Tightness of Random Walks in Infinite Dimensional Spaces and Manifolds*

Andrea C. G. Mennucci †

Abstract

In this paper we study random walks $\mathfrak{X}^\tau$; these are processes taking values in $C(\mathbb{R}^+; S)$, where $\mathbb{R}^+ = [0, \infty)$. These random walks are defined at discrete times $t \in \tau = \{t_0 = 0 < t_1 < t_2 \ldots\}$ and then interpolated for t between t_i, t_{i+1}.

The main objective is to prove tightness for the family of all $\mathfrak{X}^\tau$; by Prokhorov's Theorem, this implies that the sequence has limit points that are random functions in $C(\mathbb{R}^+; S)$.

We will provide results in three cases: $S = H$ a (possibly infinite dimensional) separable Hilbert Space; S a manifold embedded in H; and then the particular case when S is the Stiefel Manifold.

These results are motivated by problems in Probability Theory and in Shape Theory, and in particular some models of manifolds of planar immersed curves.

Keywords: infinite dimensional manifold ; Riemannian manifold ; Stiefel Manifold ; Hilbert space ; Tight family ; Random Walk ; Wiener Process ; Donsker's Theorem ; Brownian Motion ; Stochastic completeness ; .

MSC2020 subject classifications: Primary 60J99 ; 60Fxx ; 60B10 ; 58J65 ; 58D20 ; 46T05 ; , Secondary 49Q10 ; 58C35 ; 60J70 ; 60B11 ; .

1 Introduction

Let H be a separable Hilbert space, M a manifold embedded in H, possibly infinite dimensional. Let $S = H$ or $S = M$.

In this paper we study random walks $\mathfrak{X}^\tau$; these are processes, *i.e.* random functions, taking values in the Frechét space $C(\mathbb{R}^+; S)$, where $\mathbb{R}^+ = [0, \infty)$.

These random walks are defined at discrete times $t \in \tau = \{t_0 = 0 < t_1 < t_2 \ldots\}$ and then interpolated for t between t_i, t_{i+1}.

---

*This project was supported by the `RICBASE_2022_MENNUCCI` funding.

†Scuola Normale Superiore, Pisa, Italy
E-mail: `andrea.mennucci@sns.it` ORCID 0000-0002-4302-3275

The main objective is to prove tightness for the family of all $\mathfrak{X}^\tau$; by Prokhorov's Theorem, this implies that the sequence has limit points (in the narrow sense) that are random functions in $C(\mathbb{R}^+; S)$.

We will provide results in three cases: $S = H$ in Section 5, $S = M$ in Section 6 and then the particular case when S is the Stiefel Manifold in Section 7.

These results are motivated by problems in Probability Theory (in particular Stochastic Processes in manifolds) and in Shape Theory, that are detailed in Section 2. We are particularly concerned by the infinite dimensional case, since, as discussed in Section 2.2.3, in this case many standard methods cannot be applied.

We will use some definition and results from Probability Theory; since there may be ambiguity in some definitions, and some results are not easily found in the needed generality, then we have written a compendium in [18]; it is available as supplemental material.

1.1 Random walk

Here we briefly define the random walks $\mathfrak{X}^\tau_t$ that we will study (more details will be in Section 3.4).

Let $\mathbb{R}^+ = [0,\infty)$; let H, U be separable Hilbert spaces. Let

$$\tau = \{t_0 = 0 < t_1 < t_2 \ldots\} \subset \mathbb{Q} \quad .$$

We will need a source of random noise: for $t \in \mathbb{R}^+$, a family of i.i.d. r.v. Y_t taking values in U, each with law γ. We will need a Borel map

$$D = D(x, v, t, s) : H \times U \times (\mathbb{R}^+)^2 \to H$$

continuous in s and such that $D(x, v, t, 0) = x$. We fix $X^\tau_0 = \mathfrak{X}_0$ a random variable, and we define recursively

$$X^\tau_{(n+1)} = X^\tau_n + D\left(X^\tau_n\ ,\ Y_{t_n}\ ,\ t_n\ ,\ (t_{n+1} - t_n)\right) \quad . \tag{1.1}$$

Then we interpolate using

$$\mathfrak{X}^\tau_t = X^\tau_n + D\left(X^\tau_n\ ,\ Y_{t_n}\ ,\ t_n\ ,\ (t - t_n)\right) \tag{1.2}$$

for $t_n \le t \le t_{(n+1)}$; so each trajectory $t \mapsto \mathfrak{X}^\tau_t(\omega, t)$ is continuous; hence each $\mathfrak{X}^\tau$ is a r.v. taking value in $C(\mathbb{R}^+; H)$, the Frechét space of continuous functions $x : \mathbb{R}^+ \to H$.

Since U is used only in the second argument of D, and H, U are isomorphic, we can decide that $H = U$ with no loss of generality.

2 Motivation

2.1 Wiener Process, Donsker's Theorem

We recall this standard result.

Theorem 2.1. *Let $Y_1, Y_2, Y_3, \ldots$ be a sequence of i.i.d. real random variables with mean 0 and variance 1. Let*

$$S_n = \sum_{i=1}^n Y_i$$

We rescale and extend the process to continuous time $t \in [0, 1]$. Define

$$W^n(t) = \frac{S_j}{\sqrt{n}}, \qquad t = j/n$$

and then linearly interpolate

$$W^n(t) = (1-s)\frac{S_j}{\sqrt{n}} + s\frac{S_{j+1}}{\sqrt{n}} = \frac{S_j + sY_j}{\sqrt{n}} \tag{2.1}$$

for $j/n \le t \le (j+1)/n$ *and*

$$t = (1-s)j/n + s(j+1)/n \ \text{ i.e. } \ s = nt - j \quad ;$$

so W^n *is a random variable taking values in* $C([0,1])$*. Then sequence of random function* $(W^n)_n$ *converges narrowly to a random function* W *on* $C([0,1])$*, as* $n \to \infty$*; this* W *is the standard* Wiener Process.[1]

Theorem 9.1 in [5] uses the above as a method to *define* Wiener Process; by Kolmogoroff Theorem, W has a version where almost all paths are continuous. Other sources call the above result *Donsker's Theorem*. The proof may be found in Theorem 9.1 [2] in [5]; the proof is obtained in two steps:

1. show that the family W^n is tight: by Prokhorov's theorem [3], then it admits narrow limits in $C([0,1])$ as $n \to \infty$;
2. show that there is an unique narrow limit W: by a standard argument this implies that $W^n \to W$ narrowly. Indeed it is easy to argue that any narrow limit W has independent increments and the law of $W_t - W_s$ is $N(0, t-s)$ (by CLT): this uniquely identifies the Wiener Process.

The above construction of W^n in Theorem 2.1 is a special case of the random walk $\mathfrak{X}^\tau_t$ where $\tau = \{i/n : j \in \mathbb{N}\}$, $\mathfrak{X}_0 = 0$, $H = \mathbb{R}$ and $D(x,v,t,s) = \sqrt{s}\ v$. (There is a slightly different interpolation method: *cf* 2.2).

So we can imagine a form of *Donsker's theorem*, for random walks with variable time step and taking value in infinite dimensional Hilbert spaces, or manifolds; to be proven in this way:

1. show that the family $\mathfrak{X}^\tau_t$ is tight;
2. show that there is an unique the narrow limit $\mathfrak{X}$: by a standard argument this implies that $\mathfrak{X}^\tau \to \mathfrak{X}$ narrowly.

Hence one purpose of this paper is to provide a tool for the first step: this is Theorem 5.5.

Remark 2.2. If we would like to apply the linear interpolation (used in (2.1)) to our random walk then we would replace (1.2) with

$$\mathfrak{X}^\tau_t = sX^\tau_{n+1} + (1-s)X^\tau_n$$

where $s \in [0,1]$ satisfies

$$t = t_{n+1}s + t_n(1-s)$$

that is

$$s = \frac{t - t_n}{t_{n+1} - t_n} \quad .$$

We prefer the former (1.2) since it provides some technical simplifications: see Remark 3.13. Conversely, if we set $H = \mathbb{R}$, $\mathfrak{X}_0 = 0$ and $D(x,v,t,s) = \sqrt{s}v - x$ then, to be able to state that $\mathfrak{X}^\tau_t = W^n(t)$, we should define the interpolation as

$$W^n(t) = \frac{S_j + \sqrt{s}Y_j}{\sqrt{n}} \tag{2.2}$$

and this is not the definition in [5]. Note that this interpolation (2.2) has the benefit that $W^n(t) \sim N(0,t)$.

---

[1] *Wiener Process* is also known as *Brownian Motion* in some texts, as [16] or Chapter 12 in [7].

[2] Theorem 9.1 uses the linear approximations

[3] See Theorem 4.13 in [18].

At the same time, the results in this paper are valid for different interpolations. We add an important remark.

Proposition 2.3. *There is no choice of common probability space where to define Y_t and W_t and such that the approximating terms W^n defined above in* (2.1) *would converge to W in probability.*

(The proof is in Appendix B).

2.2 Manifolds

2.2.1 Finite Dimensional Manifolds

The theory of Stochastic Differential Equations in finite dimensional Riemannian Manifolds M is well developed; see *e.g.* [16]. In particular, there are multiple equivalent definitions of *Brownian Motion*; each based on different principles,

- stochastic differential equations in local charts,
- development of euclidean Brownian Motion (Example 2.6.8 in [16]),
- the heat equation and its transition probabilities;

but all leading to the same ultimate definition: see Proposition 3.2.1 in [16].
We define this concept as in Section 4.2 in [16].

Definition 2.4 (Stochastic completeness)**.** *Consider a non-compact connected manifold M, and let ∞ be the point added by the Alexandroff compactification (the one-point topological compactification). For any continuous path $x : \mathbb{R}^+ \to M \cup \{\infty\}$ let*

$$e = e(x) = \sup\{t \geq 0 : \forall s, 0 \leq s < t, x(s) \in M\} = \sup\{t \geq 0 : x(t) \in M\}$$

be the first time t such that $x(t) = \infty$; we agree that $x(s) = \infty$ for $s \geq e$. Suppose that $\mathfrak{X}_t$ is Brownian Motion*, whose paths are continuous in $M \cup \{\infty\}$. A manifold is called* stochastically complete *if $e(\mathfrak{X}_t)$ is infinite almost surely:*

$$\mathbb{P}\{e(\mathfrak{X}_t) = \infty\} = 1 \quad .$$

A thorough discussion of this problematic may be found in [14]. There is an important problem: even if the manifold is complete, it may fail to be *stochastically complete*. ([14] attributes the first such example to [1]).
There are many properties of M that ensure that the manifold is *stochastically complete*, such as as: volume growth of geodesic balls, isoperimetric inequalities, conservation of mass in the heat equation, curvature bounds, etc.; see [14]. (Indeed our Theorem 6.3 requires a kind of curvature bound).

2.2.2 Radial process

For $\mathfrak{X}$ a process taking values in M and with continuous paths, the *radial process* is

$$r_{\mathfrak{X}}(t) = d(x_0, \mathfrak{X}_t)$$

where $x_0 \in M$ is a fixed point and we agree that $d(x_0, \infty) = +\infty$. The radial process satisfies an SDE, and its evolution can be bounded by bounds on the curvature (see Section 3.5 in [16]); since

$$\{e(\mathfrak{X}_t) > T\} = \{\forall s \in [0, T] : r_{\mathfrak{X}}(t) < \infty\}$$

this can be used to prove stochastic completeness.

Proposition 2.5. *Let $x_0 \in M$ be a fixed point. Stochastic completeness is equivalent to*

$$\forall \varepsilon > 0, \forall T > 0, \exists R > 0 \text{ such that } \mathbb{P}\{\exists t \in [0,T], d(x_0, \mathfrak{X}_t) > R\} < \varepsilon \quad . \tag{2.3}$$

Since (by Hopf–Rinow theorem) a closed set is compact iff it is bounded, then (2.3) is equivalent to *tightness*, in this sense

$$\forall \varepsilon > 0, \forall T > 0, \exists C \subseteq M \text{ compact, such that } \mathbb{P}\{\exists t \in [0,T], \mathfrak{X}_t \notin C\} < \varepsilon \quad . \tag{2.4}$$

2.2.3 Infinite Dimensional Manifolds

When the Riemannian Manifold M is infinite dimensional, though, we immediately identify some obstacles.

- When the manifold M has dimension N, we have an important property: each tangent space $T_x M$ is isomorphic to $\mathbb{R}^N$; hence there is a canonical choice of Gaussian measure $N(0, \mathbb{I})$ on each one. This is, in a sense, the *"white noise"* that is driving the Brownian Motion.

  When the manifold M is modeled on a infinite dimensional Hilbert Space H then there is no Gaussian measure in H that is rotationally invariant (actually, rotations of a Gaussian measure $N(0, Q)$ tend to be mutually singular, as explained in [6]). So we will need to decide what "white noise" we will use.
- While the *heat equation* can be defined in H, an approach using this tool would have to deal with some technical difficulties; for example, the heat equation is not *Feller*, that is, it does not regularizes the initial data.

  Moreover, usually the transition probabilities of the heat kernel are used; these transition probabilities are expressed as densities with respect to the volume form; but an infinite dimensional Riemannian Manifold does not have a volume form that may be used as a reference measure.
- The Hopf–Rinow theorem is false, closed bounded sets are not necessarily compact.
- The one point compactification is not useful, since any non-empty open set contains a sequence such that $x_n \to \infty$
- The radial process is not useful, since there may be examples of complete Riemannian Manifolds where the trajectories of the Brownian Motion are bounded, but each of them would have $\mathfrak{X}_t \to \infty$ in finite time. (A key point to build such an example may be [3]).

Prokhorov's theorem, on the other end, is valid in any separable metric space (regardless of "dimension"): so a concept of *tightness* similar to (2.4) will be the key element for Theorem 6.3.

2.3 Stiefel Manifolds

The results in this paper will be valid when M is a Stiefel Manifold.

