Generalized Lyapunov exponents and aspects of the theory of deep learning

Anders Karlsson*

December 26, 2022

Abstract

We discuss certain recent metric space methods and some of the possibilities these methods provide, with special focus on various generalizations of Lyapunov exponents originally appearing in the theory of dynamical systems and differential equations. These generalizations appear for example in topology, group theory, probability theory, operator theory and deep learning.

1 Introduction

The law of large numbers states that for a sequence of independent, identically distributed (i.i.d.) random variables $X_1, X_2, ..., X_n$ with finite expectation,

$$(X_1 + X_2 + ... + X_n)/n \to \mathbb{E}[X_1] \tag{1}$$

almost surely as $n \to \infty$. Bellman [Be54], Furstenberg [Fu63] and others asked whether in some situations there could exist a similar limit law for products

$$u(n) := g_1 g_2 g_3 ... g_n$$

of i.i.d. noncommutative operations $g_{1,} g_2, ..., g_n$. Such products appear for example as solutions to difference equations with random coefficients, or from time-one maps of the solutions of continuous models, say from a stochastic PDE. In addition to mathematics and physics (early references being [Dy53, S57]), one can find papers in biology, epidemiology, medicine, and economics leading to random products of noncommuting transformations [BHS21, CDS09, IS96, Neu19]. Compositional products is also one of the key features of deep learning as will be highlighted below.

Note that in contrast to (1) it is unclear how to form an average in the noncommutative setting. Important partial answers to the above question were obtained at the end of the 1960s ([Ki68, O68]) and later as one aspect of random walks on groups, see for example [Gu80, Ka00, KM99, Er10, BQ16, MT18, Zh22]. A quite general affirmative answer to the question of a limit law for noncommuting random products was provided in [KL06, GK20], see Theorem 3 below.

In ergodic theory one formalizes the setting as follows, more general than the i.i.d. assumption. Let (Ω, μ) be a measure space with $\mu(\Omega) = 1$. Let $T : \Omega \to \Omega$ be a measurable map preserving

*The author was supported in part by the Swedish Research Council grant 104651320 and the Swiss NSF grants 200020-200400 and 200021-212864.

the measure. We furthermore assume ergodicity, which is an irreducibility assumption that states that up to measure zero there are no T-invariant subsets of Ω. Given a measurable map $g : \Omega \to G$ (assigning some measurable structure on the group G; g is what a probabilist would call a random variable), we define the following *ergodic cocycle*:

$$u(n, \omega) := g(\omega)g(T\omega)...g(T^{n-1}\omega).$$

In addition, one needs to assume that the cocycle is integrable which means that the integral over Ω of the "size" of g is finite.

For matrices, a first answer was provided by Furstenberg-Kesten for the norm of the matrices, and a more precise answer was given later in the 1960s by Oseledets in his multiplicative ergodic theorem. One can view this as a random spectral theorem, intuitively it says that the random product behaves in the same way as the powers of one single "average" matrix:

Theorem 1. (Oseledets' multiplicative ergodic theorem [O68]) *Let $A(n, \omega) = g_n g_{n-1}...g_1$ be an integrable ergodic cocycle of invertible matrices. Then there are a.s. a random filtration of subspaces $0 = V_0 \subset V_1 \subset ... \subset V_k = \mathbb{R}^d$ and numbers $\lambda_1 < \lambda_2 < ... < \lambda_k$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \|A(n, \omega)v\| = \lambda_i$$

whenever $v \in V_i \setminus V_{i-1}$.

The case $d = 1$ is the Birkhoff ergodic theorem generalizing (1). The numbers λ_i are called *Lyapunov exponents*. This is a fundamental theorem in the theory of differentiable dynamical systems and has also many other applications, the most spectacular such is Margulis' proof [M75] of his super-rigidity theorem. As for physics, Nobel laureate Parisi wrote in the foreword of [CPV93] that "The properties of random matrices and their products form a basic tool, whose importance cannot be underestimated. They play a role as important as Fourier transforms for differential equations."

Now compare the above with the following. Let Σ be an oriented closed surface of genus $g \geq 2$. Let $\mathcal{S}$ denote the isotopy classes of simple closed curves on M not isotopically trivial. For a Riemannian metric ρ on Σ, let $l_\rho(\beta)$ be the infimum of the length of curves isotopic to β. In a legendary preprint from 1976 [T88], Thurston announced the following (the details are worked out in [FLP79, Théorème Spectrale] using foliation theory):

Theorem 2. (Thurston's spectral theorem for surface diffeomorphisms [T88]) *Let f be a diffeomorphism of a surface Σ of genus $g \geq 2$. Then there is filtration of subsurfaces $Y_1 \subset Y_2 \subset ... \subset Y_k = \Sigma$ and algebraic integers $\lambda_1 < \lambda_2 < .. < \lambda_k$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \log l_\rho(f^n c) = \lambda_i$$

whenever the simple closed curve c can be isotoped to a curve contained in Y_i but not in Y_{i-1}.

This is analogous to a simple statement for linear transformations A in finite dimensions (which corresponds to the Oseledets theorem in the case $A(n, \omega) = A^n$): given a vector v there is an associated exponent λ (absolute value of an eigenvalue), such that

$$\lim_{n\to\infty} \|A^n v\|^{1/n} = \lambda.$$

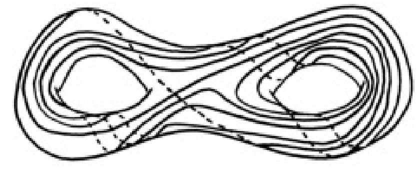

Figure 1: $f^n(c)$ (illustration from [T88])

