\mathcal{E}_k(\Pi_k)=\frac{1}{2}k(k-1)\log 3$, so that combining the inequalities, we obtain

$$\left|\mathbb{E}\left(\Xi_k(\Pi\Delta\Pi')\middle|\Delta\in Q\right)-\left(\mathbb{E}\,\Xi_k(\Pi\Delta\Pi_k)+\mathbb{E}\,\Xi_k(\Pi_k\Delta\Pi')\right)\right|\le K,$$

where $K=3C+k(k-1)\log 3$. ☐

4. Good Blocks

This section deals with *good blocks*. The strategy we follow goes back to Ledrappier and Young in the context of invertible matrices [21, Lemmas 3.3, 3.6 & 4.3], and it was later used in [10]. Lemma 7 is the main tool to control the effect of perturbations on good blocks. Lemma 8 collects standard facts about Lyapunov exponents, Oseledets splittings and their approximations via singular vectors, which are used to define good blocks. Lemma 9 establishes the conditions defining *tame perturbations*. Proposition 10 provides a lower bound on Ξ_k over a sequence of tame perturbations, comparable with Ξ_k for the unperturbed cocycle.

For each $k\in\mathbb{N}$, we define $E_k(A)$ to be the space spanned by the *images* of the singular vectors with k largest singular values under A, and $F_k(A)$ to be the space spanned by the orthogonal complement of the pre-image of $E_k(A)$ under A. Thus, $F_k(A)$ is exactly the space spanned by those singular vectors of A whose singular value is not amongst the k largest. We note that the spaces $F_k(A), E_k(A)$ are uniquely defined when the singular values $s_k(A)$ and $s_{k+1}(A)$ are distinct. We will always use our results in this setting, and therefore do not worry about the possibility of non-uniqueness.

We collect some properties of singular values and singular vectors for compact operators on Hilbert spaces and matrices.

Lemma 5. *Let A be a compact operator on a Hilbert space, H. Let the singular values be $s_1(A), s_2(A), \ldots$.*
(a) $s_j(A) = \min_{V\in\mathcal{G}^{j-1}(H)} \max_{x\in V\cap S} \|Ax\|$;
(b) $s_j(A) = \max_{V\in\mathcal{G}_j(H)} \min_{x\in V\cap S} \|Ax\|$;
(c) $|s_j(A) - s_j(B)| \le \|A-B\|_{\mathsf{op}}$;

Proof. The characterizations (a) and (b) are well known.

To show (c), using (b), let V be a j-dimensional space such that $\|Ax\| \ge s_j(A)$ for all $x \in V\cap S$. Then $\|Bx\| \ge s_j(A) - \|A-B\|_{\mathsf{op}}$ for all $x\in V\cap S$, so that using (b) again, we see $s_j(B) \ge s_j(A) - \|A-B\|_{\mathsf{op}}$. By symmetry, $s_j(A) \ge s_j(B) - \|A-B\|_{\mathsf{op}}$, giving the result. □

Lemma 6. *Let $U \in \mathcal{G}^k(H)$ and $V \in \mathcal{G}_k(H)$. Then $s_k(\Pi_{U^\perp}\Pi_V) \ge \perp(U,V)$.*

Proof. Choose $v\in V$ with $\|v\|=1$. Let $v = u+w$ with $u\in U$ and $w\in U^\perp$. Let $\hat u \in U\cap S$ be such that $u = \|u\|\hat u$ ($\hat u$ may be chosen arbitrarily if $u=0$) and let θ be the angle between $\hat u$ and v, so that $0<\theta\le\frac{\pi}{2}$. By assumption $\|\hat u - v\| \ge \sqrt{2}\perp(U,V)$. We have $\|\hat u - v\| = 2\sin\frac{\theta}{2}$. Notice that $\|w\| = \|\Pi_{U^\perp}v\| = \sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} \ge \sqrt{2}\perp(U,V)\cos\frac{\theta}{2}$. Since $\theta \le \frac{\pi}{2}$, we see $\|\Pi_{U^\perp}v\| \ge \perp(U,V)$ for all $v\in V\cap S$. □

Lemma 7. *For any $\delta < \frac12$, there exists a $K > \delta^{-(4k+3)}$ such that if (i) the kth singular value of a compact linear operator $A : X \to X$ exceeds K; (ii) the $(k+1)$st singular value of A is at most 1; and (iii) $\|B-A\| \le 1$, then the following hold:*
(a) $e^{-\delta} \le s_j(A)/s_j(B) \le e^{\delta}$ *for each* $j\le k$ *and* $s_j(B)\le 2$ *for each* $j>k$;
(b) $\angle(E_k(A), E_k(B))$ *and* $\angle(F_k(A),F_k(B))$ *are less than* δ;
(c) If V is any subspace of dimension k such that $\perp(V, F_k(A)) > \delta$, then $\angle(BV, E_k(A)) < \delta$;
(d) If V is a subspace of dimension k and $\perp(V,F_k(A)) > 2\delta$, then $|\det(B|_V)| \ge \delta^k \exp \Xi_k(B)$.

Proof. For each closed subspace W of X, let $\Pi_W : X\to W$ be the orthogonal projection onto W.

(a) For the first part, notice that by assumption, for $j\le k$, we have $s_j(A)\ge K$. Also by Lemma 5(c), we have $|s_j(A)-s_j(B)| \le 1$, so that $\frac{K}{K+1} \le s_j(A)/s_j(B) \le \frac{K}{K-1}$. The second part of the claim follows from Lemma 5(c) also.

(b) Let $K > 1 + \frac{6}{\delta}$. For symmetry, in this part, we assume only $s_k(A), s_k(B) \ge K - 1$, $s_{k+1}(A), s_{k+1}(B) \le 2$ and $\|A - B\|_{\mathsf{op}} \le 1$.

Let $v \in S$ satisfy $d(v, F_k(A) \cap S) \ge \delta$. We will show that $v \notin F_k(B)$. Let $v = u + w$ with $u \in F_k(A)$ and $w \in F_k(A)^\perp$. By assumption, $\|w\| \ge \frac{\delta}{2}$, so that $\|Bv\| \ge \|Av\| - 1 \ge \|Aw\| - 1 > (K-1)\|w\| - 1 > 2$. On the other hand, if $v \in F_k(B)$, then $\|Bv\| \le s_{k+1}(B) \le 2$. The identical argument shows that if $v \in F_k(A) \cap S$, then $d(v, F_k(B) \cap S) < \delta$

To show the closeness of the fast spaces, first let $v \in F_k(B)^\perp \cap S$, and write v as $au + w$, where $u \in F_k(A) \cap S$ and $w \in F_k(A)^\perp$. Let $u' \in F_k(B) \cap S$ satisfy $\|u - u'\| < \delta$ (such a u' exists by the paragraph above). Now $\langle v, u\rangle = \langle v, u'\rangle + \langle v, u - u'\rangle$. The first term is 0 and the second term is less than δ in absolute value. Hence $|a| < \delta$ and $\|w\| \ge \frac{1}{2}$. Now $Bv = aAu + Aw + (B - A)v$. In particular, $\|Bv - Aw\| \le 2\delta + 1 \le 2$ while $\|Bv\| \ge K - 1$. Hence if $z \in E_k(B) \cap S$, we have $d(z, E_k(A)) \le 2/(K-1)$, so $d(z, E_k(A) \cap S) \le 4/(K-1)$. The identical argument holds if the roles of A and B are reversed, so $\angle(E_k(A), E_k(B)) < 4/(K-1) < \delta$.

(c) Let $K > 4/\delta^2 + 2/\delta$. Let $v \in V \cap S$ and write $v = u + w$ with $u \in F_k(A)$ and $w \in F_k(A)^\perp$. By assumption, $\|w\| \ge \delta$. Hence $\|Aw\| \ge K\delta$, while $\|Au\| \le 1$. Since $\|B - A\|_{\mathsf{op}} \le 1$, we have $\|Bv - Aw\| \le \|Bv - Av\| + \|Av - Aw\| \le 2$, so that $\|Bv - Aw\|/\|Bv\| \le 2/(K\delta - 2)$. Hence for an arbitrary element, y of $BV \cap S$, we have $d(y, E_k(A)) \le 2/(K\delta - 2) < \frac{\delta}{2}$ and $d(y, E_k(A) \cap S) \le 4/(K\delta - 2) < \delta$. By Lemma 2, we deduce that $\angle(BV, E_k(A)) < \delta$ as required.

(d) We have that $\log|\det(B|_V)| \ge \Xi_k(\Pi_{E_k(B)} B|_V) = \Xi_k(B\Pi_{F_k(B)^\perp}|_V) = \Xi_k(B\Pi_{F_k(B)^\perp}\Pi_V) = \Xi_k(B\Pi_{F_k(B)^\perp}) + \Xi_k(\Pi_{F_k(B)^\perp}\Pi_V) \ge \Xi_k(B) + k\log\delta$. The last inequality follows from the facts that $\Xi_k(B\Pi_{F_k(B)^\perp}) = \Xi_k(B)$; and $\perp(F_k(B)^\perp, V) \ge \perp(F_k(A)^\perp, V) - \angle(F_k(A)^\perp, F_k(B)^\perp) > \delta$ so that $\|\Pi_{F_k(B)^\perp}\Pi_V v\| \ge \delta\|v\|$ for every $v \in V$ by Lemma 6, hence $\Xi_k(\Pi_{F_k(B)^\perp}\Pi_V) \ge k\log\delta$. The claim follows.

□

The following lemma underlies the definition of good blocks: Using the notation of the lemma, if $n \ge n_0$ and $\omega \in G$, and we say the block $A_\omega^{(n)}$ is *good*. See [10, Lemma 2.4] for a proof in the context of matrix cocycles, which applies without changes in our setting.

Lemma 8 (Good blocks)**.** *Let σ be an invertible ergodic measure-preserving transformation of $(\Omega,\mathbb{P})$ and let $A\colon \Omega\to \mathsf{SHS}$ be a measurable map, taking values in the strong Hilbert-Schmidt operators on H, and such that $\int \log^+\|A(\omega)\|_{\mathsf{SHS}}\,d\mathbb{P}(\omega)<\infty$. Let the Lyapunov exponents of the cocycle A be $\infty>\mu_1\ge\mu_2\ge\ldots\ge-\infty$, counted with multiplicities. Suppose $k\ge 1$ is such that $\mu_k>0>\mu_{k+1}$. Let $E_k(\omega)$ and $F_k(\omega)$ denote the k-dimensional and k-codimensional Oseledets spaces of A at ω corresponding to Lyapunov exponents $\mu_1\ge\cdots\ge\mu_k$ and $\mu_{k+1}\ge\ldots$, respectively.*

Let $\xi>0$ and $\delta_1>0$ be given. Then there exist $n_0>0$, $\tau\le\min(\delta_1,\frac14\mu_k)$ and $0<\delta\le\delta_1$ such that: for all $n\ge n_0$, there exists a set $G\subseteq\Omega$ with $\mathbb{P}(G)>1-\xi$ such that for $\omega\in G$, we have

(a) $\perp(F_k(\omega),E_k(\omega))>10\delta$*;*

(b) $\angle(E_k(A^{(n)}_\omega),E_k(\sigma^n\omega))<\delta$*;*

(c) $\angle(F_k(A^{(n)}_\omega),F_k(\omega))<\delta$*;*

(d) $e^{(\mu_k+\tau)n}>s_k(A^{(n)}_\omega)>\max(K(\delta),e^{(\mu_k-\tau)n})$ *and* $s_{k+1}(A^{(n)}_\omega)<1$*, where $K(\delta)$ is as given in Lemma 7.*

(e) $\frac1n\sum_{i=0}^{n-1}\log(1+\|A_{\sigma^i\omega}\|_{\mathsf{SHS}})<2\int\log(1+\|A_\omega\|_{\mathsf{SHS}})\,d\mathbb{P}(\omega)$.

Assume that $\epsilon>0$ is fixed. A perturbation Δ is said to be *tame* if $|\Delta_{s,t}|\le\epsilon^{-1/2}(\frac23)^{s+t}$ for all s,t (otherwise Δ is *wild*). A quick calculation shows that if Δ is tame, then $\|\epsilon\Delta\|_{\mathsf{HS}}<2\sqrt{\epsilon}$.

Lemma 9 (Good block length)**.** *Let $\sigma:(\Omega,\mathbb{P})\circlearrowleft$ be an ergodic measure-preserving transformation. Let $A:\Omega\to\mathcal{B}(H)$ be a measurable map, taking values in the bounded linear operators on H, such that $\log^+\|A(\omega)\|_{\mathsf{op}}$ is integrable. There exists $C_9>0$ such that for all $\eta_0>0$, there exists ϵ_0 such that for all $\epsilon<\epsilon_0$, there exists $G\subseteq\Omega$ of measure at least $1-\eta_0$ such that for all $\omega\in G$, if $(\Delta_n)\in\mathsf{HS}^{\mathbb{Z}}$ satisfies Δ_n is tame for each $0\le n<N$, then*

$$\|A^{\epsilon(N)}_{\bar\omega}-A^{(N)}_\omega\|_{\mathsf{op}}\le 1,$$

where $\bar\omega=(\omega,(\Delta_n))$, $N=\lfloor C_9|\log\epsilon|\rfloor$ and $A^{\epsilon(N)}_{\bar\omega}=A^\epsilon_{N-1}(\bar\omega)\ldots A^\epsilon_1(\bar\omega)A^\epsilon_0(\bar\omega)$.

The probability that one of $\Delta_0,\ldots,\Delta_{N-1}$ is wild is $O(e^{-1/(2\epsilon)})$.

Proof. Let $g(\omega)=\log^+(\|A_\omega\|_{\mathsf{op}}+1)$ and let $C>0$ satisfy $\int g(\omega)\,d\mathbb{P}(\omega)<1/(2C)$. Notice that provided $\epsilon<\frac14$ (and assuming that the perturbations $(\Delta_n)_{0\le n<N}$ are tame, so that $\|\epsilon\Delta_n\|_{\mathsf{op}}\le\|\epsilon\Delta_n\|_{\mathsf{HS}}\le 2\sqrt{\epsilon}$ for $0\le n<N$), $\log^+\|A^\epsilon_{\bar\sigma^n\bar\omega}\|\le g(\sigma^n\omega)$ for each $0\le n<N$, and

$$\begin{aligned}\|A^{\epsilon(N)}_{\bar\omega}-A^{(N)}_\omega\|_{\mathsf{op}}&\le\sum_{i=0}^{N-1}\|A^{\epsilon(N-i-1)}_{\bar\sigma^i\bar\omega}(A^\epsilon_{\bar\sigma^i\bar\omega}-A_{\sigma^i\omega})A^{(i)}_\omega\|_{\mathsf{op}}\\&\le 2N\sqrt{\epsilon}\exp(g(\omega)+\ldots+g(\sigma^{N-1}\omega)).\end{aligned}$$

There exists n_0 such that for $N \geq n_0$, $g(\omega)+\ldots+g(\sigma^{N-1}\omega) \leq N/(2C)-\log(4N)$ on a set of measure at least $1-\eta_0$, hence $2N\sqrt{\epsilon}\exp(g(\omega)+\ldots+g(\sigma^{N-1}\omega)) \leq \frac{1}{2}\sqrt{\epsilon}\exp(N/(2C))$ on a set of measure at least $1-\eta_0$. In particular, provided $\lfloor C|\log\epsilon_0|\rfloor > n_0$, taking $N=\lfloor C|\log\epsilon|\rfloor$, we have $\|A^{\epsilon(N)}_{\bar\omega} - A^{(N)}_{\omega}\|_{\mathsf{op}} \leq 1$ provided that the perturbations $\Delta_0,\ldots\Delta_{N-1}$ are all tame.