Classically, the Stiefel manifold $\mathbf{St}(p, \mathbb{R}^n)$ is defined as the set of all frames composed of p orthonormal vectors in $\mathbb{R}^n$ (with $1 \leq p \leq n$); those frames are represented as $n \times p$ matrices. Geodesics in Stiefel manifolds $\mathbf{St}(p, \mathbb{R}^n)$ are known to have closed form solutions as demonstrated by Edelman et al. [12] in Section 2.2.2 .[4] This property extends to infinite dimensional manifolds, as will be explained in Section 7.3.1.

Let $p \geq 1$ and let V be a Hilbert space with $\dim(V) \geq 2p$ (possibly infinite dimensional). Let $H = V^p$, we write

$$x \in H \quad , \quad x = (x_1, \dots x_p)$$

[4][12] credits a personal communication by R. Lippert for the final closed form formula (7.3).

and

$$\|x\|_H = \sqrt{\sum_{i=1}^{p} \|x_i\|_V^2}$$

as usual (and similarly for scalar products). By analogy to the finite dimensional space, we will call *columns* the p vectors x_i that compose x. We define

$$\pi_i : H \to V \quad , \quad \pi_i(x) = x_i \quad .$$

Definition 2.6. We define $\mathbf{St}(p, V)$ as the manifold of $x \in H$ such that

$$\langle x_i, x_j \rangle_V = \delta_{i,j}$$

and $\mathbf{St}(p, V)$ is an embedded manifold in H of codimension $p(p+1)/2$.

The sphere is the special case $\mathbf{St}(1, V)$. The geometry of Stiefel Manifolds is pretty well understood [12, 15]. See Section 7.3.1 for details.

2.4 Shape Theory, Curves

"Shapes" appear in two broad categories of applications:

- *shape optimization*, where we want to find the best shape according to a criterion;
- *shape analysis*, where we study families of shapes for purposes of statistics, (automatic) cataloging, probabilistic modeling, etc.

Shape theory is central in computer vision because shapes partially characterize objects in images. We focus on the specific case where *shapes* are represented by smoothly immersed planar curves; this is a widely studied subject, see [17] and references therein. In this case, it should be noted that the *shape space* classically used in shape optimization is more precisely identified as *the space of embedded curves, up to a choice of parameterization*, whereas in shape analysis the space is usually identified as *the space of embedded curves, up to rotation, translation, scaling and reparameterization*. We will not address this issue in this paper.

There are various reasons why it is useful to model the space of curves as a Riemannian manifold.

- In the past methods for *shape optimization* were devised that would find the solution by using appropriate *gradient flows*. Calling the minimizing flows *gradient flows*, however, implies a certain Riemannian metric on the space of curves.
- Modeling the space of curves as a Riemannian manifold has also obvious benefits in shape analysis: indeed the distance between curves can be used for clustering, the geodesic can be used to define the average of two shapes, and so on.

We concentrate on two models of "Riemannian manifolds of curves", where we disregard translation and scaling.

- A model for open immersed curves $c : [0, 1] \to \mathbb{R}^2$; using a transformation known as *square-root velocity representation"* the Riemannian Manifold is isometric to a subset of the unit sphere in $V = L^2([0, 1])$; see [20].
- A model for closed immersed curves $c : S^1 \to \mathbb{R}^2$; using an appropriate transformation the Riemannian Manifold is isometric to a subset of the Stiefel Manifold $\mathbf{St}(V, 2)$; see [25, 24, 21] .

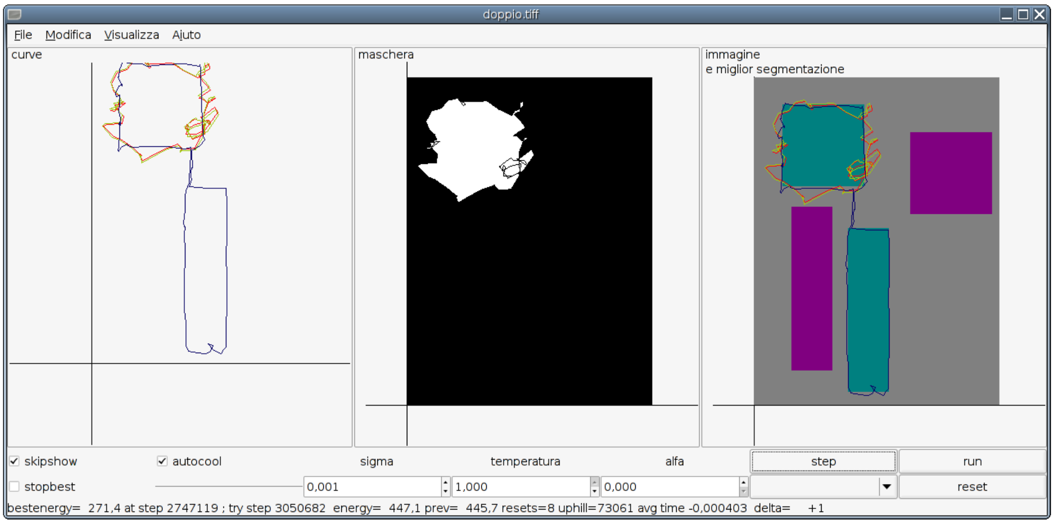

Figure 1: Stocastic minimization of curve-based segmentation energy, from [10]. At the left pane , three curves: the blue curve is the best result so far; in green and red, stochastic steps. In center pane, the current examined region for segmentation. In the right pane: the image to be segmented, with curves superimposed.

Since the sphere in the first model is the special case $\mathbf{St}(1, V)$, we are in both cases interested in infinite dimensional Stiefel Manifolds. In both cases, the space of smooth immersions is completed to a larger space of absolutely continuous curves, so that the *"shape space"* is now $\mathbf{St}(p, V)$.

Since stochastic methods play an important role in applications, we are then lead to investigate them in $\mathbf{St}(p, V)$. In particular, in [10] a stochastic minimization method was developed to seek numerically global minima for a task of image segmentation; curves would stochastically evolve by a scheme resembling the "random walk on manifold" presented later in Section 6; a pruning method (inspired by simulated annealing) would drive the random walk towards a global minimum: see Figure 1. In [21] a stochastic method was developed in $\mathbf{St}(2, V)$, similar to the classical Kalman filtering, to track a moving object.

A question remained open: could the numerical methods in [10] and in [21] be explained as a space and time discretization of an (yet to be understood) infinite dimensional stochastic method in $\mathbf{St}(2, V)$? Space discretization would not pose a problem, since it can be argued that $\mathbf{St}(p, V)$ can be approximated by $\mathbf{St}(p, \mathbb{R}^N)$ for N large (using *e.g.* Fourier series). There remain thus this question: does a discrete time random walk in $\mathbf{St}(p, V)$ somehow approximate a time continuous stochastic process in $\mathbf{St}(p, V)$? More in general: how can we define probabilities and stochastic methods in $\mathbf{St}(p, V)$? Some positive and negative results were found in [2]. In this paper we will eventually provide a positive result in Theorem 7.3: the discrete time random walk on $\mathbf{St}(p, V)$ can indeed converge to a time continuous process, when the time partition gets finer and finer. In the spirit of the Donsker's Theorem, this is a first step to an operative definition of *"Brownian Motion"* on $\mathbf{St}(p, V)$. In this paper we will not prove that there is an unique possible limit, neither will we characterize its property: this is left for a future paper. Eventually all of the above will provide a sound foundation for methods such as the ones in [10] and in [21].

3 Definitions

Let H be a separable Hilbert space. We agree that variables n, m, h, k, i, j, l are natural numbers. We will use the theory of *"nets"*; in this paper a net is a function whose domain is a partially ordered directed set with no maxima (abbreviated to *"dposet"* in the following): see [18] for details and properties.

3.1 Measures

Definition 3.1. Given measurable spaces (X_1, Σ_1) and (X_2, Σ_2), a measurable mapping $f : X_1 \to X_2$ and a measure $\mu\colon \Sigma_1 \to [0, +\infty]$, the **push forward** of μ is defined to be the measure $f_\sharp\mu : \Sigma_2 \to [0, +\infty]$ given by

$$f_\sharp\mu(B) = \mu\left(f^{-1}(B)\right) \text{ for all } B \in \Sigma_2 \quad .$$

The push forward measure is denoted also as $f_*\mu$, $\mu \circ f^{-1}$, or $f\#\mu$.

Definition 3.2 (Law *a.k.a.* Distribution)**.** If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $Y : \Omega \to X_2$ is a r.v. and γ is a probability measure on (X_2, Σ_2), we will write

$$Y \sim \gamma \quad \text{when} \quad \gamma = Y_\sharp\mathbb{P} \quad ,$$

we will say that γ is the *law* or the *distribution* of Y; similarly for $Y, Z : \Omega \to X_2$

$$Y \sim Z \quad \text{when} \quad Y_\sharp\mathbb{P} = Z_\sharp\mathbb{P} \quad .$$

We will use the *narrow convergence*. [5]

Definition 3.3 (Narrow Convergence)**.** Let S be a Hausdorff topological space. Given a net of Radon measures μ_α, μ on S, for $\alpha \in A$ a *dposet*, we will say that $\mu_\alpha \to \mu$ **narrowly** if

$$\forall f \in C_b(S) \quad , \quad \lim_\alpha \int_S f(x)\,\mathrm{d}\mu_\alpha(x) = \int_S f(x)\,\mathrm{d}\mu(x) \quad .$$

We agree on this (non standard) definition.

Definition 3.4. Let $\mathfrak{T}$ be a dposet, and S be a Hausdorff topological space. Let μ_τ be a net of Radon measures on S: it is **tight** if $\forall\varepsilon > 0$ there is a compact set $C \subseteq S$ such that

$$\limsup_{\tau\in\mathfrak{T}} \mu_\tau(S \setminus C) \le \varepsilon \quad .$$

Let $\mathfrak{X}^\tau$ a net of r.v. taking values in S, for $\tau \in \mathfrak{T}$: it is **tight** if the net of laws $\mu_\tau = \mathfrak{X}^\tau_\sharp\mathbb{P}$ is tight.

3.2 Probability setting

Hypotheses 3.5. We fix a constant $c_t > 0$ that will be used to bound temporal finess and a constant $c_3 > 0$ that will be used to control exponential decay. [6]

We will use a Borel measure γ on H satisfying:

- $$\int_H \|x\|^4 e^{4c_3c_t\|x\|}\,\mathrm{d}\gamma(x) < \infty$$
- γ is centered
$$\int_H x\,\mathrm{d}\gamma(x) = 0 \quad ;$$

[5] Other text call this *convergence in distribution* or *weak convergence*.

[6] The constants c_3, c_t will appear again in subsequent hypotheses.

- There is [7] a linear compact symmetric injective operator $K : H \to H$ such that $\gamma(K(H)) = 1$ and
- $$\int_H \|K^{-1}x\|^4 \,\mathrm{d}\gamma(x) < \infty \quad .$$

All the above hold true when $\gamma = N(0,Q)$ a Gaussian measure (as defined in next section): see Proposition 3.6 and Remark 5.2.

Let γ be such a probability on H. We need a Probability Space, so that we have i.i.d. r.v. $Y_t : \Omega \to H$ each with distribution $Y_{t\sharp}\mathbb{P} = \gamma$, for $t \in \mathbb{Q}$; to this end we may set

$$\Omega = \times_{t\in\mathbb{Q}} H \quad , \quad \mathcal{F} \quad , \quad \mathbb{P} = \otimes_{t\in\mathbb{Q}}\gamma \quad ;$$

where $\mathcal{F}$ is the σ-algebra generated by null sets of $\mathbb{P}$ and by $\mathcal{B}(\Omega)$, the Borel σ-algebra.

Proposition 3.6. *For any $c \in [0, 4c_3], \alpha \in [0, 4]$ there is a constant $\tilde{c} = \tilde{c}(\alpha, c, c_t) > 0$ such that for all $\delta \in [0, c_t]$,*

$$\mathbb{E}[g(\delta\|Y_t\|)] \le \tilde{c}\delta^\alpha \quad ,$$

where $g(s) = s^\alpha e^{cs}$. (The proof is in Section B)

3.3 Gaussian measures

Let H be a separable Hilbert space.

Definition 3.7. Suppose that $a \in H$ and $Q : H \to H$ is a linear symmetric trace-class operator such that the quadratic form $\langle x, Qx\rangle_H$ is non negative. We recall that $\gamma = N(a,Q)$ is a Gaussian measure in the Hilbert space H when the characteristic function (or Fourier transform) is

$$\forall f \in H \quad , \quad \int_H e^{i\langle f,x\rangle} \,\mathrm{d}\gamma(x) = \exp\left(i\langle a, f\rangle - \frac{1}{2}\langle x, Qx\rangle_H\right)$$

(Theorem 2.3.1 in [6]; Section 1.5 in [9]). The mean and variance are characterized by

$$\langle f, a\rangle = \int_H \langle f, x\rangle \;\mathrm{d}\gamma(x) \tag{3.1}$$

$$\langle f, Qg\rangle_H = \int_H \langle f, x-a\rangle \,\langle g, x-a\rangle \;\mathrm{d}\gamma(x) \tag{3.2}$$

for all $f, g \in H$. In particular γ is called *centered* when $a = 0$ and *non-degenerate* when the variance has empty kernel.

The proposition 3.6 is true for Gaussian Measures.

Proposition 3.8. *Let $c \ge 0, \alpha > 0$ and let $g(s) = s^\alpha e^{cs}$, suppose $Y \sim N(0,Q)$ then there is a $\tilde{c} = \tilde{c}(\alpha, c, Q, c_t) > 0$ such that for $0 \le \delta \le c_t$*

$$\mathbb{E}[g(\delta\|Y\|)] \le \tilde{c}\delta^\alpha \quad .$$

(The proof is in Section B)

[7] The last two requests are loosely connected to what is explained in Example 3.8.13 in [6].