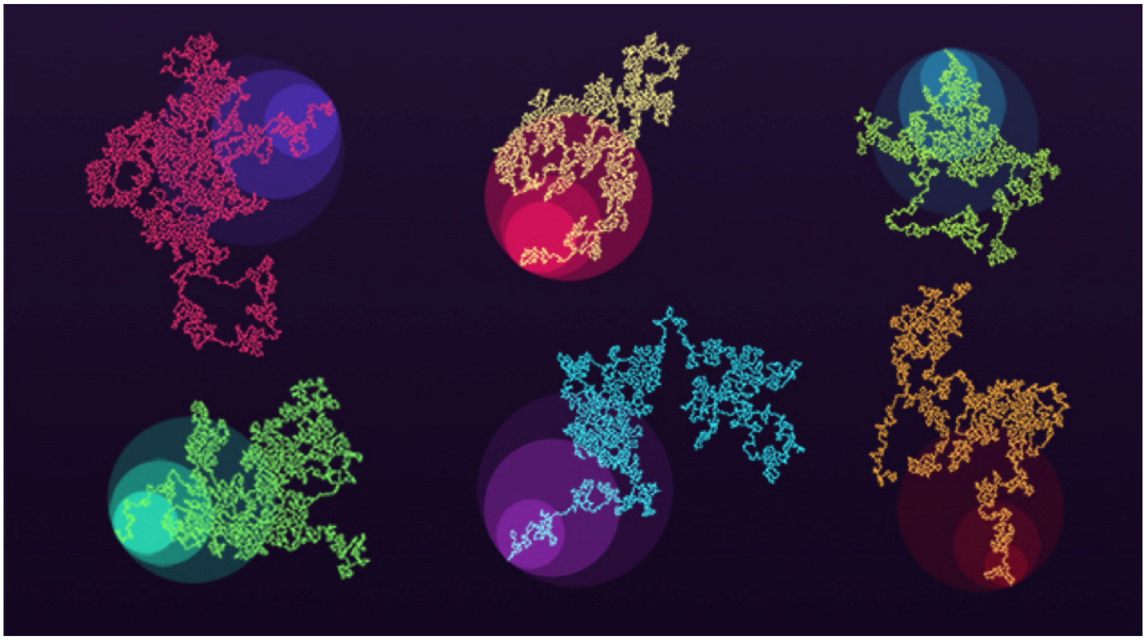

Figure 2: Illustration of the convergence in direction of $u(n,\omega)x$ as $n \to \infty$ in six experiments in a hyperbolic disk. The colored disks are horodisks that the random walks trajectories go deeper and deeper into with time. (Image by Cécile Bucher.)

To spell out the analogy: a diffeomorphism f instead of a linear transformation A, a length instead of a norm, and a curve α instead of a vector v. And the answer is given in similar terms: Lyapunov exponents and associated filtration of subspaces and subsurfaces respectively.

A *weak metric space* (following the terminology of for example [GuW22]) is a set X equipped with a function $X \times X \to \mathbb{R}$ such that

$$d(x,x) = 0$$

and

$$d(x,y) \leq d(x,z) + d(z,y)$$

for all points $x, y, z \in X$. A map $f : X \to X$ is called *nonexpansive* if

$$d(f(x), f(y)) \leq d(x,y)$$

for all $x, y \in X$.

The following multiplicative ergodic theorem was proved for isometries in [KL06] and in general in [GK20].

Theorem 3. (Ergodic theorem for noncommuting random products, [KL06, GK20]) *Let $u(n,\omega)$ be an integrable ergodic cocycle of nonexpansive maps of a weak metric space (X,d) and such that $\omega \mapsto u(n,\omega)x$ is measurable. Then there exists a.s. a metric functional h of X such that*

$$\lim_{n\to\infty} -\frac{1}{n} h(u(n,\omega)x) = \lim_{n\to\infty} \frac{1}{n} d(x, u(n,\omega)x).$$

Metric functionals are almost what is usually called horofunctions, see section 3. This theorem when specialized to X a symmetric space of nonpositive curvature, Gromov hyperbolic space, or CAT(0) space, recovers some previous results mentioned above by Oseledets, Furstenberg, Kaimanovich, and Karlsson-Margulis. It also implies random mean ergodic theorems of Ulam-von Neumann, Kakutani, and Beck-Schwartz, see [GK20], and it holds even when the traditional mean ergodic theorem fails [K21]. Theorem 3 furthermore provides generalized laws of large numbers with concave moments [KMo08], and has found application to random walks on groups and bounded harmonic functions on manifolds [KL07, KL07b] without knowing anything specific about the metric functionals in these settings. I want to emphasize that Theorem 3 applies in particular to every random walks with finite first moment on *any* finitely generated group.

Moreover, using Theorem 3, Horbez, building on the approach of [K14], could establish the following random extension of Thurston's theorem:

Theorem 4. (Random spectral theorem of surface homeomorphisms [K14, H16]) *Let $v(n,\omega) = A(T^{n-1}\omega)...A(T\omega)A(\omega)$ be an integrable i.i.d random product of homeomorphisms of a closed surface Σ of genus $g \geq 2$. Then there is a (random) filtration of subsurfaces $Y_1 \subset Y_2 \subset ... \subset Y_k = \Sigma$ and (deterministic) exponents $\lambda_1 < \lambda_2 < .. < \lambda_k$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \log l_\rho(v(n,\omega)c) = \lambda_i$$

whenever the simple closed curve c can be isotoped to a curve contained in Y_i but not in Y_{i-1}. Here l_ρ is the minimal length in the isotopy class in some fixed Riemannian metric ρ.

The exponents λ_i are a type of generalized Lyapunov exponents that perhaps could be called *topological Lyapunov exponents for surface homeomorphisms.* A different approach was provided in [K18] which showed how to get the top exponent for ergodic cocycles of homeomorphisms, using the metric ideas and a lemma in [KM99] combining it with results in [LRT12]. In both cases, the proofs use Thurston's asymmetric metric. A prior study of random walks on the mapping class groups was carried out in [KM96], which showed in particular that under a non-elementary assumption the random walk converges to uniquely ergodic foliations, in which case it follows from [K14] that there is only one exponent.

Actually the main results of Horbez' paper [H16] concern instead random walks on the outer automorphism group of free groups, giving a result very similar to Theorem 4. In order not to have to explain notations from the important subject of automorphisms of free groups, I will not state it here. The proof goes via a determination of the metric functionals of the outer space and an application of Theorem 3. To get all generalized Lyapunov exponents Horbez then studies the set of stationary measures on the boundary of outer space in parallel to works in the matrix case of Furstenberg-Kifer and Hennion.

We thus see three settings, linear transformations in finite dimensions, surface homeomorphisms, and automorphisms of free groups, that are not merely analogous but the corresponding "law of large numbers" can be deduced ultimately from the same theorem. The strategy is:

- Instead of looking at the underlying space where the linear maps, homeomorphisms, group automorphisms etc act, we lift the action to a more abstract space, a *moduli space* as it were, of positive structures on the corresponding underlying space.
- On that associated space there is often an invariant metric.
- Employ the noncommutative ergodic theorem in terms of metric functionals and interpret the result as concretely as possible.

The very last part can in fact often be done, as testified by Theorems 1 and 4 above which include no reference to metric functionals (or horofunctions), and likewise in Theorem 10 of the last section on deep learning. Sometimes, like in complex analysis, Cayley graphs, or maps of cones there is no need to pass to an auxiliary space in order to find an invariant metric for the transformations in question.

Further generalized Lyapunov exponents could perhaps also be defined for higher dimensional diffeomorphisms, using their isometric action on Ebin's space of Riemannian metrics on a fixed compact manifold [Eb68] and an investigation of the metric functionals. Some progress and possibilities are pointed out in section 5.

A different direction, using some of the arguments in [KL06], was developed by Masai [Ma21], namely for surface bundles over a circle, he showed that for pseudo-Anosov maps the translation length (="top Lyapunov exponent" in the terminology of the present paper) in a certain weak metric equals the 3-dimensional hyperbolic volume of the mapping torus that the map defines.