Recall that (i,j)th entry of Δ is distributed as $3^{-(i+j)}$ times a standard normal random variable. Hence the probability that $|\Delta_{i,j}| > \epsilon^{-1/2}(\frac{2}{3})^{i+j}$ is $\mathbb{P}(|N| > \epsilon^{-1/2}2^{i+j})$. Using a standard estimate on the tail of a normal random variable [7, Theorem 1.2.3], this is at most $\frac{2\sqrt{\epsilon}}{\sqrt{2\pi}}2^{-(i+j)}\exp(-2^{2i+2j-1}/\epsilon)$.

In particular, using the union bound, the probability that one of $\Delta_0,\ldots,\Delta_{N-1}$ is wild is $O(e^{-1/(2\epsilon)})$. □

We comment that once $\xi > 0$ and $\delta_1 > 0$ are fixed, Lemma 8 guarantees the existence of an n_0 such that for all sufficiently large n, the good set defined in the lemma has measure at least $1-\xi$. Now for ϵ sufficiently small, the length $N = \lfloor C_9|\log\epsilon|\rfloor$ exceeds n_0. For the remainder of the proof, we let G be the good set from Lemma 8 with n taken to be N (so that the good set, G, depends on ξ, δ_1 and ϵ, but this dependence will not be made explicit). We further introduce the notation $\bar G = G \cap \bigcap_{i=0}^{N-1}\{\Delta_i \text{ is tame}\}$, which we shall also use for the remainder of the proof.

Proposition 10 (Glueing good blocks)**.** *Under the assumptions of Lemma 8, suppose* $j<l$ *and* $\bar\sigma^{jN}\bar\omega, \bar\sigma^{(j+1)N}\bar\omega,\ldots,\bar\sigma^{(l-1)N}\bar\omega \in \bar G$. *Then,*

$$\Xi_k(A^{\epsilon((l-j)N)}_{\bar\omega}) \geq \Xi_k(A^{((l-j)N)}_{\omega}) + 2(l-j)k\log\delta. \tag{1}$$

Proof. Let $B_n = A^{(N)}_{\sigma^{nN}\omega}$ and $\tilde B_n = A^{\epsilon(N)}_{\bar\sigma^{nN}\bar\omega}$. This is proved by induction using Lemma 7. Recall that since B_n is a good block, $\|B_n - \tilde B_n\| \leq 1$ by Lemma 9. We let $\tilde V_j = V_j = F_k(B_j)^\perp$ and define $V_{n+1} = B_nV_n$ and $\tilde V_{n+1} = \tilde B_n\tilde V_n$.

We claim that the following hold, for each $n = j, j+1,\ldots,l-1$:

(i) $\angle(V_n,\tilde V_n) < 2\delta$;
(ii) $\perp(V_n, F_k(B_n)) > \delta$ and $\perp(\tilde V_n, F_k(B_n)) > \delta$.

Item (i) and the first part of (ii) hold immediately for the case $n=j$. The second part of (ii) holds because $\tilde V_j = V_j = F_k(B_j)^\perp$ and $\angle(F_k(B_j), F_k(\tilde B_j)) < \delta$ by Lemma 7(b).

Given that (i) and (ii) hold for $n=m$, B_m is a good block and $\tilde B_m$ is a good perturbation, Lemma 7(c) implies that $\angle(V_{m+1}, E_k(B_m)) < \delta$,

$\angle(\tilde{V}_{m+1}, E_k(B_m)) < \delta$ so that $\angle(\tilde{V}_{m+1}, V_{m+1}) < 2\delta$, yielding (i) for $n = m+1$.

Making use of the induction hypothesis and Lemma 8, we have that $\angle(E_k(\sigma^{(m+1)N}\omega), E_k(B_m)) < \delta, \angle(F_k(\sigma^{(m+1)N}\omega), F_k(B_m)) < \delta$ and $\perp (E_k(\sigma^{(m+1)N}\omega), F_k(\sigma^{(m+1)N}\omega)) > 10\delta$. Thus, we obtain (ii) for $n = m+1$.

Hence using Lemma 7(d), we see that $\log|\det(\tilde{B}_n|_{\tilde{V}_n})| \geq k\log\delta + \Xi_k(\tilde{B}_n) \geq k\log\delta - k\delta + \Xi_k(B_n) \geq \Xi_k(B_n) + 2k\log\delta$, where we made use of Lemma 7(a) for the second inequality.

Since $\Xi_k(\tilde{B}_{l-1}\cdots\tilde{B}_j) \geq \sum_{i=j}^{l-1}\log|\det(\tilde{B}_i|\tilde{V}_i)|$, summing yields

$$\begin{aligned}\Xi_k(\tilde{B}_{l-1}\cdots\tilde{B}_j) &\geq 2(l-j)k\log\delta + \sum_{i=j}^{l-1}\Xi_k(B_i)\\ &\geq 2(l-j)k\log\delta + \Xi_k(B_{l-1}\cdots B_j),\end{aligned} \tag{2}$$

as required. □

Lemma 11. *Let the Hilbert-Schmidt cocycle, $A\colon \Omega \to \mathsf{HS}$ and all parameters and perturbations be as above. If $\bar{\sigma}^{iN}\omega \in \bar{G}$ for each $0 \leq i < n$, then $\angle\left(F_k(A^{\epsilon(nN)}_{\bar{\omega}}), F_k(A^{(N)}_{\omega})\right) < \delta$.*

Proof. By the first part of (2), $\Xi_k(A^{\epsilon(nN)}_{\bar{\omega}}) > \sum_{i=0}^{n-1}\Xi_k(A^{(N)}_{\sigma^{iN}\omega}) + 2nk\log\delta$. Also, $\Xi_{k+1}(A^{\epsilon(nN)}_{\bar{\omega}}) \leq \sum_{i=0}^{n-1}\Xi_{k+1}(A^{\epsilon(N)}_{\bar{\sigma}^{iN}\bar{\omega}}) \leq \sum_{i=0}^{n-1}\Xi_k(A^{\epsilon(N)}_{\bar{\sigma}^{iN}\bar{\omega}}) + n\log 2 \leq \sum_{i=0}^{n-1}\Xi_k(A^{(N)}_{\sigma^{iN}\omega}) + n\log 2 + nk\delta \leq \sum_{i=0}^{n-1}\Xi_k(A^{(N)}_{\sigma^{iN}\omega}) - 2nk\log\delta$ by Lemma 7(a). Since we have $\log s_{k+1}(A^{\epsilon(nN)}_{\bar{\omega}}) = \Xi_{k+1}(A^{\epsilon(nN)}_{\bar{\omega}}) - \Xi_k(A^{\epsilon(nN)}_{\bar{\omega}})$, we deduce $s_{k+1}(A^{\epsilon(nN)}_{\bar{\omega}}) \leq \delta^{-4nk}$.

On the other hand, if $v \in S$ is such that $\perp(v, F_k(A^{(N)}_{\omega})) > \delta$, an inductive argument exactly like the proof of Proposition 10 shows that $\|A^{\epsilon(nN)}_{\bar{\omega}}v\| \geq (\frac{\delta}{3})^n e^{-n\delta}\prod_{i=0}^{n-1}s_k(A^{(nN)}_{\omega}) \geq (\delta^3 K(\delta))^n$. The choice of $K(\delta)$ in Lemma 7 ensures $\|A^{\epsilon(nN)}_{\bar{\omega}}v\| > s_{k+1}(A^{\epsilon(nN)}_{\bar{\omega}})$, so that $v \notin F_k(A^{\epsilon(nN)}_{\bar{\omega}})$. □

Proposition 12. *Let ω be such that $\bar{\sigma}^{iN}\bar{\omega} \in \bar{G}$ for $0 \leq i < n$. Then for any V such that $\perp(V, F_k(A^{(N)}_{\omega})) > 2\delta$, one has $\log|\det(A^{\epsilon(nN)}_{\bar{\omega}}|_V)| \geq \Xi_k(A^{\epsilon(nN)}_{\bar{\omega}}) + k\log\delta$.*

Proof. We argue as in Lemma 7(d).

$$\begin{aligned}\log|\det(A^{\epsilon(nN)}_{\bar\omega}|_V)| &\geq \Xi_k(\Pi_{E_k(A^{\epsilon(nN)}_{\bar\omega})}A^{\epsilon(nN)}_{\bar\omega}\Pi_V)\\ &= \Xi_k(A^{\epsilon(nN)}_{\bar\omega}\Pi_{F_k(A^{\epsilon(nN)}_{\bar\omega})^\perp}\Pi_V)\\ &= \Xi_k(A^{\epsilon(nN)}_{\bar\omega}\Pi_{F_k(A^{\epsilon(nN)}_{\bar\omega})^\perp}) + \Xi_k(\Pi_{F_k(A^{\epsilon(nN)}_{\bar\omega})^\perp}\Pi_V)\\ &\geq \Xi_k(A^{\epsilon(nN)}_{\bar\omega}) + k\log\perp(F_k(A^{\epsilon(nN)}_{\bar\omega}),V),\end{aligned}$$

where we used Lemma 6 for the last line. Lemma 11 and the triangle inequality allow us to conclude. ☐

5. Comparing perturbed and unperturbed bad blocks (Type I)

We distinguish two ways in which a block can be bad: types I and II. A *type I bad block* is one where the unperturbed cocycle has bad properties. On the other hand, a *type II bad block* is one where the unperturbed cocycle is well-behaved, but the perturbations are wild.

Conditional on being in a type I bad block, the perturbations are unconstrained, whereas conditional on being in a type II bad block at least one perturbation is constrained to be large. For later use with the type II bad blocks, we state some of the lemmas when one is conditioned to be in a high probability event (but the high probability event will be taken to be the whole space when dealing with type I blocks.)

Lemma 13. *Let $k > 0$. There exists a $C > 0$ with the following property. Let T be a multi-variate normal Hilbert-Schmidt-valued random operator whose entries have mean 0, let $A \in \mathsf{HS}$ and let Π and Π' be orthogonal projections onto k-dimensional subspaces of H. Then for any subset Q of HS such that $\mathbb{P}(T \in Q) \geq \frac{1}{2}$, one has*

$$\mathbb{E}\left(\left(\Xi_k(\Pi(A+T)\Pi') - \Xi_k(\Pi A\Pi')\right)^- \Big| T \in Q\right) \geq -C,$$

where x^- denotes $\min(x, 0)$.

Proof. We assume $\Xi_k(\Pi A\Pi') > -\infty$ as otherwise the result is trivial.

Let $\tilde\Pi$ be Π composed with an isometry from the range of Π to $\mathbb{R}^k$ and similarly let $\tilde\Pi'$ be an isometry from $\mathbb{R}^k$ to the range of Π'. Then we have $\Xi_k(\Pi B\Pi') = \log|\det(\tilde\Pi B\tilde\Pi')|$ for any bounded operator B on H so that we need to show

$$\mathbb{E}\left(\left(\log|\det(\tilde\Pi(A+T)\tilde\Pi')| - \log|\det(\tilde\Pi A\tilde\Pi')|\right)^- \Big| T \in Q\right) \geq -C.$$

Let $Y = \tilde\Pi A\tilde\Pi'$ and $Z = \tilde\Pi T\tilde\Pi'$, so that Y is a fixed $k \times k$ matrix and Z is a $k \times k$ matrix-valued random variable with multivariate normal entries. By our earlier assumption, Y is invertible, so let $X = ZY^{-1}$ (this also has multi-variate normal entries for unconditioned T). We then need a lower bound for $\mathbb{E}\left(\log\det(I+X)\middle|T\in Q\right)$.

The unconstrained matrix-valued random variable X can be written as $\sum_{l=1}^{d} N_l B^l$, where the B^l are fixed $k\times k$ matrices, d is the dimension of the support of X (at most k^2 depending on the pattern of entries in the unperturbed A's) and the N_l are independent standard normal random variables (see for example [7, Example 3.9.2]).

Let Ψ denote the map from $\mathbb{R}^d$ to $M_{k\times k}$ defined by $x\mapsto \sum x_l B^l$. Let $\mathcal{S}$ be the image under Ψ of the unit sphere and μ be the measure on $\mathcal{S}$ that is the push-forward of the normalized volume measure on the unit sphere. The unconditioned measure on X is then the push forward of $\mu\times C_d r^{d-1}e^{-r^2/2}\,dr$, where C_d is chosen so that $C_d\int r^{d-1}e^{-r^2/2}\,dr = 1$. The conditioned measure on X (since the event being conditioned upon is of measure at least $\frac{1}{2}$) is of the same form, but the density is multiplied by a varying factor in the range [0,2].

It then suffices to lower bound

$$2C_d\int_{\mathcal{S}} d\mu(M)\int_0^\infty \log^-|\det(I+rM)|\,r^{d-1}e^{-r^2/2}\,dr.$$

In particular, it is enough to give a uniform lower bound for

$$G(d,M)=\int_0^\infty \log^-|\det(I+rM)|\,r^{d-1}e^{-r^2/2}\,dr$$

as d ranges over the range 1 to k^2 and M ranges over $M_{k\times k}$.

For each fixed M, write $p_M(r)=\det(I+rM)$, so that p_M is a polynomial of degree k satisfying $p_M(0)=1$. Hence $p_M(r)$ can be written as a product $\prod_{i=1}^{k}(1-b_i r)$. Define

$$F(d,b)=\int_0^\infty \log^-|1-br|\,r^{d-1}e^{-r^2/2}\,dr,$$

so that $G(d,M)\geq\sum_{i=1}^{k}F(d,b_i)$. Hence it suffices to show that $F(d,b)$ is uniformly bounded below as b runs over the complex plane and as d runs over the range 1 to k^2.