3.4 Random walk

As aforementioned, in this paper a **net** is a function whose domain is a partially ordered directed set with no maxima (abbreviated to *"dposet"*). We will use these dposets.

Definition 3.9. We fix a constant $c_t > 0$, the same constant as in Hypotheses 3.5.

1. Let
$$\tau = \{t_0 = 0 < t_1 < t_2 \ldots\} \subset \mathbb{Q}$$
be such that
$$\lim_{n\to\infty} t_n = \infty \quad , \quad \sup_n (t_{n+1} - t_n) \le c_t \quad .$$
Let $\mathfrak{T}$ be the dposet of all such τ, ordered by inclusion.
2. Let $T > 0$ we define $\mathfrak{T}_T$ be the dposet of all τ of the form
$$\tau = \{t_0 = 0 < t_1 < t_2 < \ldots < t_n = T\}$$
with $t_0, t_1, \ldots t_{n-1} \in \mathbb{Q}$ and again $\max_{1\le j\le n}(t_j - t_{j-1}) \le c_t$. (Note that we do not require that $T \in \mathbb{Q}$).

We will use $\mathfrak{T}$ for processes with $t \in \mathbb{R}^+ = [0, \infty)$; while we will use $\mathfrak{T}_T$ for processes with $t \in [0, T]$. We will actually define all processes as in the first case, for simplicity; but then, up to restricting $t \in [0, T]$, we will study tightness using $\tau \in \mathfrak{T}_T$.

Let H be a separable Hilbert space.

Definition 3.10. We will need a Borel map
$$D : H^2 \times (\mathbb{R}^+)^2 \to H$$
continuous in the last argument and such that $D(x, v, t, 0) = x$.

Each random walk is a process $\mathfrak{X}^\tau = (\mathfrak{X}^\tau_t)_{t\ge0}$ taking values in H.

We fix $\mathfrak{X}_0$ a random variable taking values in H, independent of all Y_t.

To define $\mathfrak{X}^\tau$ we define auxiliary processes X^τ_n for $n \in \mathbb{N}$; where we define $X^\tau_0 = \mathfrak{X}_0$, and we define recursively
$$X^\tau_{(n+1)} = X^\tau_n + D\left(X^\tau_n \ , \ Y_{t_n} \ , \ t_n \ , \ (t_{n+1} - t_n)\right) \qquad \text{(as in (1.1);)}$$
then we interpolate using
$$\mathfrak{X}^\tau_t = X^\tau_n + D\left(X^\tau_n \ , \ \ Y_{t_n} \ , \ t_n \ , \ (t - t_n)\right) \qquad \text{(as in (1.2))}$$
for $t_n \le t \le t_{(n+1)}$; so each trajectory $t \mapsto \mathfrak{X}^\tau_t(\omega, t)$ is continuous; hence each $\mathfrak{X}_t$ is a r.v. taking value in $C(\mathbb{R}^+; H)$, the Frechét space of continuous functions $x : \mathbb{R}^+ \to H$ with $\mathbb{R}^+ = [0, \infty)$.

Proposition 3.12 (Filtration)**.** *Let $\mathcal{F}_t$ be the σ-algebra generated by $\mathfrak{X}_0$ and by Y_s for $s < t$ (augmented with the null sets of $\mathbb{P}$). The process X^τ_n for $t_n \le t$ is $\mathcal{F}_t$-measurable. The process $\mathfrak{X}^\tau_s$ for $s \le t$ is $\mathcal{F}_t$-measurable.*

Remark 3.13. The choice of interpolation (1.2) has a beneficial effect: suppose T is positive but $T \notin \tau$; define $\hat\tau = (\tau \cup \{T\}) \cap [0, T]$ so $\hat\tau \in \mathfrak{T}_T$, then
$$\mathfrak{X}^{\hat\tau}_t = \mathfrak{X}^\tau_t \quad \forall t \le T \quad .$$
This means that, up to adding T to τ, we can consider any process defined above as a process $\mathfrak{X}^\tau_t$ for $t \in [0, T]$ and $\tau \in \mathfrak{T}_T$. (Note that we do not require that $T \in \mathbb{Q}$).

We recall that $C(I;H)$ is a complete separable metric space; so when the topology associated $C(I;H)$ is the narrow topology and the family is tight (as defined in Definition 3.4), by Prokhorov's theorem[8] the set of limit points is not empty. When $I=[0,T]$ this can also be explained using sequences.

Lemma 3.14. *Let $T>0, I=[0,T]$; let $q_0=0, q_1=T$ and*

$$\{q_2, q_3, \ldots\} = (0,T)\cap\mathbb{Q}$$

be an enumeration, let

$$\theta_n = \{q_0, q_1, \ldots, q_n\}$$

then for $n\geq 2$ the sequence $(\theta_n)_n$ is cofinal *in $\mathfrak{T}_T$. So the limit points*

$$\bigcap_{\hat\tau\in\mathfrak{T}_T} \overline{\{\mathfrak{X}^\tau : \hat\tau\subseteq\tau\}} \tag{3.3}$$

along $\mathfrak{T}_T$ coincide with the limit points

$$\bigcap_{n\in\mathbb{N}} \overline{\{\mathfrak{X}^{\theta_k} : k\geq n\}} \tag{3.4}$$

along the sequence $(\theta_n)_n$. Similarly limits, limsup, liminf, tightness, etc, can be defined using that sequence $(\theta_n)_n$.

A similar result does not hold for $\mathfrak{T}$.

Proposition 3.15. *There does not exist a cofinal sequence $f:\mathbb{N}\to\mathfrak{T}$. (The proof is in Appendix B).*

3.5 Manifold

Suppose M is a manifold smoothly embedded in H; we consider it as a Riemannian Manifold, since it inherits the scalar product from H. For $x\in M, v\in T_xM$ we denote by $\exp_x(v)$ the exponential map. We require that M be a closed subset, so it is geodetically complete. We consider T_xM as a linear subspace of H, not as its affine translation containing x. For $x\in M$ we define the orthogonal projection $\pi_{T_xM}: H\to T_xM$; note that π_{T_xM} is symmetric that is $\pi_{T_xM}=\pi^*_{T_xM}$; we will call $P_x=\pi_{T_xM}$ for simplicity.

3.5.1 Random walks on manifolds

In this case each random walk is a process $\mathfrak{X}^\tau=(\mathfrak{X}^\tau_t)_{t\geq 0}$ taking values in M. We fix $\mathfrak{X}^\tau_0$ a random variable taking values in M, independent of all Y_t. We define recursively $X^\tau_0=\mathfrak{X}^\tau_0$ and

$$X^\tau_{(n+1)} = \exp_{X^\tau_n}\left(\sqrt{(t_{n+1}-t_n)}P_{X^\tau_n}Y_{t_n}\right)$$

Then again we define $\mathfrak{X}^\tau$ by interpolating along geodesics

$$\mathfrak{X}^\tau_t = \exp_{X^\tau_n}\left(\sqrt{(t-t_n)}P_{X^\tau_n}Y_{t_n}\right)$$

for $t_n\leq t\leq t_{(n+1)}$; and again $t\mapsto\mathfrak{X}^\tau_t$ is continuous. If, for $x\in M, y\in H$, we let

$$D(x,y,t,s) \stackrel{\text{def}}{=} \exp_x\left(\sqrt{s}P_x y\right) - x \tag{3.5}$$

then we obtain the same definition as in the previous section 3.4; so all comments and results apply to this case as well. Note that, in the words of Section 2.2.3, P_xY_t is the source of *"white noise"* that we are using to drive the random walk. We will come back to to manifolds in Section 6 and to Stiefel Manifolds in Section 7

[8] See the version of Prokhorov's theorem in Theorem 4.16 in [18].

4 Tightness by Ascoli-Arzelà Theorem

In the following, for $\psi : \mathbb{R} \to \mathbb{R}$, *"monotonic"* means *monotonically weakly increasing* that is $s \le t \Rightarrow \psi(s) \le \psi(t)$.

Definition 4.1. Let $I \subseteq \mathbb{R}$ an interval, E a normed vector space, for $x : I \to E$ uniformly continuous and $\eta > 0$ we define *the modulus of continuity*

$$\omega_{I,E}(x,\eta) \stackrel{\text{def}}{=} \sup\{\|x(t) - x(s)\|_E : t, s \in I, |t-s| \le \eta\} \quad ;$$

note that $\omega_{I,E}(x,\cdot)$ is continuous, sub-additive, monotonic, and $\omega_{I,E}(x,0) = 0$.

We recall that a set is called *pre-compact* if its closure is compact. We will use this version of Ascoli-Arzelà Theorem. (Recall that if I is compact then $C(I;S) = C_b(I;S)$)

Theorem 4.2. *Suppose H is a Banach space. Let $I \subseteq \mathbb{R}$ be a compact interval. Let $F \subseteq C(I;H)$ be a family of continuous functions $x : I \to H$. Consider these two clauses:*

- *there is $J \subseteq I$ countable dense subset such that for each $t \in J$ there exists a pre-compact set $C_t \subset H$ such that $\forall x \in F, x(t) \in C_t$;*
- $$\lim_{\eta \to 0} \sup_{x \in F} \omega_{I,H}(x,\eta) = 0 \quad .$$

The above two clauses hold if and only if F is pre-compact in $C(I;H)$.

For probability theory we transform the above as follows.

Theorem 4.3. *Suppose H is a Banach space. Let $I = [0,T]$. Suppose that $\mathfrak{X}_\alpha : \Omega \to C(I;H)$ is a net of processes (with $\alpha \in A$) such that*

- $\forall \varepsilon > 0$ *there exists a countable set $J = \{a_0, a_1, \ldots a_j \ldots\}$ dense in I and compact sets $C_j \subset H$ such that*
$$\forall \alpha \; \forall j, \; \mathbb{P}\{\mathfrak{X}_\alpha(a_j) \notin C_j\} \le \varepsilon 2^{-j} \tag{4.1}$$
- $\forall \varepsilon_0 > 0$, $\forall \varepsilon_1 > 0$, $\exists \eta > 0, \exists \alpha_0 \in A$ *such that*
$$\forall \alpha \in A, \alpha \ge \alpha_0 \Rightarrow \mathbb{P}\left\{\omega_{I,H}(\mathfrak{X}_\alpha, \eta) \ge \varepsilon_0\right\} \le \varepsilon_1 \tag{4.2}$$

then the sequence $\mathfrak{X}_\alpha$ is tight[9] *in $C(I;H)$.*

Proof. Fix $\varepsilon > 0, \varepsilon_0 > 0, \varepsilon_1 > 0$, let $F \subseteq C(I;H)$ be the set of all x such that

$$\forall j, \; x(a_j) \in C_j$$

and $\omega_{I,H}(x,\eta) < \varepsilon_0$. Then $\forall \alpha \ge \alpha_0$ we have

$$\mathbb{P}(\mathfrak{X}_\alpha \notin F) \le 2\varepsilon + \varepsilon_1 \quad .$$ □

Remark 4.4. The second hypothesis (4.2) may be reformulated as

$$\forall \varepsilon_0 > 0 \quad , \quad \lim_{\eta \to 0} \limsup_{\alpha} \mathbb{P}\left\{\omega_H(I, \mathfrak{X}_\alpha, \eta) \ge \varepsilon_0\right\} = 0 \tag{4.3}$$

since ω is monotonic in η.

[9] See Definition 3.4.

5 Tightness of random walks

Let again H be a separable Hilbert space. The purpose of this section is to state and prove the Theorem 5.5 on tightness of random walks in H.

5.1 Tightness operator

Let γ be as defined in Hypotheses 3.5. We required in Hypotheses 3.5 that there is $K : H \to H$ a linear symmetric compact injective operator such that $\gamma(K(H)) = 1$, equivalently

$$\mathbb{P}(Y_t \in K(H)) = 1 \quad \forall t \quad .$$

Define now

$$D^K(0, r) = \{Kx : x \in H \ , \ \|x\| \le r\} = K(B_H(0, r)) = rK(B_H(0, 1))$$

then

- they are pre-compact (this means that the closures $\overline{D^K(0, r)}$ are compact), and
- $$\bigcup_n D^K(0, n) = K(H)$$

  so

  $$\gamma\left(\bigcup_n D^K(0, n)\right) = 1 \quad ;$$
- this means that $\forall \varepsilon > 0 \ \exists n$ such that

$$\gamma\left(D^K(0, n)\right) \ge 1 - \varepsilon \quad .$$

Remark 5.1. Since H is a separable Hilbert space, then every probability measure on it is Radon hence tight; the above gives us an accessible family of sets for tightness of γ.

Up to rescaling K we will assume that

$$\forall v \in H \ , \ \|Kv\|_H \le \|v\|_H \quad .$$

Remark 5.2. In general, for any Hilbert space H and Gaussian measure $\gamma = N(0, Q)$, such operator K always exists.