Ackonwledgement: This text was written in connection with the conference "New Trends in Lyapunov Exponents" in Lisbon 2022. I thank the organizers and especially Pedro Duarte for the invitation to this very pleasant and stimulating week. I also thank Alex Blumenthal for helpful discussions related to the topics of this paper during this meeting.

2 Deep learning

Deep learning provided Artificial Intelligence (AI) with a long sought-after new tool that moreover exceeded all expectation, as was realized starting from around 2012. The development of deep neural network had begun much earlier and Bengio, Hinton, and LeCun received the 2018 Turing award for these methods. Part of the prize citation stated "By dramatically improving the ability of computers to make sense of the world, deep neural networks are changing not just the field of computing, but nearly every field of science and human endeavor."

The remarkable success of these methods indicates that real-life data tend to have a compositional structure. More precisely, given a learning task, one seeks maps $g_1, g_2 ... g_n$ such that their composition

$$u(n) := g_1 g_2 ... g_n$$

applied to the input data should be close to the desired output (possibly after applying a certain *decision function* f). The *depth* n can be several hundred. The maps are often of the form $g_i(x) = \sigma(W_i x + b_i)$ where σ is a fixed nonlinear function, called *activation function*, applied componentwise, W_i is a $d \times d$ matrix, called *weights*, and b_i is a vector in $\mathbb{R}^d$, called *bias vector*. The dimension d is called the *width.* The nonlinearity, inspired by our brains ([MP43, Ro58]), is crucial (for one thing, the composition of affine maps are again affine, and likewise the composition of polynomials is again a polynomial). Some standard choices are, sigmoid/logistic function ($1/(1+e^{-t})$), TanH ($\tanh(t)$), and ReLU (Rectified Linear Unit, $\sigma(t) = \max\{0, t\}$), with different features and advantages. ReLU has been observed to work particularly well, generally better than smooth functions.

In practice, the weights and biases in the neural network are first randomly selected (*initialization*) and then optimized by stochastic gradient descent on a chosen loss function specific to the task (*training*).

Much of the subject of deep learning consists of empirical observations, there is no or little

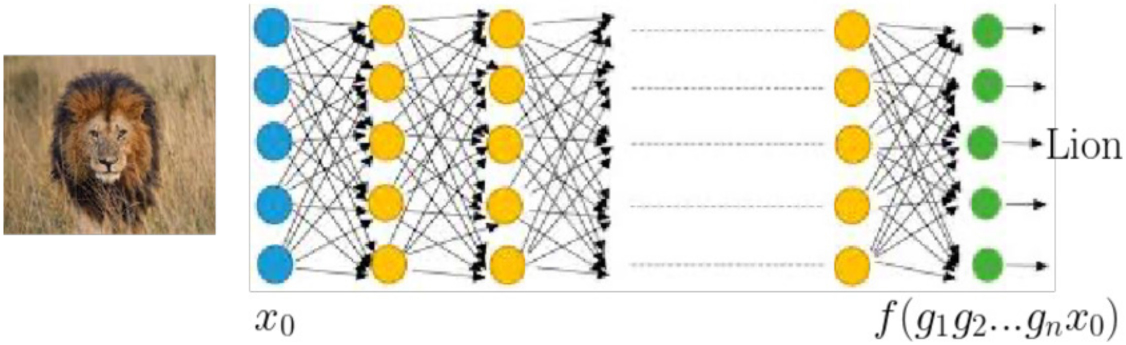

Figure 3: A deep neural network

theoretical understanding, as remarked by many authors. According to [Se20] "A mathematical theory of deep learning would illuminate how they function, allow us to assess the strengths and weaknesses of different network architectures, and lead to major improvements." Which type of layer maps to take, which σ, how many layers n, how many nodes d in each layer, how to best find the parameters, how stable the solution is under random perturbation (such as the *drop-out procedure*) are some of the questions of important practical concern. The need for a theoretical understanding, instead of relying on black-box techniques, is also expressed by practitioners, this lack of theory hinders their work.

One of the remarkable features that is not understood, is why deep neural networks generally seem to mostly avoid the problem of overfitting which is a phenomenon in traditional statistics. The latter typically happens when approximating some data with a polynomial of very high degree, the curves go through all the sample or training data, but inbetween these points of perfect fit it can fluctuate wildly, related to the Runge phenomenon. This is clearly undesirable.

There are several ways random products $u(n) := g_1 g_2 ... g_n$ of noncommuting nonlinear maps appear in deep learning:

1. *Random initialization* see [NBYS22] for a review
2. *Drop-out regularization* which in particular is used to verify robustness of the obtained error minimizer, and also a way of training the network [SHK14]
3. *Bayesian learning* [Ne12]
4. Learning that combines taking some maps g_i at random and optimize the remaining one, a procedure with apparently good performance that speeds up the training significantly [BT22]

The first two concepts are so fundamental in the current state-of-the-art that one encounters them after any couple of first lectures on deep learning. Hanin wrote in [Ha21] "Beyond illuminating the properties of networks at the start of training, the analysis of random neural networks can reveal a great deal about networks after training as well."

Moreover, as Avelin pointed out to me, not only the initialization but also the training (stochastic gradient descent) actually involves a random product of transformations. Thus we see compositional product of random operations appearing in several ways in deep learning. Since the number of layer maps can approach a thousand, it should make limit theorems discussed in this article very relevant, see the last section.

3 Elements of a metric functional analysis

A metric space is a set equipped with a distance function $d(x,y)$ that is semi-positive, symmetric and satisfy the triangle inequality. The author argued in [K21] that it is useful to develop parts of metric geometry in analogy with linear functional analysis.

Sometimes various ways of weakening the notion of a metric are useful: pseudo-metrics arise naturally in complex analysis and here we will also allow for asymmetric metrics (as in Thurston's metric). Moreover we will let d possibly to take negative values, useful for topical maps. In other words, we consider weak metrics as defined in the introduction. Weak metrics (but taking only nonnegative values) were in fact already of interest to people like Heinz Hopf in the 1940s, see [Ri43] pointed out in [PT09].

Note that a symmetrization such as $D(x,y) := \max\{d(x,y), d(y,x)\}$ (or the sum) is nonnegative following from

$$0 = d(x,x) \leq d(x,y) + d(y,x).$$

With the pseudo-metric D one defines a topology on X. In case the separation axiom holds, i.e. that $d(x,y) = 0$ implies $x = y$, then D is a genuine metric.

A map $f : X \to Y$ is *nonexpansive* if

$$d_Y(f(x_1), f(x_2)) \leq d_X(x_1, x_2)$$

for all $x_1, x_2 \in X$. Note that compositions of nonexpansive maps remain nonexpansive. Moreover, note that if we pass to a symmetrization of a weak metric, then f remains nonexpansive.

Let (X,d) be a weak metric space. We will now define the *metric compactification* of X, which will provide a weak topology with compactness properties in the metric setting. (I learnt from Cormac Walsh that this construction works without essential changes to asymmetric metrics, see [AGW09] and [W14] which inspired [K14], and see also the more recent paper [GuW22].)