Next, notice that $\log|1-br|\geq\log|1-\mathrm{Re}(b)r|$, so $F(d,b)\geq F(d,|b|)$ and it suffices to give a lower bound for positive real values of b. Also

$$\begin{aligned}F(d,b)&=\frac{1}{b^d}\int_0^\infty \log^-|1-r|\,r^{d-1}e^{-r^2/(2b^2)}\,dr\\&=\frac{1}{b^d}\int_0^2 \log|1-r|\,r^{d-1}e^{-r^2/(2b^2)}\,dr.\end{aligned}$$

For $b \geq \frac{1}{2}$, $F(b,d) \geq \frac{1}{b^d}\int_0^2 \log|1-r| r^{d-1}\,dr \geq -2^d/b^d \geq -4^d$. For $0 < b < \frac{1}{2}$, one has

$$\begin{aligned} &F(d,b)\\ &\geq \frac{1}{b^d}\int_0^{b^{d/(1+d)}} \log|1-r| r^{d-1}\,dr + \frac{1}{b^d}\int_{b^{d/(1+d)}}^2 \log|1-r| r^{d-1} e^{-r^2/(2b^2)}\,dr\\ &\geq -2\frac{1}{b^d}\int_0^{b^{d/(1+d)}} r^d\,dr + (2/b)^d \exp(-1/(2b^\alpha))\int_0^2 \log|1-r|\,dr\\ &\geq -2/(d+1) - 2^{d+1}\exp(-1/(2b^\alpha))/b^d \end{aligned}$$

where $\alpha = 2/(1+d)$. This converges to $-2/(d+1)$ as b approaches 0 from the right. By continuity and compactness, for each of the finitely many values of d, $F(d,b)$ is bounded below as b ranges over $(0, \frac{1}{2}]$. □

Proposition 14. *Let $k > 0$. Then there exists a C_{14} with the following property. For every finite sequence $A_0, \dots, A_{n-1}$ of Hilbert-Schmidt operators, let $\Delta_0, \dots, \Delta_{n-1}$ be independent copies of the perturbation Δ as described above. Let A_i^ϵ denote $A_i + \epsilon\Delta_i$.*

Then one has

$$\mathbb{E}_{\Delta_0,\dots,\Delta_{n-1}}\left(\tilde{\Xi}_k(A_{n-1}^\epsilon \cdots A_0^\epsilon) - \tilde{\Xi}_k(A_{n-1}\cdots A_0)\right)^- \geq -C_{14}n.$$

Proof. We have

$$\begin{aligned} &\mathbb{E}\left(\tilde{\Xi}_k(A_{n-1}^\epsilon \cdots A_0^\epsilon) - \tilde{\Xi}_k(A_{n-1}\cdots A_0)\right)^-\\ &\geq \sum_{j=0}^{n-1}\mathbb{E}\left(\tilde{\Xi}_k(A_{n-1}^\epsilon\cdots A_j^\epsilon A_{j-1}\cdots A_0) - \tilde{\Xi}_k(A_{n-1}^\epsilon\cdots A_{j+1}^\epsilon A_j\cdots A_0)\right)^-. \end{aligned}$$

We focus on giving a lower bound for one of the terms in the summation. We write such a term as

$$\mathbb{E}_{\Delta_j}\left(\tilde{\Xi}_k(L(A_j+\epsilon\Delta)R) - \tilde{\Xi}_k(LA_jR)\right)^-.$$

This expectation should be interpreted as being conditioned on the values of $\Delta_{j+1}, \dots, \Delta_n$, so that $L = (A_n + \epsilon\Delta_n)\cdots(A_{j+1}+\epsilon\Delta_{j+1})$.

The above expectation can be rewritten as:

$$\mathbb{E}_{\Delta,\Delta'}\,\mathbb{E}_{\Delta_j}\left[\Xi_k(\Pi_k\Delta L(A+\epsilon\Delta_j)R\Delta'\Pi_k) - \Xi_k(\Pi_k\Delta LAR\Delta'\Pi_k)\right]^-. \tag{3}$$

Once Δ and Δ' are fixed, the inner expectation is

$$\mathbb{E}_{\Delta_j}\left[\Xi_k(\Pi_k\Delta L(A+\epsilon\Delta_j)R\Delta'\Pi_k) - \Xi_k(\Pi_k\Delta LAR\Delta'\Pi_k)\right]^-. \tag{4}$$

Now let Π be the orthogonal projection onto the orthogonal complement of the kernel of $\Pi_k\Delta L$ and Π' be the orthogonal projection onto the range of $R\Delta'\Pi_k$. Then we have

$$\Xi_k(\Pi_k\Delta L(A+\epsilon\Delta_j)R\Delta'\Pi_k) = \Xi_k(\Pi_k\Delta L) + \Xi_k(\Pi(A+\epsilon\Delta_j)\Pi') + \Xi_k(R\Delta'\Pi_k);$$
$$\Xi_k(\Pi_k\Delta LAR\Delta'\Pi_k) = \Xi_k(\Pi_k\Delta L) + \Xi_k(\Pi A\Pi') + \Xi_k(R\Delta'\Pi_k);$$

Now the quantity in (4) is

$$\mathbb{E}_{\Delta_j}\left[\Xi_k(\Pi(A+\epsilon\Delta_j)\Pi') - \Xi_k(\Pi A\Pi')\right]^-$$

Applying Lemma 13 with $Q = \mathsf{HS}$, this is bounded below by $-C$, independently of Δ and Δ', so that the quantity in (3) is also bounded below by $-C$. Since there are n such terms, the statement in the lemma follows. □

6. Type II bad block perturbations

Here we give an argument for good blocks in the base that have large perturbations. We will obtain a drop in $\tilde{\Xi}_k$ over a bad block of size $O(\log\epsilon)$ at worst, that is a drop of size $O(1)$ per symbol since blocks are of length proportional to $|\log\epsilon|$. However since the frequency of these blocks is $O(e^{-C/\epsilon})$, the contribution of this drop to the singular values of a large string of blocks is minuscule.

Lemma 15. *There exists a constant $C>0$ such that if N is a standard normal random variable and $\Lambda > 2$, then for each $a \in \mathbb{C}$,*

$$\mathbb{E}\left(\log^-|1-aN|\middle|N\geq\Lambda\right) \geq -C\log\Lambda.$$

Before giving the proof, let us give a heuristic explanation for why this should be true. Conditional on $N \geq \Lambda$, the distribution of N is approximately $\Lambda + \mathrm{Exp}(\Lambda)$, that is it typically takes values that are $\Lambda + O(1/\Lambda)$. The worst case for the inequality is approximately when $a = 1/\Lambda$ and then the quantity inside the logarithm is roughly $O(1/\Lambda^2)$.

Proof. We first recall that $\int_0^a \log x\,dx = a(\log a - 1)$, so that the average value of the logarithm function over $[0,a]$ is $\log a - 1$. We claim that for any interval J, one has

$$\frac{1}{|J|}\int_J \log^-|x|\,dx \geq 2(\max_{x\in J}\log^-|x| - 1). \tag{5}$$

Indeed, this follows already for intervals $[0,a]$ with $0<a<1$, and hence for sub-intervals of $[0,1]$ and $[-1,0]$. For intervals $[-a,b]$ with $a<0<|a|\leq b\leq 1$, we have $1/(a+b)\int_{-a}^b \log^-|x|\,dx \geq 1/b\int_{-b}^b\log^-|x|\,dx = 2(\log b - 1)$. If the interval J is entirely outside $[-1,1]$, the inequality is trivial; and if J intersects $[-1,1]$, we have already established the inequality for $J\cap[-1,1]$, from which the inequality for J follows.

For $a \in \mathbb{C}$, the integrand in the statement reduced if a is replaced by $|a|$ so we may assume $a > 0$. If $a > 2/\Lambda$, the integral is 0.

If $1/(3\Lambda) \le a \le 2/\Lambda$, let $I = [\Lambda, \frac{2}{a})$, the sub-interval of $[\Lambda, \infty)$ where $\log|1-ax| < 0$; and $J = [\frac{1}{a} - \frac{1}{a\Lambda^2}, \frac{1}{a} + \frac{1}{a\Lambda^2}]$, the interval where $\log|1-ax| < -2\log\Lambda$.

The quantity to be bounded is

$$\frac{\int_I \log^-|1-ax|e^{-x^2/2}\,dx}{\int_\Lambda^\infty e^{-x^2/2}\,dx} \ge \frac{\int_I \log^-|1-ax|e^{-x^2/2}\,dx}{\int_I e^{-x^2/2}\,dx}$$

$$= \frac{\int_{I\cap J} \log^-|1-ax|e^{-x^2/2}\,dx + \int_{I\setminus J} \log^-|1-ax|e^{-x^2/2}\,dx}{\int_{I\cap J} e^{-x^2/2}\,dx + \int_{I\setminus J} e^{-x^2/2}\,dx}$$

The ratio of the two integrals over $I\setminus J$ is bounded below by $-2\log\Lambda$. Using (5), the ratio of the two integrals over $I \cap J$ is bounded below by $2(-2\log\Lambda - 1)\max_{I\cap J} e^{-x^2/2}/\min_{I\cap J} e^{-x^2/2} \ge 2e^{2/(a^2\Lambda^2)}(-2\log\Lambda - 1) \ge -2e^{18}(2\log\Lambda + 1)$. Since both ratios are bounded below by a constant multiple of $\log\Lambda$, so is the ratio of the sums.

If $a < 1/(3\Lambda)$, we argue similarly. In this case, we let $J = [\frac{1}{2a}, \frac{3}{2a}]$. On $I \setminus J$, $\log|1-ax|$ is bounded below by $-\log 2$, so that

$$\frac{\int_{I\setminus J} \log|1-ax|e^{-x^2/2}\,dx}{\int_\Lambda^\infty e^{-x^2/2}\,dx} \ge \frac{\int_{I\setminus J} \log|1-ax|e^{-x^2/2}\,dx}{\int_{I\setminus J} e^{-x^2/2}\,dx} \ge -\log 2.$$

On $I \cap J$, we have $e^{-x^2/2} \le e^{-1/(8a^2)}$. Also $\int_\Lambda^\infty e^{-x^2/2}\,dx \ge e^{-\Lambda^2/2}/(2\Lambda)$, using [7, Theorem 1.2.3]. Hence

$$\frac{\int_{I\cap J} \log^-|1-ax|e^{-x^2/2}\,dx}{\int_\Lambda^\infty e^{-x^2/2}\,dx} \ge \frac{2e^{-1/(8a^2)}(-\log 2 - 1)\frac{1}{a}}{e^{-\Lambda^2/2}/(2\Lambda)},$$

using (5). When $a = 1/(3\Lambda)$, this is $4(-\log 2 - 1)3\Lambda^2 e^{-5\Lambda^2/8}$ and the lower bound increases as a is further reduced. Minimizing this expression over Λ, we see that there is a C, independent of Λ, such that $\mathbb{E}\left(\log^-|1-aN|\middle|N \ge \Lambda\right) \ge -C$ for all $|a| < 1/(3\Lambda)$. □

Lemma 16. *Let $k > 0$ and Δ be as throughout the article. There exists $C > 0$ such that for all sufficiently small $\epsilon > 0$, for each a, b and each pair of k-dimensional orthogonal projections Π and Π',*

$$\mathbb{E}\left(\left(\Xi_k(\Pi(A+\epsilon\Delta)\Pi') - \Xi_k(\Pi A\Pi')\right)^-\middle|\mathsf{Wild}_{a,b}\right) > C(\log\epsilon - a - b),$$

where $\mathsf{Wild}_{a,b}$ is the event that Δ satisfies $|\Delta_{l,m}| < (\frac{2}{3})^{l+m}\epsilon^{-1/2}$ for each (l,m) that is lexicographically smaller than (a,b) and $|\Delta_{a,b}| \ge \epsilon^{-1/2}(\frac{2}{3})^{a+b}$ (where (l,m) is lexicographically smaller than (a,b) if $l < a$ or $l = a$ and $m < b$).

Proof. We deal with the case $\Delta_{a,b}$ positive. The case where it is negative is exactly analogous. Let $B_{a,b}$ be the collection of those Δ satisfying $\Delta_{a,b} \geq \epsilon^{-1/2}(\frac{2}{3})^{a+b}$ (and no other condition). The argument of Lemma 9 shows that $\mathbb{P}(\mathsf{Wild}_{a,b}|B_{a,b}) > \frac{1}{2}$. This allows us to deduce as in the proof of Lemma 13 that

$$\begin{aligned}&\mathbb{E}\left(\left(\Xi_k(\Pi(A+\epsilon\Delta)\Pi') - \Xi_k(\Pi A\Pi')\right)^-\Big|\Delta \in \mathsf{Wild}_{a,b}\right)\\ &> 2\,\mathbb{E}\left(\left(\Xi_k(\Pi(A+\epsilon\Delta)\Pi') - \Xi_k(\Pi A\Pi')\right)^-\Big|\Delta \in B_{a,b}\right)\end{aligned}$$

Hence it suffices to show that

$$\mathbb{E}\left(\left(\Xi_k(\Pi(A+\epsilon\Delta)\Pi') - \Xi_k(\Pi A\Pi')\right)^-\Big|\Delta \in B_{a,b}\right) > C(\log\epsilon - a - b).$$

Using the same reduction as in Lemma 13, the calculation reduces to showing that there is a C such that for sufficiently small $\epsilon > 0$, one has for an arbitrary $k \times k$ multi-variate normal matrix-valued random variable, R, whose entries have zero mean and for an arbitrary rank 1 $k \times k$ matrix Y,

$$\mathbb{E}_{N,R}\left(\Xi_k(I+R+\epsilon NY)^-\big|N > 2^{a+b}\epsilon^{-1/2}\right) \geq C(\log\epsilon - a - b),$$

where N is an independent standard normal random variable. First fixing N and taking the expectation over R using Lemma 13 (taking Q to be the full range of Δ), we obtain

$$\begin{aligned}&\mathbb{E}_{N,R}\left(\Xi_k(I+R+\epsilon NY)^-\big|N > 2^{a+b}\epsilon^{-1/2}\right)\\ &\geq \mathbb{E}_N\left(\Xi_k(I+\epsilon NY)^-\big|N > 2^{a+b}\epsilon^{-1/2}\right) - C.\end{aligned}$$

Hence it suffices to show

$$\mathbb{E}\left(\Xi_k(I+\epsilon NY)^-\big|N > 2^{a+b}\epsilon^{-1/2}\right) \geq C(\log\epsilon - a - b).$$

Since Y has rank 1, the polynomial $\det(I+tY)$ is of the form $1+at$. To see this, notice the determinant is unchanged if $I+tY$ is conjugated by an orthogonal matrix, O. Then choose O so that the first column spans the range of Y so that $O^{-1}(I+tY)O = I+t\tilde{Y}$, where $\tilde{Y}$ has only one non-zero row. $\det(I+tY)$ is then $1+t\tilde{Y}_{1,1}$. Hence we are seeking a lower bound for

$$\mathbb{E}(\log^-|1+cN|\big|N > 2^{a+b}\epsilon^{-1/2}),$$

which is of the desired form by Lemma 15. □

Proposition 17. *There exists a $C_{17} > 0$ with the following property. For any $m > 0$, let B be the event that at least one of the perturbations $\Delta_0, \ldots, \Delta_{m-1}$ is wild. Then*

$$\mathbb{E}\left(\tilde{\Xi}_k(A^{\epsilon(m)}_{\bar\omega})\Big|B\right) \geq \tilde{\Xi}_k(A^{(m)}_\omega) + C_{17}(\log\epsilon - m).$$

Proof. We write B as $B_0 \cup \ldots \cup B_{m-1}$, where B_i is the event that the ith perturbation matrix is wild, and all previous ones are tame. Since the B_i are disjoint, it suffices to establish that there is a $C > 0$ such that for each i,

$$\mathbb{E}(\tilde{\Xi}_k(A^{\epsilon(m)}_{\bar\omega})|B_i) \geq \tilde{\Xi}_k(A^{(m)}_\omega) + C(\log\epsilon - m). \tag{6}$$

We argue as in Proposition 14:

$$\begin{aligned}&\mathbb{E}\left(\tilde{\Xi}_k(A^{\epsilon(m)}_{\bar\omega}) - \tilde{\Xi}_k(A^{(m)}_\omega)\Big|B_i\right)\\ &= \sum_{j=0}^{m-1}\mathbb{E}\left(\tilde{\Xi}_k\big(A^{\epsilon(m-j)}_{\bar\sigma^j\bar\omega}A^{(j)}_\omega\big) - \tilde{\Xi}_k\big(A^{\epsilon(m-j-1)}_{\bar\sigma^{j+1}\bar\omega}A^{(j+1)}_\omega\big)\Big|B_i\right)\end{aligned}$$

As in Proposition 14, finding lower bounds for this reduces to finding lower bounds for $\mathbb{E}\left(\tilde{\Xi}_k(\Pi A^\epsilon_{\bar\sigma^j\bar\omega}\Pi') - \tilde{\Xi}_k(\Pi A_{\sigma^j\omega}\Pi')\Big|B_i\right)$.