5.2 Tightness Theorem

Definition 5.3. Given a linear continuous injective operator $K : H \to H$ we define

$$\|v\|_K \stackrel{\text{def}}{=} \begin{cases} \|K^{-1}v\|_H & \text{if } v \in K(H) \\ \infty & \text{if } v \notin K(H) \end{cases} \quad . \tag{5.1}$$

Similarly for scalar products, for $v, w \in K(H)$ we define

$$\langle v, w\rangle_K = \langle K^{-1}v, K^{-1}w\rangle_K \quad .$$

Definition 5.4. Given Banach spaces B_1, B_2, we define $\mathcal{L}(B_1; B_2)$ to be the space of linear continuous operators $A : B_1 \to B_2$. If $B_1 = B_2$ then we write $\mathcal{L}(B_1)$

Theorem 5.5. *Let $I = [0, T]$. Suppose that the random walks $\mathfrak{X}^\tau : \Omega \to C(I; H)$ above defined in Definition 3.10 satisfy*

1. *there is a $K : H \to H$ a linear compact operator satisfying the requisites in the previous section and such that*
2.
$$\mathbb{E}[\|\mathfrak{X}_0\|_K^4] < \infty \tag{5.2}$$
3. *for all $x, v \in K(H), t \in I, s \geq 0$*
$$D(x, v, t, s) \in K(H)$$
4. *there is a bounded Borel functional*
$$L(x, t) : H \times \mathbb{R}^+ \to \mathcal{L} = \mathcal{L}(H; H)$$
such that $\forall x, v \in K(H)$
$$L(x, t)v \in K(H) \quad ,$$
5. *and there are constants $c_3, c > 0$ such that for all $x, v \in K(H)$*
$$\|L(x, t)\|_K \leq c\,(\|v\|_K + \|x\|_K \|v\|_H)\, e^{c_3\|v\|_H} \tag{5.3}$$
(the constant c_3 must be the same as in Hypotheses 3.5),
6. *and such that, for all $x, v \in K(H)$, $t \in I$, $s \in [0, 1]$,*
$$\|D(x, v, t, s) - \sqrt{s}L(x, t)v\|_K \leq cs\Big(\|x\|_K \|v\|_H^2 + \|v\|_K \|v\|_H\Big) e^{c_3\sqrt{s}\|v\|_H} \quad . \tag{5.4}$$
7. *there is a constant $c_d > 1$ such that $\forall x, v \in H, \forall t \in \mathbb{R}^+, \forall s \in [0, 1]$*
$$\|D(x, v, t, s)\|_H \leq c_d\sqrt{s}\|v\|_H \quad , \tag{5.5}$$
$$\sqrt{s}\|v\| \leq 1/c_d \Rightarrow \|D(x, v, t, s) - \sqrt{s}L(x, t)v\|_H \leq c_d s \|v\|_H^2 \quad . \tag{5.6}$$

Then the family $\mathfrak{X}^\tau$, for $\tau \in \mathfrak{T}_T$, is tight in $C(I; H)$.

To prove this Theorem, we will use Theorem 4.3. The proof is developed in the following sections.

Remark 5.6. In particular, $D(x, v, t, s)$ is Frechét differentiable in v at $v = 0$, and the differential is $\sqrt{s}L(x, t)$; and by (5.5) we have

$$\|L(x, t)\|_{\mathcal{L}} \leq c_d \quad , \quad \forall x \in H \quad . \tag{5.7}$$

Remark 5.7. The third hypothesis in 3.5 and the first two hypotheses above imply that

$$\forall t \geq 0\ ,\ \mathbb{P}\{\mathfrak{X}_t^\tau \in K(H)\} = 1\ .$$

5.2.1 Corollaries

Corollary 5.8. *Suppose that the hypotheses of Theorem 5.5 hold for any $T > 0$; then the family $\mathfrak{X}^\tau$, for $\tau \in \mathfrak{T}$, is tight in $C(\mathbb{R}^+; H)$.*

Proof. We recall this fact. Let $T > 0$; consider the restriction map

$$r_T : C(\mathbb{R}^+; H) \to C([0, T]; H) \tag{5.8}$$

given by $r_T f = f_{|_{[0,T]}}$; then the topology on $C(\mathbb{R}^+; H)$ is the initial topology with respect to the maps r_n and the Banach spaces $C([0, n]; H)$, for $n \in \mathbb{N}$.

Let $\varepsilon > 0$, for any $n \in \mathbb{N}$ by Theorem 5.5 there exists a compact set $E_n \subseteq C([0,n];H)$ such that

$$\mathbb{P}\{\mathfrak{X}^\tau \notin r_n^{-1}(E_n)\} \leq \varepsilon 2^{-n}$$

let

$$E = \{f \in C(\mathbb{R}^+;H) : \forall n \in \mathbb{N}, r_n f \in E_n\}$$

then (by a diagonal argument) E is precompact in $C(\mathbb{R}^+;H)$ and

$$\mathbb{P}\{\mathfrak{X}^\tau \notin E\} = \mathbb{P}\{\mathfrak{X}^\tau \in E^c\} = \mathbb{P}\left\{\mathfrak{X}^\tau \in \bigcup_n r_n^{-1}(E_n^c)\right\} \leq$$
$$\leq \sum_n \mathbb{P}\left\{\mathfrak{X}^\tau \in r_n^{-1}(E_n^c)\right\} = \sum_n \mathbb{P}\left\{r_n \circ \mathfrak{X}^\tau \notin E_n\right\} \leq 2\varepsilon \qquad \square$$

Since $C(\mathbb{R}^+;H)$ is a Fréchet space, by Prokhorov's theorem we obtain this result.

Corollary 5.9. *Suppose that the hypotheses of Theorem 5.5 hold for any $T > 0$; the net of processes $\mathfrak{X}^\tau$, as r.v. in the Frechét space $C(\mathbb{R}^+;H)$, has narrow limit points.*

Corollary 5.10. *By Lemma 5.18 we have*

$$\mathbb{E}[\|\mathfrak{X}_t^\tau\|_K^4] < \infty \tag{5.9}$$

so the process $\mathfrak{X}_t^\tau$ can be restarted at time t using $\mathfrak{X}_t^\tau$ as initial time.

Suppose that the family $\mathfrak{X}^\tau$ is tight in $C(I;H)$ (with $I = [0,T]$ or $I = [0,\infty)$) ; then by Prokhorov's theorem, the set of limit points is not empty; obviously, being $\mathfrak{X}$ a random variable $C(I;H)$, then (almost all) paths are continuous. Something more can be said.

Corollary 5.11. *Any limit point $\mathfrak{X}$ will have a version such that almost all trajectories $t \mapsto \mathfrak{X}_t$ of $\mathfrak{X}$ are Hölder continuous functions with an arbitrary exponent smaller than $1/4$.*

Proof. We use Theorem 4.10 from [18] with $p_1 < p_2 = 4$ and use Lemma 5.18 below to state that

$$\lim_\tau \mathbb{E}[\|\mathfrak{X}_t^\tau - \mathfrak{X}_s^\tau\|_H^{p_1}] = \mathbb{E}[\|\mathfrak{X}_t - \mathfrak{X}_s\|_H^{p_1}] \leq c|t-s|^{p_1/2} \quad .$$

We use the Kolmogoroff test[10]: we apply it with $\delta = p_1, \varepsilon = 1$ and replacing

$$\rho(Z(t), Z(s))^\delta = \|\mathfrak{X}_t - \mathfrak{X}_s\|_H^{p_1}$$

so there is a version where paths are Hölder continuous functions with an arbitrary exponent smaller than $\frac{1}{2} - \frac{1}{p_1}$. $\square$

Remark 5.12. At this level of generality, we do not expect that there is an unique limit point. Consider this example. Going back to the classical Donsker Theorem 2.1, this time we define the random walk $\mathfrak{X}_t^\tau$ by setting $\mathfrak{X}_0 = 0$, $H = \mathbb{R}$ and $D(x,v,t,s) = g(t)v\sqrt{s}$ where $g(t) = 1$ if $t \in \mathbb{Q}$ otherwise $g(t) = 2$. Then the above Theorem can be applied; but setting $\tau_n = \{i/n : j \in \mathbb{N}\}$, $\theta_n = \{\pi i/n : j \in \mathbb{N}\}$, we have

$$\mathfrak{X}^{\tau_n} \to_n W \quad , \quad \mathfrak{X}^{\theta_n} \to_n 2W$$

where W is the standard Brownian Motion.

[10] See Theorem 3.3 in [8], or Theorem 4.1 in [18]

5.2.2 Proof of 5.5, step 1

In this section we prove that, under the hypotheses of Theorem 5.5, the first hypothesis in Theorem 4.3 is satisfied. We will use this Lemma in two ways, with K being the compact operator defined above in the hypotheses of 5.5, but also with K being the identity.

Lemma 5.13. *Let $K : H \to H$ a linear injective operator (not necessarily compact). Let $\|v\|_K$ be defined in eqn.* (5.1)*. We assume that for all $x, v \in K(H), t, s > 0$*

$$D(x, v, t, s) \in K(H) \ , \ L(x, t)v \in K(H)$$

(these are the hypotheses 3 and 4 from the Theorem), but we rewrite hypothesis 6 in this form: there are constants $c_{1D}, c_{2D} \geq 0$ such that for all $x, v \in K(H)$, $s \in [0, 1]$

$$\|D(x, v, t, s) - sL(x, t)v\|_K \leq s\Big(c_{1D}\|x\|_K\|v\|_H^2 + c_{2D}\|v\|_K\|v\|_H\Big)e^{c_3\sqrt{s}\|v\|_H} \quad . \tag{5.10}$$

and we rewrite hypothesis 5: there are constants $c_{1L}, c_{2L} \geq 0$ such that for all $x, v \in K(H)$,

$$\|L(x, t)v\|_K \leq (c_{1L}\|x\|_K\|v\|_H + c_{2L}\|v\|_K)\, e^{c_3\|v\|_H} \quad . \tag{5.11}$$

Define the following objects: let $\tau \in \mathfrak{T}$; fix $m \geq 0$ and [11] *$F \in \mathcal{F}_{t_m}$ with $\mathbb{P}(F) > 0$; we will write $\mathbb{E}_F$ for the expectation computed using the conditional probability $\mathbb{P}(\cdot|F)$; consider $n \geq m$; let*

$$e_m = \mathbb{E}_F[\|X_m^\tau\|_K^2] \quad , \quad b_n \stackrel{def}{=} \mathbb{E}_F[\|X_n^\tau - X_m^\tau\|_K^2] \quad .$$

Then we have two theses.

- *If $c_{1D} = c_{1L} = 0$ then*
$$b_n \leq \Big(e^{c_5(t_n - t_m)} - 1\Big) \quad . \tag{5.12}$$
- *Instead if $(c_{1L} + c_{1D}) > 0$ then*
$$b_n \leq (e_m + 1)\Big(e^{c_5(t_n - t_m)} - 1\Big) \quad . \tag{5.13}$$

where $c_5 > 0$ depends only on $c_{1L}, c_{2L}, c_{1D}, c_{2D}$, on $\tilde{c}(4, 4c_3)$ from Prop. 3.6, on K and the law γ of Y_t; but c_5 does not depend on e_m, on F, and on τ.

Remark 5.14. Recall that

$$\sqrt{\mathbb{E}_F[\|X_n^\tau\|_K^2]} \leq \sqrt{\mathbb{E}_F[\|X_n^\tau - X_m^\tau\|_K^2]} + \sqrt{\mathbb{E}_F[\|X_m^\tau\|_K^2]}$$

(or otherwise using Lemma A.1) we get

$$\mathbb{E}_F[\|X_n^\tau\|_K^2] \leq (\sqrt{e_m} + \sqrt{b_n})^2 \leq 2(e_m + b_n) \quad . \tag{5.14}$$

Proof of Lemma 5.13. For $t \geq t_m$ we define

$$a_{K,\alpha} = \mathbb{E}_F[\|Y_t\|_K^\alpha] = \mathbb{E}[\|Y_t\|_K^\alpha] \ , \ a_\alpha = \mathbb{E}_F[\|Y_t\|^\alpha] = \mathbb{E}[\|Y_t\|^\alpha] \quad . \tag{5.15}$$

where the equality derives by independence; by Hypotheses 3.5 these are finite for $\alpha \leq 4$. For readability, we write X_n instead of X_n^τ, we write $\tilde{X}_n$ instead of $X_n^\tau - X_0^\tau$ and $\delta_n = t_{n+1} - t_n$; we abbreviate

$$D_n = D\Big(X_n \ , \ Y_{t_n} \ , \ t_n \ , \ \sqrt{\delta_n}\Big) \quad , \tag{5.16}$$

$$A_n = D_n - \sqrt{\delta_n} L(t_n \ , \ X_n) Y_{t_n} \quad . \tag{5.17}$$

[11] Recall from Lemma 3.12 that $\mathcal{F}_t$ is the σ-algebra generated by $\mathfrak{X}_0$ and by Y_s for $s < t$.

By (5.14) and (5.10) using Lemma A.1 and Lemma 3.6, when $c_3 > 0$

$$\begin{aligned}\mathbb{E}_F[\|A_n\|_K^2] \le & 2c_{1D}^2\delta_n\mathbb{E}_F\left[\|X_n\|_K^2\|Y_{t_n}\|_H^4 e^{2c_3\sqrt{\delta_n}\|Y_{t_n}\|_H}\right] + \\ & 2c_{2D}^2\delta_n\mathbb{E}_F\left[\|Y_{t_n}\|_K^2\|Y_{t_n}\|_H^2 e^{2c_3\sqrt{\delta_n}\|Y_{t_n}\|_H}\right] \le \\ \le & 2c_{1D}^2\mathbb{E}_F\left[\|X_n\|_K^2\right]\mathbb{E}\left[\|\sqrt{\delta_n}Y_{t_n}\|_H^4 e^{2c_3\|\sqrt{\delta_n}Y_{t_n}\|_H}\right] + \\ & 2c_{2D}^2\delta_n\sqrt{\mathbb{E}\left[\|Y_{t_n}\|_K^4\right]}\sqrt{\mathbb{E}\left[\|\sqrt{\delta_n}Y_{t_n}\|_H^4 e^{4c_3\|\sqrt{\delta_n}Y_{t_n}\|_H}\right]} \le \\ \le & \delta_n^2\left(4c_{1D}^2(b_n+e_m)\tilde{c}(4,2c_3) + 2c_{2D}^2\sqrt{\tilde{c}(4,4c_3)\,a_{4,K}}\right)\end{aligned} \tag{5.18}$$

Summarizing we have

$$\mathbb{E}_F[\|A_n\|_K^2] \le \delta_n^2 c_4^2(c_{1D}^2(b_n+e_m)+1) \quad . \tag{5.20}$$

where c_4 depends only on c_{2D}, c_3, on $\tilde{c}$ from Corollary 3.6, on $c_t > 0$ from Hypotheses 3.5 and Definition 3.9; but but c_4 does not depend on F, on c_{1D} and τ.