Let $F(X,\mathbb{R})$ be the space of continuous functions $X \to \mathbb{R}$ equipped with the topology of pointwise convergence. Given a base point x_0 of the metric space X, let

$$\Phi : X \to \mathrm{F}(X,\mathbb{R})$$

be defined via

$$x \mapsto h_x(\cdot) := d(\cdot, x) - d(x_0, x).$$

Proposition 5. *The map Φ is a well-defined continuous map, and it is injective if d separates points. The closure $\overline{\Phi(X)}$ is compact.*

Proof. By the triangle inequality

$$h_x(y) - h_x(z) = d(y,x) - d(z,x) \leq d(y,z)$$

$$-h_x(y) + h_x(z) = -d(y,x) + d(z,x) \leq d(z,y)$$

therefore

$$|h_x(y) - h_x(z)| \leq \max\{d(y,z), d(z,y)\},$$

which in particular implies that h_x is continuous. This inequality clearly passes to the closure. The map Φ is continuous since

$$|h_x(z) - h_y(z)| = |d(z,x) - d(x_0,x) - d(z,y) + d(x_0,y)| \leq$$

$$2\max\{d(x,y), d(y,x)\}$$

by the usual triangle inequality and the one in X.

Suppose that d separates points, then given two points x and y, assume that $d(x_0,x) \geq d(x_0,y)$. If $d(x,y) > 0$, then

$$h_x(x) - h_y(x) = -d(x_0,x) - d(x,y) + d(x_0,y) \leq -d(x,y) < 0$$

shows that the two functions are different. In case $d(y,x) > 0$, then

$$h_y(x) - h_x(x) = d(y,x) - d(x_0,y) + d(x_0,x) \geq d(y,x) > 0.$$

These two cases cover all possibilities in view of the remark above about D, and proves the injectivity.

Finally note that by the triangle inequality

$$-d(x_0,y) \leq h_x(y) \leq d(y,x_0).$$

In view of this and the topology of pointwise topology which is the product topology, the Tychonov theorem implies that $\overline{\Phi(X)}$ is compact. $\square$

This proposition is the metric space analog of the Banach–Alaoglu theorem. We call $\overline{X} := \overline{\Phi(X)}$ the *metric compactification of* X and its elements *metric functionals,* which recently has been described concretely in a variety of metric spaces. This development is in parallel to the determination of dual spaces in the beginning of functional analysis a century ago. I reserve the more commonly used word *horofunction* for limits in the topology of uniform convergence on bounded sets (Gromov's choice of topology in [Gr81] considering genuine metric spaces) of h_{x_n} for sequences x_n such that $d(x_0,x_n) \to \infty$.

A note on the proof of Theorem 3 in the weak metric case: while [KL06] considered a skew-product extension to the boundary, the [GK20] paper established a new substantial refinement of the subadditive ergodic theorem [Ki68], which we feel is a refinement of the fundamental theorem of Kingman that has potential for further use. It has indeed already found independent dynamical applications in [KS19, CD20, ZC21]. In his book [Sz01] computer scientist Szpankowski explains why subadditivity and the subadditive ergodic theorem are fundamental for the analysis of algorithms.

As observed in [K14] the noncommutative ergodic theorem works with an asymmetric metric d. Here is the verification that it works even for weak metrics. First, it is of importance that

$$a(n,\omega) := d(x_0, u(n,\omega)x_0)$$

is a subadditive cocycle. This is verified as follows:

$$d(x_0, u(n+m,\omega)x_0) \leq d(x_0, u(n,\omega)x_0) + d(u(n,\omega)x_0, u(n+m,\omega)x_0)$$

$$\leq d(x_0, u(n,\omega)x_0) + d(x_0, u(m, T^n\omega)x_0).$$

Kingman's subadditive ergodic theorem then asserts that

$$\lim_{n\to\infty} \frac{1}{n} d(x, u(n,\omega)x)$$

exists a.e. under the integrability condition

$$\int_\Omega |d(x_0, u(1,\omega)x_0)|\, d\mu < \infty.$$

Note that nothing here depends on the choice of x_0 including the value of the limit therefore written with a general point x. And if we assume that (Ω, μ, T) is an ergodic system then this "top Lyapunov" exponent is deterministic, i.e. essentially constant in ω.

The proof of Theorem 3 in the weak metric setting now follows exactly [GK20, section 3] except that in that reference the order in which the metric is written is reversed.

A possible future direction: It seems plausible that often the limits

$$\lim_{n\to\infty} \frac{1}{n} h(u(n,\omega)x_0)$$

exist a.e. for any metric functional h. Evidence and discussion of this appear in some of my earlier papers, for example [K04] (this reference also contains an argument why Theorem 3 in the special case of CAT(0)-spaces is equivalent to geodesic ray approximation [Ka00, KM99]). From ray approximation (being of sublinear distance to a geodesic ray) and purely geometric reasons (any two geodesic rays have a well-defined linear rate of asymptotic divergence) all the above limits exist for proper CAT(0)-spaces and Gromov hyperbolic spaces. See also [Sa21] for a recent contribution to this topic. On the other hand, it is not true in general in view of the counterexample [KoN81, p. 272] for $\mathbb{R}^d$ with the ℓ^∞ norm and the metric functionals given for example in [K22].

4 Multiplicative ergodic theorems for linear operators

The need for multiplicative ergodic theorems for operators in infinite dimensions has been expressed in the influential articles [Ru82, ER85, LY12]. In one approach to the 2D Navier-Stokes equation and related evolution equations, the dynamics takes place in infinite dimensional Hilbert spaces. There has been an increasing interest in results on this topics, for example [LL10, GTQ15, Bl16, MN20, BHL20]. González-Tokman wrote in [GT18] that "An important motivation behind the recent work on multiplicative ergodic theorems is the desire to develop a mathematical theory which is useful for the study of global transport properties of real world dynamical systems, such as oceanic and atmospheric flows. Global features of the ocean flow include large scale structures which are important for the global climate." In a different direction, also leading to multiplicative ergodic theorems in Banach spaces is [CDS09], which deals with difference equations with random delays. Such delays are common in models of biological systems, immune response, epidemiology, and economics (see [CDS09] for references).

Given a bounded linear operator A, submultiplicativity implies that

$$\lim_{n\to\infty} \|A^n\|^{1/n}$$

exists and equals the spectral radius. On the other hand expressions

$$\|A^n v\|^{1/n}$$

as $n \to \infty$ may not converge in infinite dimensions, as is well known, see for example the introduction of [Sc06]. This puts a limitation on the validity of Oseledets theroem for operators. Kingman's theorem takes care of the regularity of the growth of the norm, and Theorem 3 may be the appropriate replacement for the second type of more directional behavior (local spectral theory).