In this case, for $j > i$, the conditional distribution of Δ_j is the same as the distribution used in Lemma 13 with $Q = \mathsf{HS}$, so that lemma gives a bound

$$\mathbb{E}\left(\tilde{\Xi}_k\big(A^{\epsilon(n-j)}_{\bar\sigma^j\bar\omega}A^{(j)}_\omega\big) - \tilde{\Xi}_k\big(A^{\epsilon(n-j-1)}_{\bar\sigma^{j+1}\bar\omega}A^{(j+1)}_\omega\big)\Big|B_i\right) \geq -C. \tag{7}$$

In the case $j < i$, Δ_j is conditioned to be tame. By Lemma 9, this is a set of probability (much) greater than $\frac{1}{2}$, so that Lemma 13 gives a similar bound to (7).

Finally, we address the term with $j = i$. Given that Δ_i is wild, the probability that the first oversized entry occurs in the (a, b) coordinate is $O(\exp(-\frac{1}{2}\epsilon^{-1}(2^{2a+2b} - 1)))$ (as seen from the estimate $\mathbb{P}(N > t) \approx (2\pi)^{-1/2}e^{-t^2/2}/t$ for large t [7, Theorem 1.2.3]).

Hence by conditioning and using Lemma 16, we obtain

$$\mathbb{E}\left(\tilde{\Xi}_k\big(A^{\epsilon(m-i)}_{\bar\sigma^i\bar\omega}A^{(i)}_\omega\big) - \tilde{\Xi}_k\big(A^{\epsilon(m-i-1)}_{\bar\sigma^{i+1}\bar\omega}A^{(i+1)}_\omega\big)\Big|B_i\right) > C(\log\epsilon - 1). \tag{8}$$

Combining equations (7) and the equation (8), we obtain the statement of the proposition. ∎

7. Joining good and bad blocks

Lemma 18. *For all $k \in \mathbb{N}$, there is a constant $C > 0$ such that for any $A \in \mathsf{HS}$, any orthogonal projections Π_1 and Π_2 onto k-dimensional subspaces, and any $Q \subset \mathsf{HS}$ such that $\mathbb{P}(\Delta \in Q) \geq \frac{1}{2}$, one has*

$$\mathbb{E}\,\Xi_k\big(\Pi_1(A+\Delta)\Pi_2\big|\Delta\in Q\big)\geq \mathbb{E}\,\Xi_k\big(\Pi_1\Delta\Pi_2\big|\Delta\in Q\big)-C.$$

Proof. Let $\tilde{\Pi}_1$ be an isometry from the range of Π_1 to $\mathbb{R}^k$. Similarly let $\tilde{\Pi}_2$ be the post-composition of Π_2 with an isometry from $\mathbb{R}^k$ to the span of the range of Π_2. Let $\tilde{A}=\tilde{\Pi}_1 A\tilde{\Pi}_2$ and let $\tilde{\Delta}=\tilde{\Pi}_1\Delta\tilde{\Pi}_2$ be the $k\times k$ multi-variate normal induced from the unconditioned distribution of Δ.

As in Lemma 13, we radially disintegrate the random variables $\tilde{\Delta}$, writing $\tilde{\Delta}$ as $t\tilde{M}$, where $\tilde{M}$ belongs to a 'unit sphere' equipped with a normalized probability measure and t having an absolutely continuous distribution on $[0,\infty)$ with density $r^{k^2-1}e^{-r^2/2}/\Gamma(k^2/2)$. On conditioning on $\Delta\in Q$, the density is bounded above by $2r^{k^2-1}e^{-r^2/2}/\Gamma(k^2/2)$ We prove that there is a $C>0$ such that for all $\tilde{M}$ of rank k,

$$\frac{2}{\Gamma(k^2/2)}\int_0^\infty\Big(\Xi_k(\tilde{A}+r\tilde{M})-\Xi_k(r\tilde{M})\Big)^- r^{k^2-1}e^{-r^2/2}\,dr>-C.$$

Notice that since the matrices are $k\times k$, Ξ_k is just the logarithm of the absolute value of the determinant. Let $p(r)=\det(\tilde{A}+r\tilde{M})/\det(r\tilde{M})$, a polynomial in powers of $1/r$ of degree at most k with constant coefficient 1. It can therefore be expressed as $p(r)=\prod_{i=1}^d(1-b_i/r)$, with $d\leq k$.

We are trying to bound

$$\int_0^\infty \log^-|p(r)|r^{k^2-1}e^{-r^2/2}\,dr\geq\sum_{i=1}^k\int_0^\infty\log^-|1-b_i/r|r^{k^2-1}e^{-r^2/2}\,dr.$$

As in the proof of Lemma 15, it suffices to give a bound in the case where $b>0$. We have

$$\int_0^\infty \log^-|1-b/r|r^{k^2-1}e^{-r^2/2}\,dr=\int_{b/2}^\infty\log^-|1-b/r|r^{k^2-1}e^{-r^2/2}\,dr.$$

The logarithm is bounded below by $-\log 2$ on $(2b,\infty)$, so that the contribution from this range is at least $-\Gamma(k^2/2)\log 2$. For the contribution from the range $[\frac{b}{2},2b]$, we have a lower bound of $-16(2b)^{k^2-2}e^{-b^2/8}$ (obtained by bounding $e^{-r^2/2}$ above by $e^{-b^2/8}$). Hence we obtain the required uniform lower bound. □

The following lemma plays a key role, as it provides an approximate super-additivity property for $\tilde{\Xi}_k$ (making strong use of the nature of the perturbations), complementing the well-known sub-additivity property of Ξ_k.

Lemma 19. *There exists $C > 0$ such that if Δ is distributed as above and Q is any subset of* HS *such that $\mathbb{P}(Q \in \Delta) \geq \frac{1}{2}$), then*

$$\mathbb{E}\left(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta \in Q\right) \geq \tilde{\Xi}_k(L) + \tilde{\Xi}_k(R) - k|\log \epsilon| - C.$$

Proof. We may assume that L and R have rank at least k as otherwise there is nothing to prove. Recalling the definition of $\tilde{\Xi}$, we have

$$\begin{aligned}
&\mathbb{E}\left(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta \in Q\right)\\
&= \mathbb{E}_{\Delta_1,\Delta_2}\,\mathbb{E}\left(\Xi_k(\Pi_k\Delta_1 L(A+\epsilon\Delta)R\Delta_2\Pi_k)\big|\Delta \in Q\right) \text{ and}\\
&\mathbb{E}\left(\tilde{\Xi}_k(L(\epsilon\Delta)R)\big|\Delta \in Q\right) = \mathbb{E}_{\Delta_1,\Delta_2}\,\mathbb{E}\left(\Xi_k(\Pi_k\Delta_1 L(\epsilon\Delta)R\Delta_2\Pi_k)\big|\Delta \in Q\right)
\end{aligned}$$

We first show that for fixed Δ_1 and Δ_2,

$$\begin{aligned}
&\mathbb{E}\left(\Xi_k(\Pi_k\Delta_1 L(A+\epsilon\Delta)R\Delta_2\Pi_k)\big|\Delta \in Q\right)\\
&\geq \mathbb{E}\left(\Xi_k(\Pi_k\Delta_1 L(\epsilon\Delta)R\Delta_2\Pi_k)\big|\Delta \in Q\right) - C.
\end{aligned} \tag{9}$$

We have $\Xi_k(\Pi_k\Delta_1 L(A+\epsilon\Delta)R\Delta_2\Pi_k) = \Xi_k(\Pi_k\Delta_1 L) + \Xi_k(\overline{\Pi}(A+\epsilon\Delta)\overline{\overline{\Pi}}) + \Xi_k(R\Delta_2\Pi_k)$ and $\Xi_k(\Pi_k\Delta_1 L(\epsilon\Delta)R\Delta_2\Pi_k) = \Xi_k(\Pi_k\Delta_1 L) + \Xi_k(\overline{\Pi}(\epsilon\Delta)\overline{\overline{\Pi}}) + \Xi_k(R\Delta_2\Pi_k)$, where $\overline{\Pi}$ is the orthogonal projection onto the k-dimensional orthogonal complement of the kernel of $\Pi_k\Delta_1 L$ and $\overline{\overline{\Pi}}$ is the orthogonal projection onto the range of $R\Delta_2\Pi_k$. Hence

$$\begin{aligned}
&\Xi_k(\Pi_k\Delta_1 L(A+\epsilon\Delta)R\Delta_2\Pi_k) - \Xi_k(\Pi_k\Delta_1 L(\epsilon\Delta)R\Delta_2\Pi_k)\\
&= \Xi_k(\overline{\Pi}(A+\epsilon\Delta)\overline{\overline{\Pi}}) - \Xi_k(\overline{\Pi}(\epsilon\Delta)\overline{\overline{\Pi}})\\
&= \Xi_k(\overline{\Pi}(\tfrac{1}{\epsilon}A+\Delta)\overline{\overline{\Pi}}) - \Xi_k(\overline{\Pi}\Delta\overline{\overline{\Pi}}).
\end{aligned}$$

Taking an expectation as Δ runs over Q and using Lemma 18, we obtain (9). Hence, taking the expectation over Δ_1 and Δ_2, we have

$$\begin{aligned}
\mathbb{E}\left(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta \in Q\right) &\geq \mathbb{E}\left(\tilde{\Xi}_k(L(\epsilon\Delta)R)\big|\Delta \in Q\right) - C\\
&= \mathbb{E}\left(\tilde{\Xi}_k(L\Delta R)\big|\Delta \in Q\right) - C + k\log\epsilon.
\end{aligned}$$

For the last part of the argument, we have

$$\begin{aligned}
&\mathbb{E}\left(\tilde{\Xi}_k(L\Delta R)\big|\Delta \in Q\right) = \mathbb{E}_{\Delta_1,\Delta_2}\,\mathbb{E}_\Delta\left(\Xi_k(\Pi_k\Delta_1 L\Delta R\Delta_2\Pi_k)\big|\Delta \in Q\right)\\
&= \mathbb{E}_{\Delta_1,\Delta_2}\left(\Xi_k(\Pi_k\Delta_1 L\overline{\Pi}) + \mathbb{E}_\Delta\left(\Xi_k(\overline{\Pi}\Delta\overline{\overline{\Pi}})\big|\Delta \in Q\right) + \Xi_k(\overline{\overline{\Pi}}R\Delta_2\Pi_k)\right),
\end{aligned}$$

where $\overline{\Pi}$ and $\overline{\overline{\Pi}}$ are as above. By Corollary 4, the middle term is $\mathbb{E}_{\Delta_3}\,\Xi_k(\overline{\Pi}\Delta_3\Pi_k) + \mathbb{E}_{\Delta_4}\,\Xi_k(\Pi_k\Delta_4\overline{\overline{\Pi}}) \pm C$. Substituting and recombining the expressions, we get

$$\begin{aligned}&\mathbb{E}\left(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta\in Q\right)\\&\geq \mathbb{E}_{\Delta_1,\Delta_3}\,\Xi_k(\Pi_k\Delta_1 L\Delta_3\Pi_k)+\mathbb{E}_{\Delta_2,\Delta_4}\,\Xi_k(\Pi_k\Delta_4 R\Delta_2\Pi_k)-C+k\log\epsilon\\&=\tilde{\Xi}_k(L)+\tilde{\Xi}_k(R)-C+k\log\epsilon,\end{aligned}$$

as required. □

Since the statement includes the case where Δ is conditioned to lie in a large set, this is sufficient to cover the case where Δ is conditioned to be tame. We need a version of this inequality to deal with the case where Δ is constrained to be wild.

Lemma 20. *There exists $C > 0$ such that for all polynomials, $p(x)$, one has*

$$\left|\int_{-\infty}^{\infty}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\log|p(x)|\,dx-\log M(p)\right|\leq C\deg(p),$$

where $M(p)$ is the Mahler measure of p: If $p(x) = a(x-z_1)(x-z_2)\cdots(x-z_k)$, then $M(p) = a\prod_{|z_i|>1}|z_i|$.

Proof. Write $p(x)$ as $a(x-z_1)\cdots(x-z_k)$. The inequality then follows from

$$\left|\int_{-\infty}^{\infty}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\log|x-z|\,dx-\log^+|z|\right|\leq C.$$

While we will not give all the details, the idea is to notice that the integral can be expressed as $\mathbb{E}\log|N-z|$ where N is a standard normal random variable. If z is small, then this is the integral of a function with a logarithmic singularity. If z is large, then since N is concentrated near 0, the integrand is close to $\log|z|$ with very high probability. □

Lemma 21. *For each $k > 0$, there exists a constant C such that for each polynomial $p(x) = \sum_{i=0}^k a_i x^i$, one has*

$$\big|\log M(p)-\max\log|a_i|\big|\leq C.$$

The proof can be found in Lang's book [20, Theorem 2.8].