Similarly using (5.11)

$$\begin{aligned}\delta_n\mathbb{E}_F[\|L(X_n,t)Y_{t_n}\|_K^2] \le & 2c_{1L}^2\delta_n\mathbb{E}_F\left[\|X_n\|_K^2\|Y_{t_n}\|_H^2 e^{2c_3\sqrt{\delta_n}\|Y_{t_n}\|_H}\right] + \\ & 2c_{2L}^2\delta_n\mathbb{E}\left[\|Y_{t_n}\|_K^2 e^{2c_3\sqrt{\delta_n}\|Y_{t_n}\|_H}\right] \le \\ \le & 2c_{1L}^2\mathbb{E}_F\left[\|X_n\|_K^2\right]\mathbb{E}\left[\|\sqrt{\delta_n}Y_{t_n}\|_H^2 e^{2c_3\|\sqrt{\delta_n}Y_{t_n}\|_H}\right] + \\ & 2c_{2L}^2\delta_n\sqrt{\mathbb{E}\left[\|Y_{t_n}\|_K^4\right]}\sqrt{\mathbb{E}\left[e^{4c_3\|\sqrt{\delta_n}Y_{t_n}\|_H}\right]} \le \\ \le & \delta_n\left(4c_{1L}^2(b_n+e_m)\tilde{c}(2,2c_3) + 2c_{2L}^2\sqrt{\tilde{c}(0,4c_3)\,a_{4,K}}\right)\end{aligned} \tag{5.21}$$

summarized to

$$\delta_n\mathbb{E}_F[\|L(X_n,t)Y_{t_n}\|_K^2] \le \delta_n c_4^2(c_{1L}^2(b_n+e_m)+1) \quad , \tag{5.22}$$

possibly enlarging c_4, that now depends also on c_{2L}.

We estimate iteratively. We begin by expressing

$$\begin{aligned}\|\tilde{X}_{n+1}\|_K^2 = \|D_n + \tilde{X}_n\|_K^2 &= \|\tilde{X}_n + \sqrt{\delta_n}L(X_n,t)Y_{t_n} + A_n\|_K^2 = \\ &= \|\tilde{X}_n\|_K^2 + \delta_n\|L(X_n,t)Y_{t_n}\|_K^2 + \|A_n\|_K^2 + \\ &+ 2\sqrt{\delta_n}\langle\tilde{X}_n, L(X_n,t)Y_{t_n}\rangle_K + 2\langle\tilde{X}_n, A_n\rangle_K + \\ &+ 2\sqrt{\delta_n}\langle A_n, L(X_n,t)Y_{t_n}\rangle_K\end{aligned}$$

we then compute the expectation; we note that

$$\mathbb{E}_F[\langle\tilde{X}_n, L(X_n,t)Y_{t_n}\rangle_K] = 0$$

because Y_{t_n} has zero average and is independent of F, of X_n and X_m; whereas

$$\begin{aligned}\mathbb{E}_F[\|\tilde{X}_n\|_K^2] =& b_n \\ \delta_n\mathbb{E}_F[\|L(X_n,t)Y_{t_n}\|_K^2] \le& \delta_n c_4^2(c_{1L}^2(b_n+e_m)+1) \\ \mathbb{E}_F[\|A_n\|_K^2] \le& \delta_n^2 c_4^2(c_{1D}^2(b_n+e_m)+1) \\ \mathbb{E}_F[\langle A_n, \sqrt{\delta}_n L(X_n,t)Y_{t_n}\rangle_K] \le& \sqrt{\mathbb{E}_F[\|A_n\|_K^2]}\sqrt{\mathbb{E}_F[\delta_n\|L(X_n,t)Y_{t_n}\|_K^2]} \\ \le& \delta_n^{3/2}c_4^2(c_1^2(b_n+e_m)+1) \\ \mathbb{E}_F[\langle\tilde{X}_n, A_n\rangle_K] \le& \sqrt{b_n}\sqrt{\mathbb{E}_F[\|A_n\|_K^2]}\end{aligned}$$

where $c_1 = \max\{c_{1L}, c_{1D}\}$.

In this last line, if $c_{1D} = c_{1L} = 0$ then again we use $\sqrt{s} \le s + 1$ and (5.20) so

$$\mathbb{E}_F[\langle \tilde{X}_n, A_n \rangle_K] \le \sqrt{b_n}\sqrt{\mathbb{E}_F[\|A_n\|_K^2]} \le \delta_n c_4 \sqrt{b_n} \le \delta_n c_4 (b_n + 1)$$

so (recalling that $\delta_n \le c_t$, the constant from Hypotheses 3.5) we estimate as follows

$$b_{n+1} \le b_n + c_5 \delta_n (1 + b_n) = b_n (1 + c_5 \delta_n) + c_5 \delta_n \tag{5.23}$$

which, by the Lemma A.5 (shifting the sequence), implies (5.12).

Instead if $c_1 > 0$ we note that

$$s(c_1^2 s + 2a) \le c_1^2 s^2 + 2as + a^2/c_1^2 = (c_1 s + a/c_1)^2$$

hence using (5.20) and setting $s = b_n, a = (c_1^2 e_m + 1)/2$

$$\mathbb{E}_F[\langle \tilde{X}_n, A_n \rangle_K] \le \delta_n c_4 \sqrt{b_n}\sqrt{(c_1^2(b_n + e_m) + 1)} \le \delta_n c_4 \left(c_1 b_n + c_1 \frac{e_m}{2} + \frac{1}{2c_1} \right) \tag{5.24}$$

Eventually we estimate as follows

$$b_{n+1} \le b_n \Big(1 + c_5 \delta_n \Big) + c_5 \delta_n \Big(e_m + 1 \Big) \tag{5.25}$$

which, by the Lemma A.5 (shifting the sequence), implies (5.13). □

Corollary 5.15. *Let $m = 0, F = \Omega$, then by* (5.14) *and* (5.13) *we obtain that*

$$\mathbb{E}[\|X_n^\tau\|_K^2] \le (e_0 + 1) e^{c_5 t_n}$$

with $e_0 = \mathbb{E}[\|\mathfrak{X}_0\|_K^2] < \infty$, assuming hypothesis (5.2)*; moreover as explained in Remark 3.13 we have*

$$\mathbb{E}[\|\mathfrak{X}_t^\tau\|_K^2] \le (e_0 + 1) e^{c_5 t}$$

for all $t \ge 0$.

Conclusion of step 1 . So to conclude the first step, we consider a process $\mathfrak{X}^\tau$ and a time $T > 0$; let $\varepsilon > 0$, for $0 \le t \le T, r > 0$ we have then by Markov inequality

$$\mathbb{P}\{\|\mathfrak{X}_t^\tau\|_K \ge r\} \le \frac{\mathbb{E}[\|\mathfrak{X}_t^\tau\|_K^2]}{r^2} \le \frac{1}{r^2} c_K$$

with

$$c_K = (e_0 + 1) e^{c_5 T}$$

so setting $r_j = \sqrt{(c_K 2^j/\varepsilon)}$

$$\mathbb{P}\{\mathfrak{X}_{a_j}^\tau \notin D_K(0, r_j)\} < \varepsilon 2^{-j}$$

and this satisfies the first hypothesis in Theorem 4.3. □

5.2.3 Proof: Lemmas for step 2

In this section we prove some powerful Lemmas that then will be used to prove that the second hypothesis in Theorem 4.3 is satisfied.

Remark 5.16. Consider hypothesis 7 in Theorem 5.5; note that, for $s \in [0, 1]$, by equations (5.5) and (5.7),

$$\|D(x, v, t, s) - \sqrt{s} L(x, t) v\|_H \le \|D(x, v, t, s)\|_H + \|\sqrt{s} L(x, t) v\|_H \le 2c_d \sqrt{s} \|v\|_H \le 2c_d^2 s \|v\|_H^2$$

for $\sqrt{s}\|v\| \ge 1/c_d$ so adding (5.6) we obtain

$$\forall x, v \in H, \forall s \in [0, 1], t \ge 0 \ , \ \|D(x, v, t, s) - \sqrt{s} L(x, t) v\|_H \le 2c_d^2 s \|v\|_H^2 \quad . \tag{5.26}$$

Lemma 5.17. *Assume hypothesis 7 in Theorem 5.5; define the objects as in Lemma 5.13, with K being the identity, recall that in this case*

$$b_n \stackrel{\text{def}}{=} \mathbb{E}_F[\|X_n^\tau - X_m^\tau\|_H^2] \quad ;$$

then for $n \geq m$

$$b_n \leq \left(e^{c_5(t_n - t_m)} - 1\right) \quad . \tag{5.27}$$

In particular for $0 \leq t_m \leq t_n \leq T$ we have

$$b_n \leq c_6(t_n - t_m) \quad . \tag{5.28}$$

Proof. Use Remark 5.16 and Remark 5.6; apply Lemma 5.13 with K being the identity, $c_{1L} = c_{1D} = c_3 = 0$, $c_{2L} = c_{2D} = 2c_d^2$; we obtain the constant $c_5 > 0$ and hence we set $c_6 = c_5 e^{c_5 T}$. □

Lemma 5.18. *Assume hypothesis 7 in Theorem 5.5; We fix $\tau \in \mathfrak{T}$; we fix $m \geq 0$ and $F \in \mathcal{F}_{t_m}$; we will write $\mathbb{E}_F$ for the expectation computed using the conditional probability $\mathbb{P}(\cdot|F)$; consider $n \geq m$; letting*

$$b_n \stackrel{\text{def}}{=} \mathbb{E}_F[\|X_n^\tau - X_m^\tau\|_H^2] \quad ,$$

suppose that there is a constant c_6 such that

$$b_n \leq c_6(t_n - t_m) \quad . \tag{5.29}$$

for $0 \leq t_m \leq t_n \leq T$ (as in (5.28)) and eventually let

$$q_n \stackrel{\text{def}}{=} \mathbb{E}_F[\|X_n^\tau - X_m^\tau\|_H^4] \quad .$$

We prove that, for $0 \leq t_n \leq t_m \leq T$,

$$q_n \leq (c_7 + 2c_8) g(t_n - t_m) \tag{5.30}$$

where

$$g(t) = \frac{e^{c_7 t} - 1 - c_7 t}{c_7^2} \quad . \tag{5.31}$$

and where c_8, c_7 depend only on c_6, c_d and the fourth moment of Y_t.

Proof of Lemma. For $t \geq t_m$ we define

$$a_\alpha = \mathbb{E}_F[\|Y_t\|^\alpha] = \mathbb{E}[\|Y_t\|^\alpha]$$

where the equality derives by independence.

Again, for readability, we write X_n instead of X_n^τ, and $\tilde{X}_n$ instead of $X_n^\tau - X_m^\tau$, $\delta_n = t_{n+1} - t_n$ and

$$D_n = D\left(X_n \ , \ Y_{t_n} \ , \ t_n \ , \ \sqrt{\delta_n}\right)$$

and $\|x\| = \|x\|_H$. We compute

$$\|\tilde{X}_{n+1}\|^4 = \|\tilde{X}_n + D_n\|^4 = \left(\|\tilde{X}_n\|^2 + 2\left\langle \tilde{X}_n, D_n \right\rangle + \|D_n\|^2\right)^2 =$$

$$= \|\tilde{X}_n\|^4 + 4\left\langle \tilde{X}_n, D_n \right\rangle^2 + \|D_n\|^4 + 4\|\tilde{X}_n\|^2 \left\langle \tilde{X}_n, D_n \right\rangle + 4\left\langle \tilde{X}_n, D_n \right\rangle \|D_n\|^2 + 2\|\tilde{X}_n\|^2 \|D_n\|^2$$

then we compute integrals; for the fourth term, since

$$\mathbb{E}_F\left[\|\tilde{X}_n\|^2 \left\langle \tilde{X}_n, L(X_n, t_n) Y_{t_n} \right\rangle\right] = 0$$

by independence, then

$$\mathbb{E}_F\left[\|\tilde{X}_n\|^2\left\langle \tilde{X}_n, D_n\right\rangle\right] = \mathbb{E}_F\left[\|\tilde{X}_n\|^2\left\langle \tilde{X}_n, D_n - \sqrt{\delta_n}L(X_n,t_n)Y_{t_n}\right\rangle\right] \tag{5.32}$$

so by (5.26)

$$\begin{aligned}\left|\mathbb{E}_F\left[\|\tilde{X}_n\|^2\left\langle \tilde{X}_n, D_n\right\rangle\right]\right| &\le \mathbb{E}_F\left[\|\tilde{X}_n\|^3\|D_n - \sqrt{\delta_n}L(X_n,t_n)Y_{t_n}\|\right] \le \\ &\le 2c_d^2\delta_n\mathbb{E}_F\left[\|\tilde{X}_n\|^3\|Y_{t_n}\|^2\right] = \\ &= 2c_d^2\delta_n\mathbb{E}_F\left[\|\tilde{X}_n\|^3\right]\mathbb{E}\left[\|Y_{t_n}\|^2\right] \le \\ &\le 2c_d^2\delta_n a_2\sqrt{\mathbb{E}_F\left[\|\tilde{X}_n\|^2\right]\mathbb{E}_F\left[\|\tilde{X}_n\|^4\right]} \le \\ &\le 2c_d^2\delta_n a_2\sqrt{q_n}\sqrt{b_n}\end{aligned}$$

(again by using independence in the third step). For the other terms, using (5.5),

$$\begin{aligned}\mathbb{E}_F\left[\left\langle \tilde{X}_n, D_n\right\rangle^2\right] \le& \mathbb{E}_F\left[\|\tilde{X}_n\|^2\|D_n\|^2\right] \le \\ &\le c_d^2\mathbb{E}_F\left[\|\tilde{X}_n\|^2\|\sqrt{\delta_n}Y_{t_n}\|^2\right] \le c_d^2 a_2 b_n\delta_n \quad , \\ \mathbb{E}_F\left[\|D_n\|^4\right] \le& c_d^4 a_4\delta_n^2 \quad , \\ \mathbb{E}_F\left[\left\langle \tilde{X}_n, D_n\right\rangle\|D_n\|^2\right] \le& \mathbb{E}_F\left[\|\tilde{X}_n\|\|D_n\|^3\right] \le \\ &\le c_d^3\delta_n^{3/2}\mathbb{E}_F\left[\|\tilde{X}_n\|\|Y_n\|^3\right] \le c_d^3 a_3\sqrt{b_n}\delta_n^{3/2} \quad ,\end{aligned}$$

Eventually we use $b_n \le c_6 t_n$ and note that $t_n \ge \delta_n$ then $t_n\delta_n \ge \sqrt{t_n}\delta_n^{3/2}$ and $t_n\delta_n \ge \delta_n^2$; hence (defining $c_7, c_8 > 0$ appropriately), summarizing

$$q_{n+1} \le q_n + (c_7\sqrt{q_n t_n} + c_8 t_n)\delta_n$$

using Lemma A.3 (shifting the sequence) we obtain (5.30). □

We recall Etemadi's inequality [13] in the version of Theorem 22.5 in [4].