The first infinite dimensional extension of Theorem 1 is Ruelle's theorem [Ru82] for compact operators, and other early results were shown by Mañe and Thieullen [Th87]. A strengthening of this for the Hilbert-Schmidt class was obtained in [KM99]:

Theorem 6. ([KM99]) *Let $u(n,\omega)$ be an ergodic cocycle of $Id + A$ operators where A is Hilbert-Schmidt. Then there is a.s. an operator Λ_ω such that*

$$\frac{1}{n}\left(\sum_i (\log \mu_i(n))^2\right)^{1/2} \to 0$$

as $n \to \infty$, where $\mu_i(n)$ are the eigenvalues of the positive part of $\Lambda_\omega^{-n} u(n,\omega)$.

The uniformity of the convergence implicit in the conclusion, thanks to the metric methods, is noteworthy since this is a much stronger statement in infinite dimensions. In finite dimensions the statement is equivalent to Oseledets' theorem. This metric approach was recently substantially extended to a von Neumann algebra setting with a finite trace in [BHL20], which used the theorem in [KM99] together with an intricate analysis, especially of completeness properties, of a space of positive operators admitting a finite trace to get nonpositive curvature.

Let Pos be the space of positive operators on a Hilbert space H. This is a convex cone in the Banach space of symmetric operators, and it has the corresponding Thompson metric:

$$d(p,q) = \sup_{v \in H, \|v\|=1} \left| \log \frac{(qv,v)}{(pv,v)} \right|.$$

Invertible bounded operators g act by isometry on this metric space via $p \mapsto gpg^*$. Unless one restricts to subspaces where there is a finite trace, this is not a CAT(0) space. On the other hand, as noted in particular in [CPR94], the fundamental Segal inequality

$$\|\exp(u+v)\| \leq \|\exp(u/2)\exp(v)\exp(u/2)\|$$

for symmetric operators u and v, can be seen as a weak form of nonpositive curvature more in Busemann's sense (with respect to a selection of geodesics). This means that the exponential map exp:Sym→Pos is distance preserving on lines from 0 and otherwise distance-increasing.

In finite dimensions, Lemmens has recently determined $\overline{Pos}$ [L21]. In infinite dimensions, the task to describe this compactification remains to be done, with some small steps done in [K22] in relation to the invariant subspace problem.

As for ergodic cocycles of invertible bounded linear operators, one has from Theorem 3:

Theorem 7. ([GK20]) *Let $v(n,\omega) = A(T^{n-1}\omega)...A(T\omega)A(\omega)$ be an integrable ergodic cocycle of bounded invertible linear operators of a Hilbert space. Denote the square of the positive part*

$$[v(n,\omega)] := v(n,\omega)^* v(n,\omega).$$

Then for a.e. ω there is a metric functional h_ω on Pos such that

$$\lim_{n\to\infty} -\frac{1}{n} h_\omega([v(n,\omega)]) = \lim_{n\to\infty} \frac{1}{n} \|\log[v(n,\omega)]\|.$$

Note that this statement by-passes the limitation mentioned above from local spectral theory. We can deduce an a priori weaker statement as follows, with a bit more information than in [GK20]. A state is a positive linear functional of norm 1 on certain types of algebras of operators. The space of states is compact in the weak*-topology.

Theorem 8. ([GK20]) *Let $v(n,\omega) = A(T^{n-1}\omega)...A(T\omega)A(\omega)$ be an integrable ergodic cocycle of bounded invertible linear operators of a Hilbert space. Denote the square of the positive part*

$$[v(n,\omega)] := v(n,\omega)^* v(n,\omega).$$

Then for a.e. ω there is a state f_ω on the space of bounded linear operators of the form,

$$f_\omega(A) = s(A\xi, \xi) + (1-s)\psi(A),$$

where ξ is a unit vector, ψ is a state on the algebra of all bounded linear operators vanishing on all compact operators and $0 \leq s \leq 1$, all depending on ω, such that

$$\lim_{n\to\infty} \frac{1}{n} \left| f_\omega(\log[v(n,\omega)]) \right| = \lim_{n\to\infty} \frac{1}{n} \left\| \log[v(n,\omega)] \right\|.$$

Proof. Let $u(n,\omega) := v(n,\omega)^*$ and hence $[v(n,\omega)] = u(n,\omega)I$ is the random orbit in Pos . Therefore, as explained in a previous section, we know that the distance

$$d(I, [v(n,\omega)])$$

is a subadditive cocycle. Moreover, thanks to the exact distance properties of the exponential map recalled above, also the distance inside Sym , which is the operator norm

$$\|\log[v(n,\omega)]\|$$

is subadditive. To see this in detail: by the distance preserving of exp on lines we have

$$d(I, [v(n,\omega)]) = \|\log[v(n,\omega)]\|$$

while

$$\|\log[v(n,\omega)] - \log[v(n+m,\omega)]\| \leq d([v(n,\omega)], [v(n+m,\omega)]) = d(I, [v(m, T^n\omega)])$$
$$= \|\log[v(m, T^n\omega)]\|.$$

Thus by the triangle inequality,

$$\|\log[v(n+m,\omega)]\| \leq \|\log[v(n,\omega)]\| + \|\log[v(m, T^n\omega)]\|.$$

Let $y_n := \log[v(n,\omega)]$ and $\epsilon_n \searrow 0$. Since y_n is a self-adjoint operator we can find a unit vector ξ_n such that

$$|(y_n\xi_n, \xi_n)| > \|y_n\| - \epsilon_n.$$

Let $f_n(A) = (A\xi_n, \xi_n)$ be the corresponding linear functional, which in other words is a vector state. We assume that the right hand side is strictly positive, that is,

$$\tau := \lim_{n\to\infty} \frac{1}{n} \|\log[v(n,\omega)]\| > 0,$$

otherwise there is nothing to prove. From the main subadditive ergodic result in [GK20] we have for almost every ω, a sequence $n_i \to \infty$ and sequence $\delta_l \to 0$ such that for every i and every $l \leq n_i$,

$$\|\log[v(n_i,\omega)]\| - \left\|\log[v(n_i - l, T^l\omega)]\right\| \geq (\tau - \delta_l)l.$$

In view of this, for any $l \leq n_i$,

$$\begin{aligned}\|y_l\| \geq |f_{n_i}(y_l)| = |f_{n_i}(y_{n_i} + y_l - y_{n_i})| = |f_{n_i}(y_{n_i}) - f_{n_i}(y_{n_i} - y_l)| \geq |f_{n_i}(y_{n_i})| - |f_{n_i}(y_{n_i} - y_l)| \\ \geq \|y_{n_i}\| - \epsilon_{n_i} - \|y_{n_i} - y_l\| \geq \|\log[v(n_i,\omega)]\| - \left\|\log[v(n_i - l, T^l\omega)]\right\| - \epsilon_{n_i} \\ \geq (\tau - \delta_l)l - \epsilon_{n_i}.\end{aligned}$$

By weak*-compactness letting $i \to \infty$ there is a state $f = f^\omega$ for which

$$\lim_{l\to\infty} \frac{1}{n} |f(y_l)| = \tau,$$

as desired. By Glimm's theorem [Gl60] this state, being a limit of vector states, must be of the form

$$f(A) = s(A\xi, \xi) + (1 - s)\psi(A),$$

where ξ is a unit vector, ψ a state on the algebra of all bounded linear operators on the Hilbert space vanishing on all compact operators and $0 \leq s \leq 1$. □

Remark. When the cocycle is composed of compact operators then s must be 1 and f_ω is a pure vector state, which then provides a result pointing in the direction of Ruelle's theorem.