Lemma 22. *Let $\Lambda > 2$ and let N be a standard normal random variable. There exists a $C > 0$ such that for all $a, b \in \mathbb{C}$,*

$$\mathbb{E}\left(\log|a+bN|\Big|N>\Lambda\right)\geq\max(\log|a|,\log|b|)-C\log\Lambda.$$

Proof. The case where $|a| > |b|$ follows from Lemma 15 (writing $\log|a+bN| = \log|a| + \log|1+\frac{b}{a}N|$). If $|b| \geq |a|$, then $|a+bN| \geq |b|\Lambda/2$ whenever $N > \Lambda$. The result follows. □

Lemma 23. *There exists a constant $C > 0$ such that for all i, j,*

$$\begin{aligned}&\mathbb{E}\,\tilde{\Xi}_k(L(A+\epsilon\Delta)R\big|\mathsf{Wild}_{i,j})\\ \geq\,&\tilde{\Xi}_k(L)+\tilde{\Xi}_k(R)-C|\log\epsilon|-C(i+j+1),\end{aligned}$$

where $\mathsf{Wild}_{i,j}$ is the event that $|\Delta_{i,j}| \geq (\frac{2}{3})^{i+j}\epsilon^{-1/2}$ and $|\Delta_{a,b}| < (\frac{2}{3})^{a+b}\epsilon^{-1/2}$ for all pairs (a, b) that are lexicographically smaller than (i, j).

Proof. As in the proof of Lemma 19, the proof reduces to showing a version of Lemma 18:

$$\mathbb{E}\,\Xi_k\big(\Pi_1(A+\epsilon\Delta)\Pi_2\big|\mathsf{Wild}_{i,j}\big) \geq \mathbb{E}\,\Xi_k(\Pi_1\epsilon\Delta\Pi_2) - C(i+j+1).$$

We first compare $\mathbb{E}\,\Xi_k\big(\Pi_1(A+\epsilon\Delta)\Pi_2\big|\mathsf{Wild}_{i,j}\big)$ to $\mathbb{E}\,\Xi_k\big(\Pi_1(A+\epsilon\Delta)\Pi_2\big|\mathsf{Tame}_{i,j}\big)$, where $\mathsf{Tame}_{i,j}$ is the event that $|\Delta_{a,b}| < (\frac{2}{3})^{a+b}\epsilon^{-1/2}$ for all pairs (a, b) that are lexicographically smaller than (i, j). Fixing all entries of Δ other than $\Delta_{i,j}$, this amounts to comparing $\mathbb{E}\,\big(\log|\det(B+NZ)|\big|N > 2^{i+j}\epsilon^{-1/2}\big)$ to $\mathbb{E}\,\big(\log|\det(B+NZ)|\big)$, where B is an invertible $k \times k$ matrix and Z is rank 1. As pointed out in Lemma 16, $\det(B + NZ) = a + bN$ for constants a and b, so that it suffices to compare $\mathbb{E}\,\big(\log|a+bN|\big|N > 2^{i+j}\epsilon^{-1/2}\big)$ to $\mathbb{E}\log|a+bN|$. By Lemma 22, the first of these is at least $\max(\log|a|, \log|b|) - C(i+j+\log\epsilon)$ and by Lemmas 20 and 21, the second of these is within C of $\max(\log|a|, \log|b|)$. We deduce that

$$\mathbb{E}\,\Xi_k\big(\Pi_1(A+\epsilon\Delta)\Pi_2\big|\mathsf{Wild}_{i,j}\big) > \mathbb{E}\,\Xi_k\big(\Pi_1(A+\epsilon\Delta)\Pi_2\big|\mathsf{Tame}_{i,j}\big) - C(i+j+\log\epsilon).$$

Hence, using the same cancellation argument that occurs in Lemma 19, we have

$$\mathbb{E}\,\tilde{\Xi}_k(L(A+\epsilon\Delta)R|\mathsf{Wild}_{i,j}) \geq \mathbb{E}\,\tilde{\Xi}_k(L(A+\epsilon\Delta)R|\mathsf{Tame}_{i,j}) - C(i+j+\log\epsilon).$$

Finally using Lemma 19 to bound $\mathbb{E}\,\tilde{\Xi}_k(L(A+\epsilon\Delta)R|\mathsf{Tame}_{i,j})$, the result follows. □

Proposition 24. *There exists $C_{24} > 0$ with the following property: Let L, R, and A be Hilbert-Schmidt operators and let Δ be the multivariate normal perturbation described earlier. Then each of $\mathbb{E}\,\tilde{\Xi}_k(L(A+\epsilon\Delta)R)$, $\mathbb{E}\,\big(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta$ is wild$\big)$ and $\mathbb{E}\,\big(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta$ is tame$\big)$ is bounded below by $\tilde{\Xi}_k(L) + \tilde{\Xi}_k(R) + C_{24}\log\epsilon$.*

Proof. The cases of $\mathbb{E}\,\tilde{\Xi}_k(L(A+\epsilon\Delta)R)$, $\mathbb{E}\,\big(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta$ is tame$\big)$ are handled by Lemma 19. The case of $\mathbb{E}\,\big(\tilde{\Xi}_k(L(A+\epsilon\Delta)R)\big|\Delta$ is wild$\big)$ is handled using Lemma 23 by conditioning on the first entry of Δ that is large analogously to the end of the proof of Proposition 17. □

8. Comparison of Ξ_k and $\tilde{\Xi}_k$

Lemma 25. *Let C_k be the expected value of $\log|\det N_k|$ where N_k is a $k \times k$ matrix-valued random variable with independent standard normal entries. Let $n \geq k$, let A be an $n \times n$ matrix and let N be a $k \times n$ matrix-valued random variable with independent standard normal entries. Then $\mathbb{E}\,\Xi_k(NA) \geq \Xi_k(A) + C_k$.*

Proof. Write $A = UDV$ where U and V are orthogonal and D is diagonal with decreasing entries. Then by an argument like that in Lemma 3 (computing covariances between elements) NU has the same distribution as N, so that we have $\mathbb{E}\,\Xi_k(NA) = \mathbb{E}\,\Xi_k(NUDV) = \mathbb{E}\,\Xi_k(ND) \geq \mathbb{E}\,\Xi_k(ND\Pi_k)$. Notice that since D is diagonal, $ND\Pi_k$ has the form $\big(N_kD_k|0\big)$, where N_k is the left $k \times k$ submatrix of N and D_k is the top left $k \times k$ submatrix of D. Hence $\mathbb{E}\,\Xi_k(ND\Pi_k) = \mathbb{E}\,\Xi_k(N_kD_k) = C_k + \Xi_k(D_k) = C_k + \Xi_k(A)$ as required. □

Lemma 26. *Let A, B and C be Hilbert-Schmidt matrices, and let $A_n = \Pi_n A \Pi_n$. Then $\Xi_k(BA_nC) \to \Xi_k(BAC)$ as $n \to \infty$.*

Proof. Let $R_n = A - A_n$, so that $\|R_n\| \to 0$. We have $|s_i(BA_nC) - s_i(BAC)| \leq \|B\| \cdot \|R_n\| \cdot \|C\|$ for each i so that $s_i(BA_nC) \to s_i(BAC)$ for each i. The conclusion follows. □

Proposition 27. *Let $k > 0$. Then there exists a constant C_{27} such that for an arbitrary Hilbert-Schmidt operator A on H,*

$$\tilde{\Xi}_k(A) \geq \Xi_k(\mathcal{D}_3 A \mathcal{D}_3) - C_{27}.$$

Proof. We have $\tilde{\Xi}_k(A) = \mathbb{E}_{\Delta,\Delta'}\,\Xi_k(\Pi_k \Delta A \Delta' \Pi_k)$ where Δ and Δ' are independent copies of the perturbation operator. Since $\Xi_k(\Pi_k\Delta A_n \Delta'\Pi_k) \leq k \log \|\Pi_k \Delta A_n \Delta' \Pi_k\|_{\mathsf{op}} \leq k\log(\|\Delta\|_{\mathsf{op}} \cdot \|A_n\|_{\mathsf{op}} \cdot \|\Delta'\|_{\mathsf{op}})$; $\|A_n\|_{\mathsf{op}} \leq \|A\|_{\mathsf{HS}}$ and $\mathbb{E}\log\|\Delta\|_{\mathsf{op}} < \mathbb{E}\,\|\Delta\|_{\mathsf{op}} \leq \mathbb{E}\,\|\Delta\|_{\mathsf{HS}} < \infty$, we see that the family of functions, $(\Delta, \Delta') \mapsto \Xi_k(\Pi_k \Delta A_n \Delta' \Pi_k)$ is dominated by an integrable function. Hence, by the Reverse Fatou Lemma and Lemma 26, we have

$$\limsup_{n\to\infty} \tilde{\Xi}_k(A_n) = \limsup_{n\to\infty} \mathbb{E}\,\Xi_k(\Pi_k \Delta A_n \Delta' \Pi_k) \leq \tilde{\Xi}_k(A).$$

However, we have

$$\mathbb{E}_{\Delta,\Delta'}\,\Xi_k(\Pi_k \Delta A_n \Delta' \Pi_k) = \mathbb{E}_{\Delta,\Delta'}\,\Xi_k(\Delta_{k\times n} A_n \Delta'_{n\times k}),$$

where $\Delta_{k\times n}$ denotes the random Hilbert Schmidt operator Δ with all entries outside the top left $k \times n$ corner replaced by 0's (and $\Delta'_{n\times k}$ similarly). Hence

$$\begin{aligned}
\tilde{\Xi}_k(A_n) &= \mathbb{E}_{N,N'}\,\Xi_k\big((\mathcal{D}_3)_{k\times k} N_{k\times n} (\mathcal{D}_3)_{n\times n} A_n (\mathcal{D}_3)_{n\times n} N'_{n\times k} (\mathcal{D}_3)_{k\times k}\big) \\
&= \mathbb{E}_{N,N'}\,\Xi_k\big(N_{k\times n} (\mathcal{D}_3)_{n\times n} A_n (\mathcal{D}_3)_{n\times n} N'_{n\times k}\big) - k(k-1)\log 3
\end{aligned}$$

Applying Lemma 25 twice, we deduce $\tilde{\Xi}_k(A_n) \geq \Xi_k(\mathcal{D}_3 A_n \mathcal{D}_3) + C$, so that on taking the limit, we deduce $\tilde{\Xi}_k(A) \geq \Xi_k(\mathcal{D}_3 A \mathcal{D}_3) + C$ as required. □

Corollary 28. *There is a C_{28} with the following property. Let L, R, A and A' be Hilbert-Schmidt operators and Δ and Δ' be independent copies of the standard perturbation. Then we have*

$$\mathbb{E}\,\tilde{\Xi}_k(L(A'+\epsilon\Delta')(A+\epsilon\Delta)R) \geq \tilde{\Xi}_k(L) + \tilde{\Xi}_k(R) + C_{28}\log\epsilon.$$

The same inequality holds if either or both of Δ and Δ' are constrained to be either tame or wild (or one of each).

Proof. Let $L' = L(A'+\epsilon\Delta') = L(A'+\epsilon\Delta')I$. By Proposition 24, $\mathbb{E}_{\Delta'}\,\tilde{\Xi}_k(L') \geq \tilde{\Xi}_k(L) + \tilde{\Xi}_k(I) + C_{24}\log\epsilon$, with this inequality still satisfied if Δ' is constrained to be tame or wild. By Proposition 27, $\tilde{\Xi}_k(I)$ is a finite constant. Finally, $\mathbb{E}_\Delta\,\tilde{\Xi}_k(L'(A+\epsilon\Delta)R) \geq \tilde{\Xi}_k(L') + \tilde{\Xi}_k(R) + C_{24}\log\epsilon$. Combining the inequalities, the result is proved. □

Lemma 29. *Let $f(t) = \sum_{i=1}^n a_i e^{b_i t}$ where $a_i > 0$ for each i. Then $f(t)$ is log-convex.*

Proof. We have $(\log f)' = f'/f$, so that $(\log f)'' = (ff'' - (f')^2)/f^2$. Now

$$\begin{aligned} ff'' - (f')^2 &= \sum_{i\neq j} a_i a_j e^{(b_i+b_j)t}(b_j^2 - b_i b_j) + \sum_i a_i^2 e^{2b_i t}(b_i^2 - b_i^2) \\ &= \sum_{i<j} a_i a_j e^{(b_i+b_j)t}(b_i^2 + b_j^2 - 2b_i b_j) \\ &\geq 0. \end{aligned}$$

□

Lemma 30. *Let V be a k-dimensional subspace of H and let Π_V be the orthogonal projection onto V. Then $f(s) := \Xi_k(\mathcal{D}_{e^s} \circ \Pi_V)$ is a convex function.*

Proof. We first prove that for $0 < s < t$, $f(s) \leq \frac{s}{t} f(t)$. To see this, let $v_1, \dots, v_k$ be an orthogonal basis of V such that $\mathcal{D}_{e^t} v_1, \dots, \mathcal{D}_{e^t} v_k$ are orthogonal. Then $f(s) \leq \sum_{i=1}^k \log\|\mathcal{D}_{e^s} v_i\|$. By Lemma 29, $s \mapsto \log\|\mathcal{D}_{e^s} v_i\| = \frac{1}{2}\log(\sum_j e^{-2sj}{(v_i)_j}^2)$ is convex, so that $\log\|\mathcal{D}_{e^s} v_i\| \leq \frac{s}{t}\log\|\mathcal{D}_{e^t} v_i\|$. Hence $f(s) \leq \frac{s}{t} f(t)$ as claimed.