Lemma 5.19 (Etemadi's inequality)**.** *Suppose that S_n is a process taking value in normed space, and it is the sum $S_n = Y_1 + \cdots + Y_n$ of i.i.d. r.v. $(Y_n)_n$; then for $\varepsilon > 0$ we have*

$$\mathbb{P}\left(\max_{1\le k\le l}|S_k| \ge 3\varepsilon\right) \le 3\max_{1\le k\le l}\mathbb{P}\left(|S_k| \ge \varepsilon\right) \quad .$$

This is a keys step in the proof of Donsker Theorem, but we cannot use it in this form. To conclude the proof of Theorem 5.5 we need to prove a similar result, adapted to our process and hypotheses.

Lemma 5.20. *In the hypotheses of Theorem 5.5, with $\tau \in \mathfrak{T}$ (as in previous Lemmas), we fix $T, \varepsilon > 0$, we fix $l \ge m > 0$ integers such that $t_l \le T$, then*

$$\mathbb{P}\left(\max_{m\le k\le l}\|X_k^\tau - X_{m-1}^\tau\| > 3\varepsilon\right) \le \frac{c_{10}}{\varepsilon^4}g(t_l - t_{m-1}) \tag{5.33}$$

where again g was defined in (5.31)*; and c_{10} depends only on the constants in previous Lemmas.*

Proof. Let $X = X^\tau$ for simplicity, and

$$\tilde{X}_k = X_k - X_{m-1} \quad ,$$

note that

$$\tilde{X}_n - \tilde{X}_j = X_n - X_j \quad .$$

Let

$$A_m = \left\{ \|\tilde{X}_m\| > 3\varepsilon \right\}$$

and for $j = m+1, \dots l$ let

$$A_j = \left\{ \max_{m \le i \le j-1} \|\tilde{X}_i\| \le 3\varepsilon \wedge \|\tilde{X}_j\| > 3\varepsilon \right\}$$

so

$$\bigcup_{j=m}^{l} A_j = \left\{ \max_{m \le k \le l} \|\tilde{X}_k\| > 3\varepsilon \right\}$$

then, further disintegrating,

$$\begin{aligned} \mathbb{P}\left(\max_{m \le k \le l} \|\tilde{X}_k\| > 3\varepsilon \right) &\le \mathbb{P}\left(\|\tilde{X}_l\| \ge \varepsilon \right) + \sum_{j=m}^{l} \mathbb{P}\left(A_j \cap \{\|\tilde{X}_l\| < \varepsilon\} \right) \le \\ &\le \mathbb{P}\left(\|\tilde{X}_l\| \ge \varepsilon \right) + \sum_{j=m}^{l} \mathbb{P}\left(A_j \cap \{\|X_l - X_j\| > 2\varepsilon\} \right) = \\ &= \mathbb{P}\left(\|\tilde{X}_l\| \ge \varepsilon \right) + \sum_{j=1}^{l} \mathbb{P}(A_j) \mathbb{P}\left(\{\|X_l - X_j\| > 2\varepsilon\} \mid A_j \right) \end{aligned}$$

by Markov

$$\mathbb{P}\left(\{\|X_l - X_j\| > 2\varepsilon\} \mid A_j \right) \le \frac{1}{2^4 \varepsilon^4} \mathbb{E}\left[\|X_l - X_j\|^4 \ \mid \ A_j \right] \quad .$$

We use Lemma 5.17 with K the identity and $F = A_j$; note that indeed $A_j \in \mathcal{F}_{t_j}$; having set $T > 0$ we obtain (5.27) that is (5.29). So (5.29) is satisfied and we can apply Lemma 5.18 to obtain the eqn. (5.30) in the thesis in Lemma 5.18, that we rewrite as

$$\mathbb{E}\left[\|X_n - X_j\|^4 \mid A_j \right] \le (c_7 + 2c_8) g(t_n - t_j) \quad . \tag{5.34}$$

Plugging it all in

$$\sum_{j=m}^{l} \mathbb{P}(A_j) \mathbb{P}\left(\{\|X_l - X_j\| > 2\varepsilon\} \mid A_j \right) \le \frac{1}{2^4 \varepsilon^4} (c_7 + 2c_8) g(t_l - t_{m-1}) \sum_{j=m}^{l} \mathbb{P}(A_j) \quad .$$

Similarly we deal with the first term $\mathbb{P}\left(\|\tilde{X}_l\| \ge \varepsilon \right)$. □

We then prove the same result for the process $\mathfrak{X}^\tau$.

Corollary 5.21. *In the hypotheses of the previous Lemma, fix $\varepsilon > 0$, then for $t_m \in \tau$ and $t_m \le s \le T$*

$$\mathbb{P}\Big(\sup_{t_m < t \le s} \|\mathfrak{X}^\tau_t - \mathfrak{X}^\tau_{t_m}\| > 3\varepsilon \Big) \le \frac{c_{10}}{\varepsilon^4} g(s - t_m) \tag{5.35}$$

Proof. Fix $\tau \in \mathfrak{T}_T$ and $t_m \in \tau$. For $t_m \le s \le T$ let

$$A_s^\tau = A_s \stackrel{\text{def}}{=} \left\{ \sup_{t_m \le t \le s} \|\mathfrak{X}_t^\tau - \mathfrak{X}_{t_m}^\tau\| > 3\varepsilon \right\}$$

then

$$A_s = \left\{ \sup_{t_m \le t \le s \ ,\ t \in \mathbb{Q}} \|\mathfrak{X}_t^\tau - \mathfrak{X}_{t_m}^\tau\| > 3\varepsilon \right\}$$

since trajectories are continuous. Then for $s_1 < s_2$ we have $A_{s_1} \subseteq A_{s_2}$ and moreover

$$\bigcup_{s_1 < s_2} A_{s_1} = A_{s_2}$$

again using the fact that trajectories are continuous; hence we obtain left-continuity

$$\sup_{s_1 < s_2} \mathbb{P}(A_{s_1}) = \lim_{s_1 \to s_2 -} \mathbb{P}(A_{s_1}) = \mathbb{P}(A_{s_2}) \quad .$$

As noted in Remark 3.13 if $\hat t > 0$ (and not necessarily $\hat t \in \mathbb{Q}$), if we add $\hat t$ to τ and obtain $\hat\tau = \tau \cup \{\hat t\}$ then

$$\mathfrak{X}_t^{\hat\tau} = \mathfrak{X}_t^\tau \quad \forall t \le \hat t \quad ,$$

so

$$A_t^{\hat\tau} = A_t^\tau \quad \forall t \le \hat t \quad ,$$

but then we can apply the previous Lemma and the above left-continuity to say that

$$\mathbb{P}(A_t^{\hat\tau}) = \mathbb{P}(A_t^\tau) \le \frac{c_{10}}{\varepsilon^4} g(t - t_m) \quad .$$

□

We recall this other fundamental Lemma (that is key to Theorem 8.3 in [5]).

Lemma 5.22. *Suppose E is a normed vector space, $I = [a, b]$; let $\eta, \varepsilon, v > 0$ with $v \in \mathbb{N}$; suppose μ is a probability measure on the space $C = C(I; E)$, let $a = s_0 < s_1 < \dots s_v = b \in I$ with*

$$(s_{i+1} - s_i) \ge \eta \quad \textit{for} \quad i = 2, \dots v - 2 \tag{5.36}$$

then

$$\mu\{x \in C : \omega(x, \eta) \ge 3\varepsilon\} \le \sum_{i=0}^{v-1} \mu \left\{ x \in C : \sup_{s_i \le s < s_{i+1}} \|x(s) - x(s_i)\|_E \ge \varepsilon \right\} \tag{5.37}$$

(For the proof, see the Corollary after Theorem 8.3 in [5]). Note that the inequality (5.36) need not hold for $i = 1, i = v - 1$.

5.2.4 Proof of 5.5, step 2

Now that we have proved the powerful Lemmas, we prove the second hypothesis in Theorem 4.3, that is eqn. (4.2); to this end, we fix $\varepsilon_0 > 0, \varepsilon_1 > 0$; there is an $\eta > 0$ with $\eta \in \mathbb{Q}, \eta < c_t, \eta < T/2$ such that

$$9^4 \lceil T/\eta \rceil c_{10} \frac{g(\eta)}{\varepsilon_0^4} < \varepsilon_1 \tag{5.38}$$

where g was defined in eqn. (5.31) in Lemma 5.18; eqn. (4.2) will be satisfied with $A = \mathfrak{T}_T$ and

$$\alpha_0 = \tau_0 = \{\eta i : 0 \le i < v\} \cup \{T\} \quad . \tag{5.39}$$

where $v = \lceil T/\eta \rceil$.

Let $\varepsilon = \varepsilon_0/9$. Define for convenience $s_i = \eta i$ (that are equispaced for $i < v$) while $s_v = T$. Consider any $\tau \supseteq \tau_0$; let $\mathfrak{X}^\tau$ be a process; by (5.37)

$$\mathbb{P}\big\{\omega(\mathfrak{X}^\tau,\eta) \geq 9\varepsilon\big\} \leq \sum_{i=0}^{v-1} \mathbb{P}\left\{\sup_{s_i \leq t < s_{i+1}} \|\mathfrak{X}^\tau(t) - \mathfrak{X}^\tau(s_i)\|_H \geq 3\varepsilon\right\} \tag{5.40}$$

For the terms in the sum in (5.40) we use our version (5.35) of Etemadi's estimate to obtain

$$\mathbb{P}\big\{\omega(\mathfrak{X}^\tau,\eta) \geq 9\varepsilon\big\} \leq \sum_{i=0}^{v-1} \mathbb{P}\left\{\sup_{s_i \leq t < s_{i+1}} \big\|\mathfrak{X}^\tau_t - \mathfrak{X}^\tau_{s_i}\big\|_H > 3\varepsilon\right\} \leq v\frac{c_{10}}{\varepsilon^4} g(\eta) < \varepsilon_1 \tag{5.41}$$

by (5.38). So we have satisfied the second hypothesis of Theorem 4.3, in the form expressed in eqn. (4.3).

This concludes the proof of Theorem 5.5.

6 Results on manifolds

6.1 Hypotheses for manifolds

We again define, for $v \in K(H)$, $\|v\|_K \stackrel{\text{def}}{=} \|K^{-1}v\|_H$ as in Definition 5.3; similarly for scalar products. The following Theorem uses the following hypotheses on the manifold M and its embedding in H. Let $\exp_x v$ be the exponential mapping of M. For convenience we denote by $P_x : H \to T_xM$ the orthogonal projection $P_x = \pi_{T_xM}$.

Hypotheses 6.1. We assume what follows.

1. The manifold M is isometrically embedded in the Hilbert space H and it is a closed subset of it.
2. The second fundamental form of the embedding of the manifold M is uniformly bounded.

We suppose that there exists a compact operator K satisfying the requisites in the previous section 5.1, and constants $c_e, c_p > 0$, such that:

3. $$\mathbb{P}\{\mathfrak{X}_0 \in M\} = 1$$
   and
   $$\mathbb{E}[\|\mathfrak{X}_0\|_K^4] < \infty \quad . \tag{6.1}$$
4. If $x \in M \cap K(H)$ and $v \in K(H)$ then $P_x v \in K(H)$
5. and
   $$\|P_x v\|_K \leq c_p \left(\|v\|_K + \|x\|_K \|v\|_H\right) \quad ; \tag{6.2}$$
6. if $x \in M \cap K(H)$ and $v \in T_xM \cap K(H)$ then $\exp_x v \in M \cap K(H)$
7. and
   $$\|(\exp_x v) - (v + x)\|_K \leq c_e \Big(\|x\|_K \|v\|_H^2 + \|v\|_K \|v\|_H\Big) e^{c_3 \|v\|_H} \quad . \tag{6.3}$$

The second hypothesis can be reformulated as follows.

Proposition 6.2. *The following are equivalent:*

1. *The second fundamental form of the embedding of manifold M is uniformly bounded.*
2. $\exists c_e > 0$ *such that* $\forall x \in M$, $\forall v \in T_xM$,
   $$\|v\|_H \leq 1/c_e \Rightarrow \|(\exp_x v) - (v + x)\|_H \leq c_e \|v\|_H^2 \quad . \tag{6.4}$$

6.2 Tightness of random walk

Theorem 6.3. *Consider the random walks $\mathfrak{X}^\tau$ defined as in Section 3.5.1; restrict each $\mathfrak{X}^\tau_t$ to $t \in I = [0,T]$. Under Hypotheses 6.1, these $\mathfrak{X}^\tau$, for $\tau \in \mathfrak{T}_T$, are a tight family in $C(I;M)$.*

Since this Theorem is proved using Theorem 5.5, then all corollaries of the latter hold also for the former. We have moreover this result.