5 Diffeomorphisms

In the introduction topological Lyapunov exponents for surface homeomorphisms were explained. Thurston's powerful measured foliation theory is presumably difficult to generalize to dimensions greater than two (and so far cannot treat the case of products of random homeomorphisms). The metric perspective can on the other hand more easily be generalized. For example, Ebin's Riemannian manifold of Riemannian metrics on a compact manifold is one possibility [Eb68]. The diffeomorphisms act by isometry. This is the replacement for the Teichmüller spaces. There are two variants, one general and one restricted to Riemannian metrics sharing the same volume form, and we would then consider volume preserving diffeomorphisms. These spaces have nonpositive curvature but not always complete.

Here is a related metric taken from [AK22], and which perhaps has not been considered before. Let M be a compact submanifold of a finite dimensional vector space equipped with a norm $\|\cdot\|$. Consider the following weak metric (which can be symmetrized if needed) on the set X of distance function functions bi-Lipschitz equivalent to $d_0(x, y) = \|x - y\|$:

$$D(d_1, d_2) = \log \sup_{x \neq y} \frac{d_2(x, y)}{d_1(x, y)}.$$

If $T : M \to M$ is a diffeomorphism, it will preserve D-distances, considered as map $(T^*d)(x, y) := d(Tx, Ty)$ since it just permutes the underlying set M. Note that T^* is an adjoint type of map, it reverses the order of composition (and if it is desired to keep the orientation we could instead use the inverse if it exists).

The following was shown in [AK22], with one ingredient being the main results of [GK20]:

Theorem 9. (Existence of a point with maximal stretch [AK22]) *Let*

$$v(n,\omega) = A(T^{n-1}\omega)...A(T\omega)A(\omega)$$

be an integrable ergodic cocycle of diffeomorphisms of M. Then there is a number λ such that

$$\lim_{n\to\infty}\left(\sup_{x\neq y}\frac{\|v(n,\omega)x - v(n,\omega)y\|}{\|x-y\|}\right)^{1/n} = e^{\lambda}.$$

In the case that $\lambda > 0$ then there exists a point $z \in M$ and a sequence $w_i = (x_i, y_i) \in \{(x,y) \in M \times M : x \neq y\}$ such that $w_i \to (z,z)$ and for any $\epsilon > 0$ there is $N > 0$ such that for all $n > N$

$$\frac{\|v(n,\omega)x_i - v(n,\omega)y_i\|}{\|x_i - y_i\|} \geq e^{(\lambda-\epsilon)n}$$

for all i sufficiently large for a fixed n.

In words, it means that given a cocycle there is a.e. a random point z such that nearby this point the cocycle is stretching at a near maximal rate.

Another possibility of measuring distances in 1-dimensional dynamics is the total variation of the logarithm of the derivative as in [EBN22]. This can be extended to a weak metric on diffeomorphisms groups on compact manifolds M via

$$d(f,g) = \sup_{x\in M}\left|\log\frac{|J_g(x)|}{|J_f(x)|}\right|$$

where $|J_f|$ is the determinant of the Jacobian derivative. See [AK22] for further definitions of this type and their use.

6 Applications of metrics and ergodic theorems in deep learning

In [AK22] Avelin and I introduced new geometric frameworks to the theory of neural networks, that also enabled the application of the noncommutative ergodic theorem. More specifically, we suggested several metrics on the data set making various choices of layer maps nonexpansive. This includes the most standard choices of activation functions (those mentioned above), and with positive, unitary or invertible features for the weights. Since the composition of nonexpansive maps remains nonexpansive, already this guarantees some regularity and absence of wild fluctuations.

- *In several standard models of neural networks it is possible to find semi-invariant metrics. This may help to explain phenomena that have been observed emprically, such as a certain stability that ensures good generalization as opposed to overfitting.*

In addition, in view of the noncommutative ergodic theorem above, when the layer maps are selected at random and the number of layers is large, their compositions are close to being constant functions in some cases, see for example Theorem 10 below. As Dherin and his colleagues from Google and DeepMind informed us this fits very well their theory aimed at explaining why deep learning generalize well and do not overfit data, instead the trained network has a bias towards simple functions [DMRB22]. They measure simplicity with their Geometric Complexity notion inspired by the Dirichlet energy. Constant maps have zero complexity. And as the authors argue in [BD21] and [DMRB22] initializations that deliver near-constant functions are an advantage.

- *When the noncommutative ergodic theorem, Theorem 3, is applicable to the neural network, random initilialization gives near-constant maps. According to [BD21, DMRB22] this is something observed in practice that may be highly desirable by contributing to the simple nature of the functions that the stochastic gradient method tends to find.*

Thus one can say that from this point of view it seems good to choose a network architecture for which Theorem 3 applies. Such situations are discussed in more detail in [AK22].

To illustrate the above points, here is a sample corollary of Theorem 3: maps used in a popular layer model called ResNets, see [HZRS16], are treated in the following result:

Theorem 10. ([AK22]) *Let $X = \mathbb{R}^d$ with the standard scalar product. Consider the layer maps $T(x) = W^T\sigma(Wx + b)$, with b a general vector and W having operator norm at most 1, and the activation function being either ReLU, TanH or the sigmoid function. When such layer maps are selected i.i.d. under a finite moment condition, it holds a.s. as $n \to \infty$ and any $x_0 \in X$ that there is a (random) vector v such that*

$$\frac{1}{n}T_1T_2...T_nx_0 \to v.$$

The vector v does not depend on the input x_0 thus the limiting map is a (random) constant function, and in the case of a large, but finite, fixed number n of layers, the composed function should be nearly constant.

The infinite-width limit as been more studied than the infinite-depth case here discussed. For investigations about the case when both the width and he depth go to infinity, see [HaN20, LNR22].

References

[AGW09] Akian, Marianne; Gaubert, Stéphane; Walsh, Cormac The max-plus Martin boundary. Doc. Math. 14 (2009), 195–240.

[AK22] Avelin, B. Karlsson, A. Deep limits and cut-off phenomena for neural networks, J. Mach. Learn. Res. 23 (191) (2022) 29 pp

[BD21] Barrett, David G.T.; Dherin, Benoit, Implicit Gradient Regularization, ICLR2021

[BT22] Basteri, A.; Trevisan, D. Quantitative Gaussian Approximation of Randomly Initialized Deep Neural Networks, arXiv:2203.07379 [cs.LG]

[Be54] Bellman, Richard, Limit theorems for non-commutative operations. I. Duke Math. J. 21 (1954), 491–500.