Now if $0 < a < b < c$, let $W = \mathcal{D}_{e^a} V$, let $s = b - a$ and $t = c - a$. Let $\alpha = \Xi_k(\mathcal{D}_{e^a}\Pi_V)$. Now we have $f(a) = \alpha$, $f(b) = \Xi_k(\mathcal{D}_{e^b}\Pi_V) = \Xi_k(\mathcal{D}_{e^{b-a}}\mathcal{D}_{e^a}\Pi_V) = \Xi_k(\mathcal{D}_{e^{b-a}}\Pi_W) + \Xi_k(\mathcal{D}_{e^a}\Pi_V) = \alpha + \Xi_k(\mathcal{D}_{e^s}\Pi_W)$. Similarly $f(c) = \alpha + \Xi_k(\mathcal{D}_{e^t}\Pi_W)$ and the result follows from the above. □

Lemma 31. *Let A be a Hilbert-Schmidt operator on H. Then $g(s)\colon = \Xi_k(\mathcal{D}_{e^s}A)$ is a convex function. Similarly $h(s)\colon = \Xi_k(A\mathcal{D}_{e^s})$ is convex.*

Proof. Let $0 < a < b < c$. Let V be the k-dimensional space spanned by the top k right singular vectors of $\mathcal{D}_{e^b}A$ and Π_V be the orthogonal projection onto V. Let $W = A(V)$ and Π_W be the orthogonal projection onto W. Then we have $\Xi_k(\mathcal{D}_{e^t}A\Pi_V) = \Xi_k(\mathcal{D}_{e^t}\Pi_W) + \Xi_k(A\Pi_V)$, the sum of a convex function and a constant by Lemma 30. Now $g(b) = \Xi_k(\mathcal{D}_{e^b}A) = \Xi_k(\mathcal{D}_{e^b}A\Pi_V) \leq \frac{c-b}{c-a}\Xi_k(\mathcal{D}_{e^a}A\Pi_V) + \frac{b-a}{c-a}\Xi_k(\mathcal{D}_{e^c}A\Pi_V) \leq \frac{c-b}{c-a}g(a) + \frac{b-a}{c-a}g(c)$ as required.

We have $h(s) = \Xi_k(A\mathcal{D}_{e^s}) = \Xi_k(\mathcal{D}_{e^s}A^*)$, which is convex by the above. □

Proposition 32. *Let A be a Hilbert-Schmidt operator on H. Then*

$$\Xi_k(\mathcal{D}_3A\mathcal{D}_3) - \Xi_k(A) \geq \left(\frac{\log 3}{\log 2}\right)^2 \big(\Xi_k(\mathcal{D}_2A\mathcal{D}_2) - \Xi_k(A)\big).$$

Proof. Let $f(s,t) = \Xi_k(\mathcal{D}_{e^s}A\mathcal{D}_{e^t}) - \Xi_k(A)$. Since $\mathcal{D}_a$ is contractive for $a > 1$, we have $f(\log 3, 0) \leq 0$ and $f(0, \log 2) \leq 0$. Now Lemma 31 applied to $\Xi_k(\mathcal{D}_3A\mathcal{D}_{e^t}) - \Xi_k(A)$ implies that $f(\log 3, \log 2) \leq \frac{\log 2}{\log 3}f(\log 3, \log 3)$. Applying the lemma to $\Xi_k(\mathcal{D}_{e^s}A\mathcal{D}_2) - \Xi_k(A)$ implies

$$f(\log 2, \log 2) \leq \frac{\log 2}{\log 3}f(\log 3, \log 2) \leq \left(\frac{\log 2}{\log 3}\right)^2 f(\log 3, \log 3),$$

as required. □

Lemma 33. *Let σ be an ergodic measure-preserving transformation of $(\Sigma, \mathbb{P})$. Let (f_n) be a sub-additive sequence of functions (that is $f_{n+m}(\omega) \leq f_n(\sigma^m\omega) + f_m(\omega)$ for each $\omega \in \Omega$ and $n, m > 0$) such that $\inf_{n>0}\int \frac{1}{n}f_n\,d\mathbb{P} > -\infty$. For any $\epsilon > 0$, there exist $\chi > 0$ and n_0 such that if $M \geq n_0$ and A is any set with $\mathbb{P}(A) < \chi$ then $\int_A f_M\,d\mathbb{P} > -\epsilon M$.*

Proof. Let $\alpha = \lim \int (f_n/n)\,d\mathbb{P}$. Let $\epsilon > 0$ be given. Let χ be small enough that $\int_B f_1\,d\mathbb{P} < \frac{\epsilon}{3}$ for any set B with $\mathbb{P}(B) \leq \chi$ and so that $2\chi(\alpha + \frac{\epsilon}{3}) > -\frac{\epsilon}{3}$. By the Kingman sub-additive ergodic theorem, there exists m_0 such that for $M \geq m_0$, $\mathbb{P}(\{\omega\colon f_M(\omega) > (\alpha + \frac{\epsilon}{3})M\}) < \chi$.

Now let A be an arbitrary set with $\mathbb{P}(A) < \chi$. We split Ω into three sets: A, $G = \{\omega \in A^c\colon f_M(\omega) \leq (\alpha + \frac{\epsilon}{3})M\}$ and $B = A^c \setminus G$ (and note that $\mathbb{P}(G^c) \le 2\chi$). Now we have

$$\begin{aligned}\alpha M &\le \int_\Omega f_M \, d\mathbb{P} \\ &= \int_A f_M \, d\mathbb{P} + \int_B f_M \, d\mathbb{P} + \int_G f_M \, d\mathbb{P} \\ &\le \int_A f_M \, d\mathbb{P} + \int_B (f_1 + \ldots + f_1 \circ \sigma^{M-1}) \, d\mathbb{P} + (\alpha + \tfrac{\epsilon}{3}) M \mathbb{P}(G).\end{aligned}$$

Hence we see

$$\begin{aligned}\int_A f_M \, d\mathbb{P} &\ge \alpha M - M\tfrac{\epsilon}{3} - (\alpha + \tfrac{\epsilon}{3}) M (1 - \mathbb{P}(G^c)) \\ &= -\tfrac{2\epsilon}{3} M + (\alpha + \tfrac{\epsilon}{3}) M \mathbb{P}(G^c) \ge -\epsilon M,\end{aligned}$$

as required. □

Lemma 34. *For all k, there exists a C_{34} such that for any bounded operator A one has*

$$\Xi_k(A) \ge \tilde{\Xi}_k(A) - C_{34}.$$

Proof. We have $\tilde{\Xi}_k(A) = \mathbb{E}_{\Delta_1,\Delta_2} \Xi_k(\Pi_k \Delta_1 A \Delta_2 \Pi_k) \le 2\,\mathbb{E}\,\Xi_k(\Delta) + \Xi_k(A) \le 2k\,\mathbb{E}\log\|\Delta\|_{\mathsf{op}} + \Xi_k(A)$, where we used sub-additivity of Ξ_k for the first inequality and the fact that $s_i(B) \le \|B\|_{\mathsf{op}}$ for the second. Hence it suffices to show that $\mathbb{E}\log\|\Delta\|_{\mathsf{op}} < \infty$. But $\mathbb{E}\log\|\Delta\|_{\mathsf{op}} \le \mathbb{E}\,\|\Delta\|_{\mathsf{op}} \le \mathbb{E}\,\|\Delta\|_{\mathsf{HS}} \le \sum_{i,j} \mathbb{E}\,|\Delta_{ij}| = \sum_{i,j} 3^{-(i+j)}\,\mathbb{E}\,|N| < \infty$. □

9. Convergence of the Lyapunov exponents

Proof of Theorem A. Rather than control the exponents directly, it is more straightforward, and clearly equivalent, to control the partial sums of the exponents. Let $\mu_1(A) \ge \mu_2(A) \ge \ldots$ denote the Lyapunov exponents of the cocycle A listed with multiplicity in decreasing order. We then let $\Lambda_k(A) = \mu_1(A) + \ldots + \mu_k(A)$. We are aiming to show that $\Lambda_k(A^\epsilon) \to \Lambda_k(A)$ for each k. By an argument of Ledrappier and Young [21], explained slightly differently in our earlier paper [10], it suffices to show that $\epsilon \mapsto \Lambda_k(A^\epsilon)$ is upper semi-continuous for each k; and lower semi-continuous for those k such that $\mu_{k+1}(A) < \mu_k(A)$.

9.1. Upper semi-continuity.

We shall show $\limsup_{\epsilon\to 0} \Lambda_k(A^\epsilon) \le \Lambda_k(A)$. To see this, let $\eta > 0$. By the sub-additive ergodic theorem, there exists an n such that $\frac{1}{n}\int \Xi_k(A^{(n)}_\omega)\, d\mathbb{P}(\omega) < \Lambda_k(A) + \eta$. As $\epsilon \to 0$, we have $\|A^{\epsilon(n)}_{\bar\omega} - A^{(n)}_\omega\| \to 0$ and hence $\Xi_j(A^{\epsilon(n)}_{\bar\omega}) \to \Xi_j(A^{(n)}_\omega)$ for all $\bar\omega \in \bar\Omega$. Set $g(\bar\omega) = 1 + \|A(\omega_0)\|$ and $h(\bar\omega) = \|\Delta_0\|$. Then for $\epsilon < 1$, $\log\|A^{\epsilon(n)}_{\bar\omega}\| \le \sum_{i=0}^{n-1} \log(g+h)(\bar\sigma^i\bar\omega)$. Since this is integrable, the Reverse

Fatou Lemma implies that $\limsup_{\epsilon\to 0}\frac{1}{n}\int \Xi_j(A^{\epsilon}{}^{(n)}_{\bar\omega})\,d\bar{\mathbb{P}}(\bar\omega) < \Lambda_k(A)+\eta$. Hence $\Lambda_k(A^\epsilon) < \Lambda_k(A)+\eta$ for sufficiently small ϵ.

9.2. **Choice of Parameters.** Now we move to showing the lower semi-continuity of $\Lambda_k(A^\epsilon)$ in the case where $\mu_{k+1}(A) < \mu_k(A)$. We assume without loss of generality (by scaling the entire cocycle by a constant if necessary) that $\mu_{k+1}(A) < 0 < \mu_k(A)$.

Let $\eta > 0$. We are seeking an ϵ_0 such that for $\epsilon < \epsilon_0$, $\Lambda_k(A^\epsilon) > \Lambda_k(A)-\eta$. First, choose an n_0 and χ such that the following inequalities are satisfied:

$$\chi < \min\left(\frac{C_9\eta}{48\max(C_{24},C_{28})}, \frac{\eta}{18\max(C_{14},C_{17}(1+\frac{2}{C_9}))}\right);$$

$$\chi < \frac{\eta}{72k\int\log(1+\|A_\omega\|_{\mathsf{SHS}})\,d\mathbb{P}(\omega)};$$

$$\int_B \Xi_k(A_\omega^{(N)})\,d\mathbb{P}(\omega) > -\frac{\eta N}{72} \text{ for } N\ge n_0 \text{ if } \mathbb{P}(B)<\chi;$$

$$\int_B \log^+\|A_\omega\|_{\mathsf{SHS}}\,d\mathbb{P}(\omega) < \frac{\eta}{108k} \text{ if } \mathbb{P}(B)<\chi.$$

That n_0 and χ can be chosen to satisfy the third inequality is a consequence of Lemma 33. Let δ be chosen so that $\mathbb{P}(G^c) < \chi/2$, where G is the event that the block $A_\omega^{(N)}$ is good as in Lemma 8. Let ϵ_1 be chosen so that $N_\epsilon := \lfloor C_9|\log\epsilon|\rfloor > n_0$ for all $\epsilon<\epsilon_1$. Let ϵ_2 be such that the probability that an N_ϵ-block of Δ's contains a wild perturbation is less than $\chi/2$ for all $\epsilon<\epsilon_2$ (such an ϵ_2 exists by Lemma 9). Let $\bar G = \{\bar\omega\in\bar\Omega : \omega\in G; \Delta_0,\ldots,\Delta_{N-1} \text{ are tame}\}$. We will only consider ϵ's that are smaller than ϵ_1 and ϵ_2 for the remainder of the argument. In particular $\bar{\mathbb{P}}(\bar G^c) < \chi$.

We need to control $\mathbb{E}\,\Xi_k(A_{\bar\omega}^{(nN)})$, where N is the length of a block (as given by Lemma 9), and we let $n\to\infty$. Here and below, the superscript ϵ indicates that we are studying the perturbed cocycle.

9.3. **Replacing Ξ_k with $\tilde\Xi_k$.** We have

$$\Xi_k(A^{\epsilon}{}^{(nN)}_{\bar\omega}) \ge \tilde\Xi_k(A^{\epsilon}{}^{(nN)}_{\bar\omega}) - C_{34}, \tag{10}$$

by Lemma 34. The advantage of $\tilde\Xi_k$ over Ξ_k is that it admits a lower bound in terms of sub-blocks.

9.4. **Splitting $A^{\epsilon}{}^{(nN)}_{\bar\omega}$ into good and bad blocks.** Recall a block $A^{\epsilon}{}^{(N)}_{\sigma^{jN}\bar\omega}$ is said to be good if $\sigma^{jN}\bar\omega\in\bar G$, that is the unperturbed cocycle is well-behaved, and the perturbations are tame. Given $\bar\omega$, we split up $A^{\epsilon}{}^{(nN)}_{\bar\omega}$ into blocks of length N. Whenever three or more consecutive blocks are good, we form a *super-block*, G^ϵ, consisting of the concatenation of the good blocks other than the first and last good blocks. All of the remaining blocks are called *filler* blocks. The B^ϵ are the filler blocks stripped of their first and last matrices.

We have

$$\begin{aligned}(11)\qquad &\mathbb{E}\left(\tilde{\Xi}_k(A^{\epsilon(nN)}_{\omega})\right) \geq \\ &\mathbb{E}\left(\tilde{\Xi}_k(B^\epsilon) + \tilde{\Xi}_k(G^\epsilon) + \tilde{\Xi}_k(B^\epsilon) + \tilde{\Xi}_k(B^\epsilon) + \tilde{\Xi}_k(B^\epsilon) + \dots\right) - E_1,\end{aligned}$$

where the splitting in the last line is into super-blocks (of variable length, all a multiple of N), here designated by G^ϵ, and filler blocks, B^ϵ, all of length $N-2$ and E_1 denotes an expected error term that we now estimate.

To obtain (11), we split the concatenation of n blocks of length N into the super-blocks and filler blocks as described above by repeatedly applying Proposition 24, which sacrifices a single matrix as 'glue' at each splitting site (or Corollary 28 in the case of two consecutive filler blocks when two matrices are sacrificed). Since the expected number of non-good N-blocks is less than χn and each such block gives rise to at most 4 transitions between adjacent blocks in the concatenation (the worst case happens when two super-blocks are joined by three fillers), we deduce $E_1 \leq 4\chi n \max(C_{24}, C_{28})|\log\epsilon|$. From Lemma 9, $|\log\epsilon| \leq 2N/C_9$, so that

$$(12)\qquad E_1 \leq 8\chi nN \max(C_{24}, C_{28})/C_9 \leq \tfrac{1}{6}\eta nN.$$

9.5. **Comparison of $\tilde{\Xi}_k(G^\epsilon)$ and $\Xi_k(G^\epsilon)$.** To bound one of the $\tilde{\Xi}_k(G^\epsilon)$, the contribution from one of the super-blocks, we first compare to $\Xi_k(G^\epsilon)$, the corresponding contribution to the genuine singular values; and then compare to $\Xi_k(G^0)$, the singular values of the unperturbed block. Recall that each G^ϵ is preceded by an N-block L^ϵ and followed by an N-block R^ϵ such that the enlarged block $L^\epsilon G^\epsilon R^\epsilon$ consists entirely of good blocks.