Corollary 6.4. *Any limit point $\mathfrak{X}$ of $\mathfrak{X}^\tau$ is a process taking values in $C(I;M)$ a.s.*

Proof. Since $M \subset H$ was assumed to be closed, then $C(I;M)$ is a closed subset of $C(I;H)$; by construction $\mathfrak{X}^\tau_t \in M$ for all $t \in I$, hence

$$\mathbb{P}(\mathfrak{X}^\tau \in C(I;M)) = 1$$

so by Alexandrov's Theorem [12]

$$\mathbb{P}(\mathfrak{X} \in C(I;M)) = 1 \quad .$$

□

Remark 6.5. Nothing is specifically *"infinite dimensional"* in this approach: this theorem can be applied to finite dimensional manifolds as well. Recall that, by Nash embedding theorems, any finite dimensional Riemannian manifold can be isometrically embedded in $H = \mathbb{R}^N$; and in this case we set K to be the identity; moreover (6.2) is trivially true. We then require that

$$\mathbb{P}\{\mathfrak{X}_0 \in M\} = 1 \ \ , \ \ \mathbb{E}[|\mathfrak{X}_0|^4] < \infty \ ; \tag{6.5}$$

then we require the bound on the second fundamental form, that implies (6.4) that in turn implies (6.3) (see Remark 5.16): under this conditions Theorem 6.3 holds.

Proof of Theorem 6.3. As in eqn. (3.5) in Section 3.5.1 we define

$$D(x,y,t,s) \stackrel{\text{def}}{=} \exp_x\left(\sqrt{s} P_x y\right) - x \qquad \text{(seen in (3.5))}$$

and we define

$$L(x,t)v = P_x v \tag{6.6}$$

In this way, the random walk on the manifold can be seen as a special case of a random walk in H. We recall the following hypothesis 7 for Theorem 5.5:

$$\|D(x,v,t,s)\|_H \le c_d \sqrt{s} \|v\|_H \ \ , \qquad \text{(seen in (5.5))}$$

$$\sqrt{s}\|v\| \le 1/c_d \Rightarrow \|D(x,v,t,s) - \sqrt{s} L(x,t)v\|_H \le c_d s \|v\|_H^2 \quad . \qquad \text{(seen in (5.6))}$$

The first one, when $c_d \ge 1$, is true for any embedded manifold, since the length of a curve is less than the distance between its end points, and $\|P_x v\|_H \le \|v\|_H$. For the second one, for $\sqrt{s}\|v\| \le 1/c_e$ we write

$$\|D(x,v,t,s) - \sqrt{s}L(x,t)v\|_H = \|\exp_x(\sqrt{s}P_x v) - (\sqrt{s}P_x v + x)\|_H \le c_e s \|P_x v\|_H^2 \le c_e s \|v\|_H^2$$

using (6.4) from Proposition 6.2.

For hypothesis 6 for Theorem 5.5: substitute in (6.3) to obtain

$$\|(\exp_x P_x v) - (P_x v + x)\|_K \le c_e \Big(\|x\|_K \|P_x v\|_H^2 + \|P_x v\|_K \|P_x v\|_H \Big) e^{c_3 \|P_x v\|_H} \quad .$$

then we use $\|P_x v\|_H \le \|v\|_H$ again, and we use (6.2) so

$$\|\exp_x(P_x v) - (P_x v + x)\|_K \le c_e \|x\|_K \|v\|_H^2 e^{c_3\|v\|_H} + c_e c_p \left(\|v\|_K + \|x\|_K \|v\|_H\right) \|v\|_H e^{c_3 \|v\|_H} =$$

$$= \Big(c_e(1+c_p) \|x\|_K \|v\|_H^2 + c_e c_p \|v\|_K \|v\|_H \Big) e^{c_3\|v\|_H}$$

then replacing $\sqrt{s}v$ for v this last satisfies (5.4) with $c = c_e(1+c_p)$. So Theorem 6.3 can be straightforwardly seen as a corollary of Theorem 5.5. □

[12] In the version in Theorem 3.8.2 in [6], or Theorem 3.5 in [22]; see Theorem 4.5 in [18] for convenience.

7 Results on Stiefel Manifolds

In the following two sections we will show that the above hypotheses 6.1 are satisfied when M is the Stiefel manifold: so the family of random walks is tight.

We recall that $H = V^p$. We will use these definitions with $E = H$ or $E = H^2$.

Definition 7.1. If E is a vector space with a scalar product, we agree that, for $x, v \in E^p$, $A = x^\top v$ is the $p \times p$ matrix defined by

$$A_{i,j} = \langle x_i, v_j \rangle_E \quad .$$

We also agree that, given $x \in E^p$ and $A \in \mathbb{R}^{p \times p}$ the right product

$$y = xA$$

is the vector $y \in E^p$

$$y_i = \sum_{j=1}^{p} x_j A_{j,i} \quad .$$

7.1 Probabilities on Stiefel Manifolds

When $M = \mathbf{St}(p, V)$ is a Stiefel Manifold, it will be convenient to build the probabilistic infrastructure in Sec. 3.2 in this specific way.

Suppose that $\tilde{\gamma} = N(0, \tilde{Q})$ is a centered non-degenerate Gaussian measure in the separable Hilbert space V. We will then define the operator $Q : H \to H$ by tensor product

$$\langle x, Qy \rangle_H = \sum_{i=1}^{p} \langle x_i, \tilde{Q} y_i \rangle_V$$

so $\gamma = N(0, Q)$ is a centered non-degenerate Gaussian measure in the separable Hilbert space H, given by the measure product

$$\gamma = \tilde{\gamma} \otimes \ldots \otimes \tilde{\gamma}$$

Equivalently, if we consider $x \in H$ as a r.v. with distribution γ, then the columns of $x \in H$ will be independent r.v. each with distribution $\tilde{\gamma}$.

Proposition 7.2. *Given $A \in O(p)$, the action*

$$A : H \to H \quad , \quad x \mapsto xA$$

maps identically

$$\gamma = A_\sharp \gamma$$

the probability γ to itself.

7.2 Tightness operator

As noted in Remark 5.2, in the space V starting from $\tilde{Q}$ we can define a compact operator $\tilde{K} : V \to V$ such that $\tilde{K}^{-1} \tilde{Q} \tilde{K}^{-1}$ is still trace class. We eventually define $K : H \to H$ as

$$y = Kx \quad \text{when} \quad y_i = \tilde{K} x_i \tag{7.1}$$

So K commutes with the right multiplication by matrixes

$$(Kx)A = K(xA) \tag{7.2}$$

defined in Definition 7.1.

7.3 Tightness in Stiefel Manifolds

Theorem 7.3. *Suppose that $Y_t \sim N(0, Q)$ as defined in previous Section 7.1. Consider random walks $\mathfrak{X}^\tau$ defined as in Section 3.5.1 when M is a Stiefel Manifold, having*

$$\mathbb{P}\{\mathfrak{X}_0 \in M\} = 1 \ , \ \mathbb{E}[\|\mathfrak{X}_0\|_K^4] < \infty$$

and restrict each $\mathfrak{X}_t^\tau$ to $t \in I = [0, T]$: these $\mathfrak{X}^\tau$, for $\tau \in \mathfrak{T}_T$, are a tight family in $C(I; M)$.

In the following sections we will indeed show that all Hypotheses 6.1 are satisfied in Stiefel Manifolds.

Since this Theorem is proved using Theorem 5.5, and this latter using Theorem 6.3, then all corollaries of 6.3 and 5.5 will hold also for 7.3.

We have moreover this result.

Corollary 7.4. *The law of $\mathfrak{X}^\tau$ is invariant for right actions of $A \in O(p)$, and so this is true for any limit point $\mathfrak{X}$ of $\mathfrak{X}^\tau$. So the above can be interpreted as a result for Grassmann Manifolds as well.*

7.3.1 More on geodesics

The results from [12] still hold, with minor adjustments in notation.

Lemma 7.5. *Given $x \in M = \mathbf{St}(p, V)$ and $v \in H$, we have that $v \in T_x M$ iff $x^\top v$ is an asymmetric matrix.*

Given $x \in M = \mathbf{St}(p, V)$ and $A \in O(p)$ orthogonal matrix, then $xA \in M$. The action $x \mapsto xA$ is an isometry in H and hence in M.

Proposition 7.6 (Critical geodesics in $\mathbf{St}(p, V)$)**.** *Let $\mathbf{St}(p, V)$ be endowed with the induced metric from V^p. Let $\gamma : [0, 1] \to \mathbf{St}(p, V)$ be a path. Then the geodesic equation is $\ddot{\gamma} + \gamma(\dot{\gamma}^\top \dot{\gamma}) = 0$. Solutions to the geodesic equation exist for all time and are given by*

$$(\gamma(t)e^{At}, \dot{\gamma}(t)e^{At}) = (\gamma(0), \dot{\gamma}(0)) \exp t \begin{pmatrix} A & -S \\ \mathbb{I} & A \end{pmatrix} \tag{7.3}$$

where $\mathbb{I}$ is the $p \times p$ identity matrix and $A = \gamma(0)^\top \dot{\gamma}(0)$, $S = \dot{\gamma}(0)^\top \dot{\gamma}(0)$.

Note that A is asymmetric and S is symmetric; moreover $A = \gamma(t)^\top \dot{\gamma}(t)$, $S = \dot{\gamma}(t)^\top \dot{\gamma}(t)$ are constant along the geodesic; and $e^{At} \in O(p)$.

Further properties of infinite dimensional Stiefel and Grassmann manifolds are discussed in [15]. In particular it is proven that any two points in those manifolds are connected by a minimal length geodesic.

We will now add more analysis to achieve the desired results.

Remark 7.7. Starting from the geodesic equation (7.3), let $\lambda \in \mathbb{R}, \lambda \neq 0$, having fixed $x \in M, v \in T_x M$ and setting $(\gamma(0), \dot{\gamma}(0)) = (x, v)$, defining $A = x^\top v$, $S = v^\top v$ as above, we multiply as follows

$$(\gamma(t)e^{At}, \lambda\dot{\gamma}(t)e^{At}) = (x, \lambda v) \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \lambda^{-1}\mathbb{I} \end{pmatrix} \exp \left(t \begin{pmatrix} A & -S \\ \mathbb{I} & A \end{pmatrix} \right) \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \lambda\mathbb{I} \end{pmatrix}$$

so

$$(\gamma(t)e^{At}, \lambda\dot{\gamma}(t)e^{At}) = (x, \lambda v) \exp t \begin{pmatrix} A & -\lambda S \\ \lambda^{-1}\mathbb{I} & A \end{pmatrix} = \tag{7.4}$$

$$= (x, \lambda v) \exp t\lambda^{-1} \begin{pmatrix} \lambda A & -\lambda^2 S \\ \mathbb{I} & \lambda A \end{pmatrix}$$

We can use this relation as follows. Let now $\theta = \|v\|, \hat{v} = v/\theta$, $\hat{A} = x^\top \hat{v}$, $\hat{S} = \hat{v}^\top \hat{v}$ then setting $\lambda = 1/\theta$

$$(\gamma(t)e^{t\theta\hat{A}}, \theta^{-1}\dot{\gamma}(t)e^{t\theta\hat{A}}) = (x, \hat{v}) \exp t\theta \begin{pmatrix} \hat{A} & -\hat{S} \\ \mathbb{I} & \hat{A} \end{pmatrix} \tag{7.5}$$

this formula decouples v into the *initial direction* $\hat{v}$ and the *initial speed* θ.

7.3.2 Estimates

Define $\|v\|_K$ as in 5.3.

Lemma 7.9. *Recall the Definitions 7.1. For $W \in \mathbb{R}^{k\times k}$ we use the norm*

$$|W| = \sqrt{\sum_{i,j=1}^{k} |W_{i,j}|^2}$$

and remark that

$$|WV| \leq |V|\,|W| \quad ;$$

we will use it with $k = p$ or $k = 2p$; if E is a vector space with a scalar product, then for $v, w \in E$ we have

$$|v^\top w| \leq \|v\|_E \|w\|_E$$

and for $v \in E^k$

$$\|vW\|_{E^k} \leq \|v\|_{E^k} |W|$$

by Cauchy-Schwarz inequality.

Lemma 7.10. *Consider the orthogonal projection $\pi_T : H \to H$ to a hyperplane*

$$T = \{x \in H : \forall i \leq v, \langle w_i, x\rangle_H = 0\}$$

orthogonal to $w_1, \dots w_v \in H$, where those vectors are orthogonal but not orthonormal: then

$$\pi_T v = v - \sum_{i=1}^{v} w_i \frac{\langle w_i, v\rangle_H}{\|w_i\|_H^2} \quad . \tag{7.6}$$

we immediately note that if v, w_i are in a vector subspace E of H, then $\pi_T v$ will be in the same E. Starting from (7.6) *we estimate*

$$\|\pi_T v\|_K \leq \|v\|_K + \sum_{i=1}^{v} \|w_i\|_K \frac{\|v\|_H}{\|w_i\|_H} \quad .$$

The tangent plane $T = T_x M$ to the Stiefel Manifold is such a plane, with

- $$w_i = (0, \dots 0, x_i, 0, \dots 0)$$

  containing the i-th column of x in position i-th; this for $i = 1, \dots p$;
- *and then for $i = p + 1, \dots p(p + 1)/2$*

  $$w_i = (0, \dots 0, x_h, 0, \dots 0, -x_k, 0, \dots 0)$$

  containing the h-th column of x in position k-th and vice versa, and with a minus sign;

so by the diagonal structure (7.1) *of* K *and by* (7.6) *above we obtain this: if* $x \in M \cap K(H)$ *and* $v \in K(H)$ *then* $\pi_{T_xM} v \in K(H)$.