[BQ16] Benoist, Yves; Quint, Jean-François Random walks on reductive groups. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics, 62. Springer, Cham, 2016. xi+323 pp.

[BHS21] Blath, Jochen; Hermann, Felix; Slowik, Martin, A branching process model for dormancy and seed banks in randomly fluctuating environments. J. Math. Biol. 83 (2021), no. 2, Paper No. 17, 40 pp.

[Bl16] Blumenthal, Alex A volume-based approach to the multiplicative ergodic theorem on Banach spaces. Discrete Contin. Dyn. Syst. 36 (2016), no. 5, 2377–2403.

[BHL20] Bowen, Lewis; Hayes, Ben; Lin, Yuqing, A multiplicative ergodic theorem for von Neumann algebra valued cocycles, Commun. Math. Phys. (2021), arXiv:2006.13293.

[CD20] Cantat, Serge; Dujardin, Romain, Random dynamics on real and complex projective surfaces, arXiv:2006.04394

[CPR94] Corach, G.; Porta, H.; Recht, L. Convexity of the geodesic distance on spaces of positive operators. Illinois J. Math. 38 (1994), no. 1, 87-94.

[CDS09] Crauel, H.; Doan, T. S.; Siegmund, S. Difference equations with random delay. J. Difference Equ. Appl. 15 (2009), no. 7, 627–647.

[CPV93] Crisanti, A.; Paladin, G.; Vulpiani, A. Products of random matrices in statistical physics. With a foreword by Giorgio Parisi. Springer Series in Solid-State Sciences, 104. Springer-Verlag, Berlin, 1993. xiv+166 pp.

[DMRB22] Dherin, Benoit; Munn, Michael; Rosca, Mihaela, Barrett, David G.T., Why neural networks find simple solutions: the many regularizers of geometric complexity, NeurIPS (2022)

[Dy53] Dyson,F.J. The dynamics of a disordered linear chain. Phys. Rev. 92, (1953) 1331–1338

[Eb68] Ebin, David G. On the space of Riemannian metrics. Bull. Amer. Math. Soc. 74 1968 1001-1003.

[ER85] Eckmann, J.-P.; Ruelle, D. Ergodic theory of chaos and strange attractors. Rev. Modern Phys. 57 (1985), no. 3, part 1, 617–656.

[Er10] Erschler, Anna Poisson-Furstenberg boundaries, large-scale geometry and growth of groups. Proceedings of the International Congress of Mathematicians. Volume II, 681–704, Hindustan Book Agency, New Delhi, 2010.

[EBN22] Eynard-Bontemps, Hélène, Navas, Andrés, (Arc-)connectedness for the space of smooth ℤ^d-actions on 1-dimensional manifolds, https://arxiv.org/abs/2103.06940

[FLP79] Fathi, A, Laudenbach, F, and Poénaru, V. *Travaux de Thurston sur les surfaces.* Astérisque, 66-67. Société Mathématique de France, Paris, 1979. 284 pp.

[Fu63] Furstenberg, Harry Noncommuting random products. Trans. Amer. Math. Soc. 108 1963 377–428.

[Gl60] Glimm, James, A Stone-Weierstrass theorem for C∗-algebras, Ann. of Math. 72 (1960), 216–244.

[GTQ15] González-Tokman, Cecilia; Quas, Anthony A concise proof of the multiplicative ergodic theorem on Banach spaces. J. Mod. Dyn. 9 (2015), 237–255.

[GT18] González-Tokman, Cecilia Multiplicative ergodic theorems for transfer operators: towards the identification and analysis of coherent structures in non-autonomous dynamical systems. Contributions of Mexican mathematicians abroad in pure and applied mathematics, 31–52, Contemp. Math., 709, Aportaciones Mat., Amer. Math. Soc., Providence, RI, 2018.

[GK20] Gouëzel, S; Karlsson, A. Subadditive and multiplicative ergodic theorems, J. Eur. Math. Soc. 22 (2020), no. 6, 1893–1915.

[Gu80] Guivarc'h, Y. Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire. (French) Conference on Random Walks (Kleebach, 1979) (French), pp. 47–98, 3, Astérisque, 74, Soc. Math. France, Paris, 1980.

[GuW22] Gutiérrez, Armando W.; Walsh, Cormac; Firm non-expansive mappings in weak metric spaces. Arch. Math. (Basel) 119 (2022), no. 4, 389–400.

[Gr81] Gromov, M. Hyperbolic manifolds, groups and actions. Riemann surfaces and related topics: Pro- ceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, N.Y., 1978), pp. 183Ð213, Ann. of Math. Stud., 97, Princeton Univ. Press, Princeton, N.J., 1981.

[HaN20] Hanin, Boris; Nica, Mihai Products of many large random matrices and gradients in deep neural networks. Comm. Math. Phys. 376 (2020), no. 1, 287–322.

[Ha21] Hanin, B. Random Neural Networks in the Infinite Width Limit as Gaussian Processes, arXiv:2107.01562 [math.PR]

[HZRS16] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[H16] Horbez, Camille, The horoboundary of outer space, and growth under random automorphisms. Ann. Sci. Éc. Norm. Supér. (4) 49 (2016), no. 5, 1075–1123.

[IS96] Iasemidis, L.D. Sackellares J.C. REVIEW: Chaos Theory and Epilepsy, The Neuroscientist, (1996) 2 issue 2, page(s): 118-126.

[KM96] Kaimanovich, Vadim A.; Masur, Howard The Poisson boundary of the mapping class group. Invent. Math. 125 (1996), no. 2, 221–264.

[Ka00] Kaimanovich, Vadim A. The Poisson formula for groups with hyperbolic properties. Ann. of Math. (2) 152 (2000), no. 3, 659–692.

[KS19] Kalinin, Boris; Sadovskaya, Victoria Periodic approximation of Lyapunov exponents for Banach cocycles. Ergodic Theory Dynam. Systems 39 (2019), no. 3, 689–706.

[K04] Karlsson, Anders, Linear rate of escape and convergence in directions, In: Proceedings of a Workshop at the Schrödinger Institute, Vienna 2001, (Ed. by V.A. Kaimanovich, in collab. with K. Schmidt, W. Woess) de Gruyter, 2004.

[K14] Karlsson, Anders, Two extensions of Thurston's spectral theorem for surface diffeomorphisms. Bull. Lond. Math. Soc. 46 (2014), no. 2, 217–226.

[K18] Karlsson, Anders, Elements of a metric spectral theory, In: Dynamics, Geometry and Number Theory: the Impact of Margulis on Modern Mathematics The University of Chicago Press, 2022

[K21] Karlsson, Anders, From linear to metric functional analysis. Proc. Natl. Acad. Sci. USA 118 (2021), no. 28, Paper No. e2107069118, 5 pp.