For the first comparison, we have

$$\begin{aligned}(13)\qquad \tilde{\Xi}_k(G^\epsilon) &\geq \Xi_k(\mathcal{D}_3 G^\epsilon \mathcal{D}_3) - C_{27} \\ &\geq \Xi_k(G^\epsilon) + 3(\Xi_k(\mathcal{D}_2 G^\epsilon \mathcal{D}_2) - \Xi_k(G^\epsilon)) - C_{27},\end{aligned}$$

using Propositions 27 and 32 respectively. Now

$$(14)\qquad \Xi_k(\mathcal{D}_2 G^\epsilon \mathcal{D}_2) \geq \Xi_k(L^\epsilon G^\epsilon R^\epsilon) - \Xi_k(L^\epsilon \mathcal{D}_2^{-1}) - \Xi_k(\mathcal{D}_2^{-1} R^\epsilon)$$

by sub-additivity, and

$$
\begin{aligned}
&\Xi_k(L^\epsilon G^\epsilon R^\epsilon) \geq \log|\det(L^\epsilon G^\epsilon R^\epsilon|_{F^\perp(R_0)})| \\
&\quad = \log|\det(L^\epsilon|_{G^\epsilon R^\epsilon(F^\perp(R_0))})| + \log|\det(G^\epsilon|_{R^\epsilon(F^\perp(R_0))})| \\
&\qquad + \log|\det(R^\epsilon|_{F^\perp(R_0)})| \\
&\quad \geq \Xi_k(L^\epsilon) + \Xi_k(G^\epsilon) + \Xi_k(R^\epsilon) + 3k\log\delta,
\end{aligned} \tag{15}
$$

where we made use of Proposition 12 for the second inequality (Lemmas 7(c) and 8(a), (b) and (c) were used to ensure the hypotheses of that Proposition are satisfied). Combining inequalities (13), (14) and (15), we obtain

$$
\begin{aligned}
\tilde{\Xi}_k(G^\epsilon) \geq \Xi_k(G^\epsilon) + 3\Big(&\Xi_k(L^\epsilon) + \Xi_k(R^\epsilon) - \Xi_k(L^\epsilon \mathcal{D}_2^{-1}) \\
&- \Xi_k(\mathcal{D}_2^{-1} R^\epsilon) + 3k\log\delta\Big) - C_{27}.
\end{aligned}
$$

By Lemmas 5(c), 8(d) and 9, we have $\Xi_k(L^\epsilon)$ and $\Xi_k(R^\epsilon)$ are non-negative. By Lemma 8(e), using sub-additivity, we have $\Xi_k(L^\epsilon \mathcal{D}_2^{-1}), \Xi_k(\mathcal{D}_2^{-1} R^\epsilon) \leq 2kN \int \log(1 + \|A_\omega\|_{\mathsf{SHS}})\, d\mathbb{P}(\omega)$. Hence for each good block, we have

$$
\tilde{\Xi}_k(G^\epsilon) \geq \Xi_k(G^\epsilon) - \eta N/(6\chi) + 9k\log\delta - C_{27}.
$$

9.6. **Comparison of $\Xi_k(G^\epsilon)$ and $\Xi_k(G^0)$.** Next, by Proposition 10, we have $\Xi_k(G^\epsilon) \geq \Xi_k(G^0) + 2k\ell\log\delta$, where ℓ is the number of blocks forming the G^ϵ super-block, so that overall, for each good block, we have

$$
\tilde{\Xi}_k(G^\epsilon) \geq \Xi_k(G^0) - \eta N/(6\chi) + 11k\ell\log\delta - C_{27}, \tag{16}
$$

where G^0 is the corresponding unperturbed block.

In summary,

$$
\begin{aligned}
&\mathbb{E}\left(\tilde{\Xi}_k(A_\omega^{\epsilon(nN)})\right) \geq \\
&\mathbb{E}\left(\tilde{\Xi}_k(B^\epsilon) + \Xi_k(G^0) + \tilde{\Xi}_k(B^\epsilon) + \tilde{\Xi}_k(B^\epsilon) + \tilde{\Xi}_k(B^\epsilon) + \dots\right) - E_1 - E_2,
\end{aligned} \tag{17}
$$

where E_2 is the combined contribution of the errors coming from good blocks via (16).

9.7. **Comparison of $\mathbb{E}\,\tilde{\Xi}_k(B^\epsilon)$ and $\tilde{\Xi}_k(B^0)$.** We next work on giving a lower bound for the terms of the form $\mathbb{E}\,\tilde{\Xi}_k(B^\epsilon)$. It turns out to be convenient to bound this in the opposite order than the way we obtained bounds for $\mathbb{E}\,\tilde{\Xi}_k(G^\epsilon)$. Namely, we show $\mathbb{E}\,\tilde{\Xi}_k(B^\epsilon) \gtrsim \tilde{\Xi}_k(B^0) \gtrsim \Xi_k(B^0)$.

If the filler block $B^\epsilon = A^{\epsilon(N-2)}_{\bar\sigma^{jN+1}\bar\omega}$ is not type II bad, we have $\mathbb{E}\,\tilde\Xi_k(B^\epsilon) \geq \tilde\Xi_k(B^0) - C_{14}N$ by Proposition 14, where $B^0 = A^{(N-2)}_{\sigma^{jN+1}\omega}$, the unperturbed block. When B^ϵ is type II bad, we have $\mathbb{E}\,\tilde\Xi_k(B^\epsilon) \geq \tilde\Xi_k(B^0) + C_{17}(\log\epsilon - N)$ by Proposition 17. Since by Lemma 9, we have $\log\epsilon > -2N/C_9$, we get $\mathbb{E}\,\tilde\Xi_k(B^\epsilon) \geq \tilde\Xi_k(B^0) - C_{17}N(1+2/C_9)$ in this case. We therefore have in either case that

$$\mathbb{E}\,\tilde\Xi_k(B^\epsilon) \geq \tilde\Xi_k(B^0) - \eta/(18\chi)N, \tag{18}$$

9.8. **Comparison of $\tilde\Xi_k(B^0)$ and $\Xi_k(B^0)$.** For the estimate $\tilde\Xi_k(B^0) \gtrsim \Xi_k(B^0)$, we use an argument similar to that in (13) and (14) above. Namely, let the matrices preceding and following B^0 in the unperturbed cocycle be L^0 and R^0. We also write $\bar B^0 = A^{(N)}_{\sigma^{jN}\omega}$ for the N-block, $L^0B^0R^0$. Then as before, we have

(19)
$$\begin{aligned}
&\tilde\Xi_k(B^0) \geq \Xi(B^0) + 3(\Xi_k(\mathcal{D}_2B^0\mathcal{D}_2) - \Xi_k(B^0)) - C_{27}\\
&\geq \Xi_k(B^0) + 3\big(\Xi_k(\bar B_0) - \Xi_k(L^0\mathcal{D}_2^{-1}) - \Xi_k(\mathcal{D}_2^{-1}R^0) - \Xi_k(B^0)\big) - C_{27}\\
&= \Xi_k(\bar B^0) + 2\big(\Xi_k(\bar B^0) - \Xi_k(B^0)\big) - 3\big(\Xi_k(L^0\mathcal{D}_2^{-1}) + \Xi_k(\mathcal{D}_2^{-1}R^0)\big) - C_{27}.
\end{aligned}$$

We have the estimate for the subtracted terms in (19):

$$2\Xi_k(B^0) + 3(\Xi_k(L^0\mathcal{D}_2^{-1}) + \Xi_k(\mathcal{D}_2^{-1}R^0)) \leq 3kF(\sigma^{jN}\omega),$$

where $F(\omega) = \sum_{i=0}^{N-1}\log^+\|A(\sigma^i\omega)\|_{\mathsf{SHS}}$. This is a consequence of subadditivity of Ξ_k, the fact that $\|A\mathcal{D}_2^{-1}\|_{\mathsf{op}}, \|A\|_{\mathsf{op}} \leq \|A\|_{\mathsf{SHS}}$ for every $A \in \mathsf{SHS}$ and $\Xi_k(A) \leq k\log\|A\|_{\mathsf{op}}$. By the choice of χ, we have $\int_{\bar G^c} F(\omega)\,d\bar{\mathbb{P}}(\bar\omega) < \eta N/(108k)$. The combined contribution from the subtracted terms in (19) to all of the $\tilde\Xi_k(B^\epsilon)$ terms in (11) is bounded above by

$$3k\sum_{j=0}^{n-1}\mathbf{1}_{\mathsf{Filler}}(\bar\sigma^{jN}\bar\omega)F(\sigma^{jN}\omega),$$

where Filler is $\bar G^c \cup \bar\sigma^{-N}\bar G^c \cup \bar\sigma^N\bar G^c$, the set of points which are the first index of a filler block. Hence the expectation of the contribution of the subtracted terms in (19) is at most $\eta nN/12$.

We use a similar argument to give a lower bound for the sum of the added $2\Xi_k(\bar B_0)$ terms in (19). These terms are

$$2\sum_{j=0}^{n-1}\mathbf{1}_{\mathsf{Filler}}(\bar\sigma^{jN}\bar\omega)\Xi_k(A^{(N)}_{\sigma^{jN}\omega}). \tag{20}$$

By the choice of χ, $\int_B \Xi_k(A_\omega^{(N)}) \geq -\eta N/72$ for any set, B, of measure at most χ. Hence, the expected value of the expression in (20) is bounded below by $-\eta nN/12$.

Combining these estimates along all filler blocks occuring in (11), we see

$$\mathbb{E}\left(\sum_{j=0}^{n-1} \mathbf{1}_{\mathsf{Filler}}(\bar\sigma^{jN}\bar\omega)\big(\tilde\Xi_k(A^{(N-2)}_{\sigma^{jN+1}\omega}) - \Xi_k(A^{(N)}_{\sigma^{jN}\omega})\big)\right) \geq -\eta nN/6. \tag{21}$$

9.9. **Combining the inequalities.** At this point, we have (combining inequalities (11), (16), (18) and (21)),

$$\begin{aligned}&\mathbb{E}\left(\tilde\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(nN)})\right) \geq \\ &\mathbb{E}\left(\Xi_k(\bar B^0) + \Xi_k(G^0) + \Xi_k(\bar B^0) + \Xi_k(\bar B^0) + \Xi_k(\bar B^0) + \dots\right) - E_1 - E_2 - E_3,\end{aligned} \tag{22}$$

where E_3 comes from the contributions of (18) and (21). Then using (10),

$$\begin{aligned}&\mathbb{E}\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(nN)}) \\ &\geq \mathbb{E}\left(\Xi_k(\bar B^0) + \Xi_k(G^0) + \Xi_k(\bar B^0) + \Xi_k(\bar B^0) + \Xi_k(\bar B^0) + \dots\right) - C_{34} \\ &\quad - \left(\tfrac{1}{6}\eta nN\right) - \left(\tfrac{1}{6}(\eta N/\chi)\,\mathbb{E}\,n_{\mathsf{Super}} + C_{27}\,\mathbb{E}\,n_{\mathsf{Super}} - 11kn\log\delta\right) \\ &\quad - \left(\tfrac{1}{18}(\eta N/\chi)\,\mathbb{E}\,n_{\mathsf{Filler}} + \tfrac{1}{6}\eta nN\right),\end{aligned}$$

where n_{Filler} and n_{Super} are the number of filler and super-blocks respectively in $A^{\epsilon}{}_{\bar\omega}^{(nN)}$. By sub-additivity, the first term in parentheses is at least $\mathbb{E}\,\Xi_k(A_\omega^{(nN)})$. We have $\mathbb{E}\,n_{\mathsf{Filler}} < 3\chi n$ and $\mathbb{E}\,n_{\mathsf{Super}} < \chi n$,

$$\mathbb{E}\,\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(nN)}) \geq \mathbb{E}\,\Xi_k(A_\omega^{(nN)}) - C_{34} - \tfrac{2}{3}\eta nN - C_{27}\chi n + 11kn\log\delta.$$

As ϵ is reduced to 0, δ does not grow, but $N \to \infty$ so that for sufficiently small ϵ, we have

$$\mathbb{E}\,\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(nN)}) \geq \mathbb{E}\,\Xi_k(A_\omega^{(nN)}) - \eta nN.$$

Hence we deduce $\Lambda_k(A^\epsilon) \geq \Lambda_k(A) - \eta$, as required. □

10. Convergence of the Oseledets spaces

Proof of Theorem B. Let $k = D_i$ be as in the statement of the theorem. Let us assume, by possibly rescaling the cocycle by a constant, that $\mu_k > 0 > \mu_{k+1}$. Let $\delta_0 < 1$ and

$$U_\epsilon = \left\{\bar\omega\colon \angle\big(E_k^\epsilon(\bar\omega), E_k(\omega)\big) > 2\delta_0\right\}.$$

We will show that for every $0 < \eta < 1$ and every sufficiently small $\epsilon > 0$, $\bar{\mathbb{P}}(U_\epsilon) < \eta$.

Once this is established, convergence in probability of the Oseledets spaces $Y_k^\epsilon(\bar\omega)$ to $Y_k^0(\omega)$ follows via the identity $Y_k^\epsilon(\bar\omega) = E_k^\epsilon(\bar\omega) \cap F_{k-1}^\epsilon(\bar\omega)$, and the fact that $F_{k-1}^\epsilon(\bar\omega)$ coincides with the orthogonal complement of the top k-dimensional Oseledets space of the adjoint cocycle $(A^\epsilon)^*$, which converges in probability by the same argument. See [10, §4] for details.

In what follows, we will repeatedly apply Lemma 8, assuming $\xi < \frac{\eta}{2}, \delta_1 < \min\{\delta_0, \frac{\mu_k \eta}{10k}\}$, and so the value of τ provided by Lemma 8 satisfies $\tau \le \delta_1 \le \frac{\mu_k \eta}{10k}$.

Let $W_\epsilon = \bar\sigma^{-N} U_\epsilon \cap \bar G$, where N depends on ϵ as in Lemma 9. For sufficiently small ϵ, we have $\bar{\mathbb{P}}(\bar G \cap \bar\sigma^{-N} \bar G) \ge 1 - \frac{\eta}{2}$, so that once we show $\bar{\mathbb{P}}(W_\epsilon) < \frac{\eta}{2}$, we will be able to conclude that $\bar{\mathbb{P}}(U_\epsilon) = \bar{\mathbb{P}}(\bar\sigma^{-N} U_\epsilon) \le \bar{\mathbb{P}}(W_\epsilon) + \bar{\mathbb{P}}(\bar G^c) < \eta$.