[K22] Karlsson, Anders, A metric fixed point theorem and some of its applications, preprint 2022.

[KL06] Karlsson, Anders; Ledrappier, François On laws of large numbers for random walks. Ann. Probab. 34 (2006), no. 5, 1693–1706.

[KL07] Karlsson, Anders; Ledrappier, François Linear drift and Poisson boundary for random walks. Pure Appl. Math. Q. 3 (2007), no. 4, Special Issue: In honor of Grigory Margulis. Part 1, 1027–1036.

[KL07b] Karlsson, Anders; Ledrappier, François Propriété de Liouville et vitesse de fuite du mouvement brownien. C. R. Math. Acad. Sci. Paris 344 (2007), no. 11, 685–690.

[KM99] Karlsson, Anders; Margulis, Gregory A., A multiplicative ergodic theorem and nonpositively curved spaces. Comm. Math. Phys. 208 (1999), no. 1, 107–123.

[KMo08] Karlsson, Anders, Monod, Nicolas, Strong law of large numbers with concave moments, unpublished note 2008

[Ki68] Kingman, J. F. C. The ergodic theory of subadditive stochastic processes. J. Roy. Statist. Soc. Ser. B 30 1968 499–510.

[KoN81] Kohlberg, Elon; Neyman, Abraham Asymptotic behavior of nonexpansive mappings in normed linear spaces. Israel J. Math. 38 (1981), no. 4, 269–275.

[L21] Lemmens, Bas, Horofunction compactifications of symmetric cones under Finsler distances, arXiv:2111.12468.

[LRT12] Lenzhen, Anna; Rafi, Kasra; Tao, Jing Bounded combinatorics and the Lipschitz metric on Teichmüller space. Geom. Dedicata 159 (2012), 353–371.

[LNR22] Li, Mufan Bill; Nica, Mihai; Roy, Daniel M. The Neural Covariance SDE: Shaped Infinite Depth-and-Width Networks at Initialization, NeurIPS 2022

[LL10] Lian, Zeng; Lu, Kening Lyapunov exponents and invariant manifolds for random dynamical systems in a Banach space. Mem. Amer. Math. Soc. 206 (2010), no. 967, vi+106 pp.

[LY12] Lian, Zeng; Young, Lai-Sang Lyapunov exponents, periodic orbits, and horseshoes for semiflows on Hilbert spaces. J. Amer. Math. Soc. 25 (2012), no. 3, 637–665.

[MT18] Maher, Joseph, Tiozzo, Giulio, Random walks on weakly hyperbolic groups, J. Reine Angew. Math. 742 (2018), 187–239.

[M75] Margulis, G. A. Discrete groups of motions of manifolds of nonpositive curvature. (Russian) Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974), Vol. 2, pp. 21–34. Canad. Math. Congress, Montreal, Que., 1975

[MN20] Martirosyan, D.; Nersesyan, V. Multiplicative ergodic theorem for a non-irreducible random dynamical system. J. Differential Equations 268 (2020), no. 7, 3564–3598.

[Ma21] Masai, Hidetoshi, Compactification and distance on Teichmüller space via renormalized volume, arXiv:2108.06059.

[MP43] McCulloch, W and W. Pitts. A Logical Calculus of Ideas Immanent in Nervous Activity, Bull. Math. Biophys., 5 (1943) 115-133.

[NBYS22] Narkhede, M.V.; Bartakke, P.P.; and Sutaone, M.S. A review on weight initialization strategies for neural networks, Artificial intelligence review 55 (2022), no. 1, 291–322.

[Ne12] Neal, R.M. Bayesian learning for neural networks, volume 118. Springer Science & Business Media, 2012.

[Neu19] Neusser, Klaus, Time–varying rational expectations models. J. Econom. Dynam. Control 107 (2019), 103731, 17 pp.

[O68] Oseledec, V. I. A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. (Russian) Trudy Moskov. Mat. Obšč. 19 1968 179–210.

[PT09] Papadopoulos, Athanase; Troyanov, Marc Weak Finsler structures and the Funk weak metric. Math. Proc. Cambridge Philos. Soc. 147 (2009), no. 2, 419–437.

[Ri43] Ribeiro, Hugo Sur les espaces à métrique faible. Portugal. Math. 4 (1943), 21–40.

[Ro58] Frank Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., Psychological review 65 (1958), no. 6, 386.

[Ru82] Ruelle, David Characteristic exponents and invariant manifolds in Hilbert space. Ann. of Math. (2) 115 (1982), no. 2, 243–290.

[Sa21] Sampaio, Luís Miguel, Regularity of the drift for random walks in groups acting on Gromov hyperbolic spaces, https://arxiv.org/abs/2109.13060

[Sc06] Schaumlöffel, Kay-Uwe, Multiplicative ergodic theorems in infinite dimensions, In: Infinite Dimensional Random Dynamical Systems, Lyapunov Exponents, Springer Lecture Notes in Mathematics, volume 1486, (2006) pp 187–195

[S57] Schmidt,H. Disordered one-dimensional crystals. Phys.Rev.105,425(1957) 425–441

[Se20] Sejnowski, T.J. The unreasonable effectiveness of deep learning in artificial intelligence, Proc. Nat. Acad. Sci. U.S.A. December 1, 2020 117 (48) 30033-30038.

[SHK14] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I. and Salakhutdinov, R.. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1):1929–1958, 2014.

[Sz01] Szpankowski, Wojciech Average case analysis of algorithms on sequences. With a foreword by Philippe Flajolet. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2001. xxii+551 pp.

[Th87] Thieullen, P. Fibrés dynamiques asymptotiquement compacts. Exposants de Lyapounov. Entropie. Dimension. Ann. Inst. H. Poincaré Anal. Non Linéaire 4 (1987), no. 1, 49–97

[T88] Thurston, William P. On the geometry and dynamics of diffeomorphisms of surfaces. Bull. Amer. Math. Soc. (N.S.) 19 (1988), no. 2, 417–431.

[W14] Walsh, Cormac The horoboundary and isometry group of Thurston's Lipschitz metric. Handbook of Teichmüller theory. Vol. IV, 327–353, IRMA Lect. Math. Theor. Phys., 19, Eur. Math. Soc., Zürich, 2014.

[Zh22] Zheng, Tianyi, Asymptotic behaviors of random walks on countable groups. To appear in the Proceedings of ICM 2022.

[ZC21] Zou, Rui; Cao, Yongluo Livšic theorems for Banach cocycles: existence and regularity. J. Funct. Anal. 280 (2021), no. 5, Paper No. 108889, 37 pp.

Section de mathématiques, Université de Genève, Case postale 64, 1211 Genève, Switzerland; Mathematics department, Uppsala University, Box 256, 751 05 Uppsala, Sweden.