Lemma 35. *Suppose that* $\bar\omega \in \bar G$, *and that* $\angle(E_k^\epsilon(\bar\sigma^N \bar\omega), E_k(\sigma^N \omega)) > 2\delta$. *Then* $\perp(E_k^\epsilon(\bar\omega), F_k(A_\omega^{(N)})) < 4\delta^{-1} e^{-(\mu_k - \tau)N}$.

Proof. We suppose the contrapositive: assume that $\perp(E_k^\epsilon(\bar\omega), F_k(A_\omega^{(N)})) \ge 4\delta^{-1} e^{-(\mu_k - \tau)N}$. Then let $v \in E_k^\epsilon(\bar\omega) \cap S$. Write $v = u + w$ with $u \in F_k(A_\omega^{(N)})$ and $w \in F_k(A_\omega^{(N)})^\perp$. Then $A^{\epsilon}{}_{\bar\omega}^{(N)} v = A_\omega^{(N)} u + A_\omega^{(N)} w + z$, where $z = (A^{\epsilon}{}_{\bar\omega}^{(N)} - A_\omega^{(N)}) v$.

By Lemmas 8(d) and 9, $\|A_\omega^{(N)} u\| \le 1$ and $\|z\| \le 1$, while $\|A_\omega^{(N)} w\| \ge 4\delta^{-1}$. Recalling that $A^{\epsilon}{}_{\bar\omega}^{(N)} E_k^\epsilon(\bar\omega) = E_k^\epsilon(\bar\sigma^N \bar\omega)$ and normalizing, we see that an arbitrary point of $E_k^\epsilon(\bar\sigma^N \bar\omega) \cap S$ lies within $\delta/2$ of $E_k(A_\omega^{(N)})$, and so within δ of $E_k(A_\omega^{(N)}) \cap S$. By Lemma 2, we deduce $\angle(E_k^\epsilon(\bar\sigma^N \bar\omega), E_k(A_\omega^{(N)})) < \delta$. Now by Lemma 8(b), we see that $\angle(E_k^\epsilon(\bar\sigma^N \bar\omega), E_k(\sigma^N \omega)) < 2\delta$. ☐

Lemma 36. *If* ϵ *is sufficiently small so that* $4\delta^{-1} + 2 < e^{k\tau N}$, $\bar\omega \in \bar G$ *and* $\perp(E_k^\epsilon(\bar\omega), F_k(A_\omega^{(N)})) < 4\delta^{-1} e^{-(\mu_k - \tau)N}$, *we have*

$$\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(N)}|_{E_k^\epsilon(\bar\omega)}) \le (\mu_1 + \ldots + \mu_{k-1} + 2k\tau)N.$$

Proof. By hypothesis, there exists a unit length $v \in E_k^\epsilon(\bar\omega)$ such that $v = f + f^\perp$, with $f \in F_k(A_\omega^{(N)})$, $f^\perp \in F_k(A_\omega^{(N)})^\perp$ and $\|f^\perp\| < 4\delta^{-1} e^{-(\mu_k - \tau)N}$.

Now, since $E_k^\epsilon(\bar\omega)$ is k-dimensional, $\Xi_k(A^{\epsilon}{}_{\bar\omega}^{(N)}|_{E_k^\epsilon(\bar\omega)})$ is the logarithm of the volume growth of any k-dimensional parallelepiped in $E_k^\epsilon(\bar\omega)$ under $A^{\epsilon}{}_{\bar\omega}^{(N)}$. Let $v, v_2, \ldots, v_k$ be an orthonormal basis for $E_k^\epsilon(\bar\omega)$. Then,

$$\begin{aligned}\mathrm{Vol}(A^{\epsilon}{}_{\bar\omega}^{(N)} v, A^{\epsilon}{}_{\bar\omega}^{(N)} v_2, \ldots, A^{\epsilon}{}_{\bar\omega}^{(N)} v_k) &\le \mathrm{Vol}(A^{\epsilon}{}_{\bar\omega}^{(N)} f, A^{\epsilon}{}_{\bar\omega}^{(N)} v_2, \ldots, A^{\epsilon}{}_{\bar\omega}^{(N)} v_k) \\ &\quad + \mathrm{Vol}(A^{\epsilon}{}_{\bar\omega}^{(N)} f^\perp, A^{\epsilon}{}_{\bar\omega}^{(N)} v_2, \ldots, A^{\epsilon}{}_{\bar\omega}^{(N)} v_k).\end{aligned}$$

By the choice of f,

$$\begin{aligned}\mathrm{Vol}(A^{\epsilon(N)}_{\bar\omega} f, A^{\epsilon(N)}_{\bar\omega} v_2, \dots, A^{\epsilon(N)}_{\bar\omega} v_k) &\leq \|A^{\epsilon(N)}_{\bar\omega} f\| e^{\Xi_{k-1}(A^{\epsilon(N)}_{\bar\omega})} \\ &\leq 2e^{(\mu_1+\ldots+\mu_{k-1}+(k-1)\tau)N},\end{aligned}$$

where we have used that $\|A^{\epsilon(N)}_{\bar\omega} f\| \leq \|A^{(N)}_{\omega} f\| + \|A^{\epsilon(N)}_{\bar\omega} f - A^{(N)}_{\omega} f\| \leq 2$.

Since $\|f^\perp\| < 4\delta^{-1}e^{-(\mu_k-\tau)N}$, then $\mathrm{Vol}(A^{\epsilon(N)}_{\bar\omega} f^\perp, A^{\epsilon(N)}_{\bar\omega} v_2, \dots, A^{\epsilon(N)}_{\bar\omega} v_k) \leq \|f^\perp\| e^{\Xi_k(A^{\epsilon(N)}_{\bar\omega})} < 4\delta^{-1}e^{(\mu_1+\ldots+\mu_{k-1}+k\tau)N}$. □

Lemma 37. *There exists $\epsilon_0 > 0$ and $M \in \mathbb{N}$ such that for every $\epsilon < \epsilon_0, N \geq M$ and $B \subset \bar\Omega$, we have that*

$$\int_B \Xi_k(A^{\epsilon(N)}_{\bar\omega})\, d\bar{\mathbb{P}} < N(\mu_1 + \cdots + \mu_k)\bar{\mathbb{P}}(B) + 2\tau N.$$

In particular, for all sufficiently small ϵ, the above holds for N chosen as in Lemma 9.

Proof. By the L^1 convergence in the sub-additive ergodic theorem, there exists $M > 0$ be such that $\|\Xi_k(A^{(n)}_\omega) - n(\mu_1 + \cdots + \mu_k)\|_1 \leq n\tau$ for every $n \geq M$. In particular, for every $n \geq M$ and every $B \subset \bar\Omega$,

$$\int_B \Xi_k(A^{(n)}_\omega)\, d\bar{\mathbb{P}} < n(\mu_1 + \cdots + \mu_k)\bar{\mathbb{P}}(B) + n\tau.$$

Notice that $\Xi_k(A^{\epsilon(n)}_{\bar\omega}) \leq k \log^+ \|A^{\epsilon(n)}_{\bar\omega}\|_{\mathsf{op}} \leq k \sum_{j=0}^{n-1} (\log^+ \|A_{\sigma^j\omega}\|_{\mathsf{op}} + \epsilon\|\Delta_j\|_{\mathsf{op}})$, where we have used the fact that $\log^+(x+y) \leq \log^+(x) + |y|$. For a fixed n, this shows that the family of functions $g_\epsilon(\bar\omega) = \Xi_k(A^{\epsilon(n)}_{\bar\omega})$ for $0 \leq \epsilon < 1$ is dominated, and converges as $\epsilon \to 0$ to $\Xi_k(A^{(n)}_\omega)$. Hence, by the reverse Fatou lemma, for sufficiently small $\epsilon > 0$, $n \in \{M, \dots, 2M-1\}$ and every $B \subset \bar\Omega$,

$$\int_B \Xi_k(A^{\epsilon(n)}_{\bar\omega})\, d\bar{\mathbb{P}} < n(\mu_1 + \cdots + \mu_k)\bar{\mathbb{P}}(B) + 2\tau n.$$

Using sub-additivity of Ξ_k, we conclude that for every $N \geq M$, and every $B \subset \bar\Omega$,

$$\int_B \Xi_k(A^{\epsilon(N)}_{\bar\omega})\, d\bar{\mathbb{P}} < N(\mu_1 + \cdots + \mu_k)\bar{\mathbb{P}}(B) + 2\tau N.$$

□

Notice that if $\bar\omega \in W_\epsilon$, then by Lemmas 35 and 36 (the first lemma establishing the hypothesis of the next one), then $\Xi_k(A^{\epsilon(N)}_{\bar\omega}|_{E^\epsilon_k(\bar\omega)}) \leq$

$(\mu_1 + \ldots + \mu_{k-1} + 2k\tau)N$. Combining this with Lemma 37, we see

$$\begin{aligned}\mu_1^\epsilon + \ldots + \mu_k^\epsilon &= \lim_{n\to\infty} \frac{1}{n} \int \Xi_k(A^{\epsilon(n)}_{\bar\omega}|_{E_k^\epsilon(\bar\omega)})\, d\bar{\mathbb{P}}(\bar\omega) \\ &\le \frac{1}{N} \int_{W_\epsilon} \Xi_k(A^{\epsilon(N)}_{\bar\omega}|_{E_k^\epsilon(\bar\omega)})\, d\bar{\mathbb{P}}(\bar\omega) + \frac{1}{N} \int_{W_\epsilon^c} \Xi_k(A^{\epsilon(N)}_{\bar\omega})\, d\bar{\mathbb{P}}(\bar\omega) \\ &\le (\mu_1 + \ldots + \mu_{k-1} + 2k\tau)\bar{\mathbb{P}}(W_\epsilon) + (\mu_1 + \cdots + \mu_k)\bar{\mathbb{P}}(W_\epsilon^c) + 2\tau.\end{aligned}$$

Hence,

$$\mu_k \bar{\mathbb{P}}(W_\epsilon) \le (\mu_1 + \ldots + \mu_k) - (\mu_1^\epsilon + \ldots + \mu_k^\epsilon) + 4k\tau.$$

In particular, in view of the convergence of the exponents, for all sufficiently small ϵ, we have $\bar{\mathbb{P}}(W_\epsilon) \le 5k\tau/\mu_k < \frac{\eta}{2}$. □

Acknowledgements

GF and AQ acknowledge partial support from the Australian Research Council (DP150100017). The research of CGT has been supported by the Australian Research Council (DE160100147). AQ acknowledges the support of NSERC. The authors are grateful to the the School of Mathematics and Statistics at the University of New South Wales, the School of Mathematics and Physics at the University of Queensland and the Department of Mathematics and Statistics at the University of Victoria for their hospitality, allowing for research collaborations which led to this project.

References

[1] V. Baladi, A. Kondah, and B. Schmitt. Random correlations for small perturbations of expanding maps. *Random Comput. Dynam.*, 4(2-3):179–204, 1996.

[2] A. Blumenthal, J. Xue, and L.-S. Young. Lyapunov exponents for random perturbations of some area-preserving maps including the standard map. *Ann. of Math. (2)*, 185(1):285–310, 2017.

[3] J. Bochi. Genericity of zero Lyapunov exponents. *Ergodic Theory and Dynamical Systems*, 22(6):1667–1696, 12 2002.

[4] J. Bochi and M. Viana. The Lyapunov exponents of generic volume-preserving and symplectic maps. *Ann. Math.*, 161:1423–1485, 2005.

[5] C. Bocker-Neto and M. Viana. Continuity of Lyapunov Exponents for Random 2D Matrices. *ArXiv e-prints*, Dec. 2010.

[6] T. Bogenschütz. Stochastic stability of invariant subspaces. *Ergodic Theory Dynam. Systems*, 20(3):663–680, 2000.

[7] R. Durrett. *Probability: Theory and Examples (4th Edition)*. Cambridge Univ. Press, 2010.

[8] G. Froyland and C. González-Tokman. Stability and approximation of invariant measures of Markov chains in random environments. *Stoch. Dyn.*, 16(1):1650003, 23, 2016.

[9] G. Froyland, C. González-Tokman, and A. Quas. Stability and approximation of random invariant densities for Lasota-Yorke map cocycles. *Nonlinearity*, 27:647–660, 2014.

[10] G. Froyland, C. González-Tokman, and A. Quas. Stochastic stability of Lyapunov exponents and Oseledets splittings for semi-invertible matrix cocycles. *Comm. Pure Appl. Math.*, 68:2052–2081, 2015.

[11] G. Froyland, S. Lloyd, and A. Quas. Coherent structures and isolated spectrum for Perron-Frobenius cocycles. *Ergodic Theory Dynam. Systems*, 30:729–756, 2010.

[12] G. Froyland, S. Lloyd, and A. Quas. A semi-invertible Oseledets theorem with applications to transfer operator cocycles. *Discrete Contin. Dyn. Syst.*, 33(9):3835–3860, 2013.

[13] G. Froyland, S. Lloyd, and N. Santitissadeekorn. Coherent sets for nonautonomous dynamical systems. *Physica D: Nonlinear Phenomena*, 239(16):1527–1541, 2010.

[14] C. González-Tokman and A. Quas. A semi-invertible operator Oseledets theorem. *Ergodic Theory Dynam. Systems*, 34:1230–1272, 2014.

[15] C. González-Tokman and A. Quas. A concise proof of the multiplicative ergodic theorem on Banach spaces. *J. Mod. Dyn.*, 9:237–255, 2015.

[16] H. Hennion. Loi des grands nombres et perturbations pour des produits réductibles de matrices aléatoires indépendantes. *Z. Wahrsch. Verw. Gebiete*, 67(3):265–278, 1984.

[17] T. Kato. *Perturbation Theory for linear operators*. Springer Verlag, 1966.

[18] Y. Kifer. Perturbations of random matrix products. *Z. Wahrsch. Verw. Gebiete*, 61(1):83–95, 1982.

[19] Y. Kifer and E. Slud. Perturbations of random matrix products in a reducible case. *Ergodic Theory Dynam. Systems*, 2(3-4):367–382 (1983), 1982.

[20] S. Lang. *Fundamentals of Diophantine Geometry*. Springer-Verlag, Berlin, 1983.

[21] F. Ledrappier and L.-S. Young. Stability of Lyapunov exponents. *Ergodic Theory Dynam. Systems*, 11:469–484, 1991.

[22] G. Ochs. Stability of Oseledets spaces is equivalent to stability of Lyapunov exponents. *Dynam. Stability Systems*, 14(2):183–201, 1999.

[23] D. Ruelle. Analycity properties of the characteristic exponents of random matrix products. *Adv. in Math.*, 32(1):68–80, 1979.

[24] L.-S. Young. Random perturbations of matrix cocycles. *Ergodic Theory Dynam. Systems*, 6(4):627–637, 1986.

E-mail address: g.froyland@unsw.edu.au

E-mail address: cecilia.gt@uq.edu.au

E-mail address: aquas@uvic.ca