Moreover such w_i *are mutually orthogonal; and* $\|w_i\|_H = 1$ *for* $i = 1, \dots p$, *while* $\|w_i\|_H = 2$ *otherwise; while* $\|w_i\|_K \le \|x\|_K$ *in all cases; so in conclusion*

$$\|\pi_{T_xM} v\| \le c_p(\|v\|_K + \|x\|_K \|v\|_H) \tag{7.7}$$

for a $c_p > 0$ *independent of* x, v. *This proves estimate* (6.2) *in Hypotheses 6.1.*

The above suggests that (6.2) in Hypotheses 6.1 may hold for other manifolds, as long as the embedding in H has finite codimension.

Lemma 7.11. *Let* $\tilde{K} : V \to V$ *a linear continuous injective operator. Recall that* $H = V^p$ *and we defined* $K : H \to H$ *in* (7.1) *as*

$$y = Kx \quad \text{when} \quad y_i = \tilde{K} x_i \quad .$$

So K *is a linear continuous operator and commutes with the right multiplication by matrixes*

$$(Kx)A = K(xA) \quad . \qquad \text{(as defined in (7.2))}$$

There is a constant $c > 0$ *such that for all* $x \in M \cap K(H), v \in T_xM \cap K(H)$ *and the geodesic with*

$$(\gamma(0), \dot\gamma(0)) = (x, v)$$

we have

$$(\|\gamma(1) - x - v\|_K + \|\dot\gamma(1) - v + Sx\|_K) \le c\Big(\|x\|_K \|v\|_H^2 + \|v\|_K \|v\|_H\Big) e^{c\|v\|_H} \quad . \tag{7.8}$$

Note that c *does not depend on* K *but only on* p.

This proves (6.3) *in in Hypotheses 6.1.*

Proof. Fix $x \in M, v \in T_xM$ and set $\theta = \|v\|, \hat v = v/\theta$. We will use the formula seen in eqn. (7.5) with $t = 1$. We define

$$\hat A = x^\top \hat v \quad , \quad \hat S = \hat v^\top \hat v \quad ,$$

$$Z = \begin{pmatrix} \hat A & -\hat S \\ \mathbb{I} & \hat A \end{pmatrix} \quad , \quad B = \begin{pmatrix} -\hat A & 0 \\ 0 & -\hat A \end{pmatrix} \quad , \quad \Theta = \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \theta \mathbb{I} \end{pmatrix}$$

so the formula (7.5) becomes

$$(\gamma(t), \dot\gamma(t)) = (x, \hat v) \exp(t\theta Z) \exp(t\theta B) \Theta \tag{7.9}$$

then

$$(\gamma(t), \dot\gamma(t)) = (x, \hat v) \left(\sum_{i \ge 0, j \ge 0} (t\theta)^{i+j} \frac{Z^i}{i!} \frac{B^j}{j!} \right) \Theta =$$

$$(x, \hat v) \left(\mathbb{I} + t\theta(Z + B) + \sum_{i,j,i+j \ge 2} (t\theta)^{i+j} \frac{Z^i}{i!} \frac{B^j}{j!} \right) \Theta =$$

$$= (x + tv, v - t\theta^2 xS) + (x, \hat v) \left(\sum_{i,j,i+j \ge 2} (t\theta)^{i+j} \frac{Z^i}{i!} \frac{B^j}{j!} \right) \Theta \tag{7.10}$$

Now

$$|\hat A| \le p \quad , \quad |\hat S| \le 1 \quad , \quad |\hat B| \le 2p \quad , \quad |Z| \le \sqrt{3p + 1}$$

and setting $t = 1$

$$\sqrt{\|\gamma(1) - x - v\|_K^2 + \|\dot\gamma(1) - v + \theta^2 x\hat S\|_K^2} \le \sqrt{\|x\|_K^2 + \|\hat v\|_K^2} \left| \sum_{i,j,i+j\ge 2} \theta^{i+j} \frac{Z^i}{i!}\frac{B^j}{j!} \right| |\Theta| \le$$

$$\sqrt{\|x\|_K^2 + \|\hat v\|_K^2} \sum_{i,j,i+j\ge 2} |\theta|^{i+j} \frac{|Z|^i}{i!}\frac{|B|^j}{j!} \sqrt{(1+p\theta^2)} \le$$

$$\sqrt{\|x\|_K^2 + \|\hat v\|_K^2}\theta^2 g(\theta)$$

where we are using the fact (7.2) that left multiplication by K and right multiplication by a matrix are associative; and

$$g(s) = \sum_{i,j,i+j\ge 2} s^{i+j-2} \frac{(3p+1)^{i/2}}{i!}\frac{(2p)^j}{j!} =$$

$$\sum_{2\le n}\sum_{0\le k\le n} \frac{s^{n-2}}{n!}\left(\binom{n}{k}(3p+1)^{k/2}\binom{n}{n-k}(2p)^{n-k}\right) = \sum_{2\le n}\frac{s^{n-2}}{n!}\left(\sqrt{3p+1}+2p\right)^n =$$

$$= \frac{1}{s^2}(\exp a - 1 - a) \quad \text{with} \quad a = s\left(\sqrt{3p+1}+2p\right) \quad \square$$

The first and second hypothesis in 6.1 are obviously true for Stiefel Manifolds: indeed the curvatures and second fundamental form are uniformly bounded, since Stiefel Manifolds are homogeneous space. Nonetheless we can provide this estimate that satisfies (6.4).

Corollary 7.14. *There is a constant $c > 0$ such that for all $x \in M, v \in T_x M$ and the geodesic with*

$$(\gamma(0), \dot\gamma(0)) = (x, v)$$

we have

$$\|\gamma(1) - x - v\|_H \le c \min\{\|v\|_H, \|v\|_H^2\} \quad .$$

Proof. We note that the Stiefel Manifold $\mathbf{St}(p, V)$ has diameter d, so that for $v \ge d$ we can estimate

$$\|\gamma(1) - x - v\|_H \le d + \|v\|_H$$

while for $v \le d$ we use the above lemma 7.11 with K being the identity, recalling that $\|x\|_K = \sqrt{p}$ in this case. $\square$

8 Future Developments

We now know that, under appropriate hypotheses, the random walks $\mathfrak{X}^\tau$ have narrow limit points $\mathfrak{X}$ when the partition τ becomes finer and finer; these $\mathfrak{X}$ are random functions in $C(\mathbb{R}^+; S)$ with $S = H$ or $S = M$ an embedded manifold.

There are multiple questions left unanswered, material for future research.

- Do the limit points enjoy some standard property? It seems plausible that they may enjoy some kind of *Markov Property*, for example.
- Under which additional hypotheses can we say that there is an unique limit point?
- Can we then characterize the limit points as solutions to a kind of SDE?

  (These two questions are in synergy).
- Can we expand the results to more general cases of random walks, for example, where the constants in the hypothesis are not “uniform” but rather they may grow (*e.g.* be bounded by a function of the distance from a given point)?

- Consequently, are there other infinite dimensional manifolds where the present results hold true?

All the above questions are starting point for future research.

A Useful Lemmas

In this section we have collected the technical Lemmas used here and there in the paper.

Lemma A.1. *For E space with scalar product and $v, w \in E$,*

$$|v + w|^2 \le 2|v|^2 + 2|w|^2 \quad .$$

Proof.

$$|v + w|^2 \le (|v| + |w|)^2 \le 2|v|^2 + 2|w|^2$$

□

In the following *monotonic* means *monotonically weakly increasing* that is $s \le t \Rightarrow g(s) \le g(t)$.

Lemma A.2. *Let*

$$t_0 = 0 < t_1 < t_2 < \dots$$

and let $\delta_n = t_{n+1} - t_n$. Suppose b_n is a real valued sequence with $b_n \ge \beta$ for all n. Suppose that

$$\varphi = \varphi(t, x) : [0, \infty) \times [\beta, \infty) \to [0, \infty)$$

is a continuous non negative function, such that $\varphi(\cdot, x)$ and $\varphi(t, \cdot)$ are monotonic. Let $f : [0, \infty) \to [\beta, \infty]$ be a solution of

$$\begin{cases} f'(t) = \varphi(t, f(t)) \\ f(0) = b_0 \end{cases} \tag{A.1}$$

(possibly $f(t) = \infty$ for large t). If

$$b_{n+1} \le b_n + \varphi(t_n, b_n)\delta_n \tag{A.2}$$

holds then

$$b_n \le f(t_n) \quad . \tag{A.3}$$

Proof. Proof by induction. Note that f is monotonic since $f' \ge 0$ but then it is convex since f' is monotonic.

$$b_{n+1} \le b_n + \varphi(t_n, b_n)\delta_n \le f(t_n) + \varphi(t_n, f(t_n))\delta_n = f(t_n) + f'(t_n)\delta_n \le f(t_{n+1})$$

□

Note that indeed (A.2) can be rewritten as

$$\frac{b_{n+1} - b_n}{t_{n+1} - t_n} \le \varphi(t_n, b_n) \quad .$$

Lemma A.3. *Let*

$$t_0 = 0 < t_1 < t_2 < \dots$$

and let $\delta_n = t_{n+1} - t_n$; fix $c_7 > 0, c_8 > 0$. Suppose b_n is a real valued sequence with $b_0 = 0, b_n \ge 0$ that satisfies

$$b_{n+1} \le b_n + 2\delta_n(c_7\sqrt{b_n}\sqrt{t_n} + c_8 t_n) \quad ; \tag{A.4}$$

then

$$b_n \le (c_7 + 2c_8) g(t_n) \;\; \textit{with} \;\; g(t) = \frac{e^{c_7 t} - 1 - c_7 t}{c_7^2} \quad .$$

Moreover, set

$$\hat\varepsilon = \frac{2}{c_7 + 2c_8} \quad ,$$

then

$$\forall n \; , \; t_n \le \hat\varepsilon \Rightarrow b_n \le t_n^2 (c_7 + c_8) \tag{A.5}$$

and note that $b_0 = b_1 = 0$.

Proof. Consider a solution of the ODE

$$\begin{cases} f'(t) = 2c_7 \sqrt{f(t)}\sqrt{t} + 2c_8 t \\ f(0) = 0 \end{cases} \quad ; \tag{A.6}$$

(that is (A.1) for this special case). Since

$$\sqrt{ab} \le \frac{a+b}{2}$$

then

$$f'(t) \le c_7 f(t) + (c_7 + 2c_8) t$$

substituting $f(t) = g(t) e^{c_7 t}$ and with some calculations we obtain

$$f(t) \le (c_7 + 2c_8) g(t) \;\; \text{with} \;\; g(t) = \frac{e^{c_7 t} - 1 - c_7 t}{c_7^2} \quad .$$

Since

$$f'(t) \ge 2c_8 t$$

then

$$f(t) \ge c_8 t^2 \tag{A.7}$$

in particular for any solution we have $f(t) > 0$ and $f'(t) > 0$ for $t > 0$. We have $f'(0) = 0$ so $f(t) \le t$ for $t \le \varepsilon$ with ε small; more precisely, note that $g(t)$ is convex and increasing and $g(0) = g'(0) = 0$ so we set ε to be the unique positive solution of

$$(c_7 + 2c_8) g(t) = t \quad ;$$

moreover $g(t) \ge t^2/2$ so we know that

$$\varepsilon \ge \frac{2}{c_7 + 2c_8} \quad .$$

Now we set

$$\hat s = \sup\{s \ge 0 : t \le s \Rightarrow f(t) \le t\}$$

note that $\varepsilon \le \hat s \le 1/c_8$; for $t \in [0, \hat s]$

$$f'(t) \le 2(c_7 + c_8) t$$

so

$$f(t) \le (c_7 + c_8) t^2 \quad . \tag{A.8}$$

□

Lemma A.5. *Let*

$$t_0 = 0 < t_1 < t_2 < \dots$$

and let $\delta_n = t_{n+1} - t_n$. *Suppose* b_n *is a real valued sequence. If, for* $c_7 > 0, c_8 \geq 0, c_9 \geq 0$,

$$b_{n+1} \leq b_n(1 + c_7\delta_n) + \delta_n(c_8 + c_9 t_n) \tag{A.9}$$

holds then

$$\begin{aligned} b_n \leq & b_0 e^{c_7 t_n} + (e^{c_7 t_n} - 1)(c_8/c_7 + c_9/c_7^2) - c_9 t_n/c_7 = \\ & e^{c_7 t_n}\Big(b_0 + c_8/c_7 + c_9/c_7^2\Big) - \Big((c_8 + c_9 t_n)/c_7 + c_9/c_7^2\Big) \quad . \end{aligned} \tag{A.10}$$

Proof. Indeed (A.9) can be rewritten as

$$\frac{b_{n+1} - b_n}{t_{n+1} - t_n} \leq c_7 b_n + c_8 + c_9 t_n$$

and the associated differential equation is

$$f'(t) = c_7 f(t) + c_8 + c_9 t$$

that has solution

$$f(t) = e^{c_7 t} f(0) + (e^{c_7 t} - 1)(c_8/c_7 + c_9/c_7^2) - c_9 t/c_7$$

so this proves the result. □

B Proofs

Proof of Proposition 2.3. This proof comes from [23]. Suppose that there is convergence in probability $W_1^n \to W_1$; consider the equality

$$\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}} = \frac{1}{\sqrt{2}} \frac{S_{2n} - S_n}{\sqrt{n}} - \left(1 - \frac{1}{\sqrt{2}}\right) \frac{S_n}{\sqrt{n}}$$

then the LHS would converge to the zero constant in probability , whereas on the RHS the random variables

$$\frac{S_n}{\sqrt{n}} \quad \text{and} \quad \frac{S_{2n} - S_n}{\sqrt{n}}$$

are independent and both converge narrowly to $N(0,1)$. □

Proof of Proposition 3.6. If $c = 0$ then

$$\mathbb{E}[g(\delta\|Y\|)] = \delta^\alpha \mathbb{E}\left[\|Y\|^\alpha\right]$$

so we set

$$\tilde{c} = \mathbb{E}\left[\|Y\|^\alpha\right] \quad .$$

Otherwise we set

$$\tilde{c} = \mathbb{E}\left[\|Y\|^\alpha e^{c c_t \|Y\|}\right] \quad ,$$

so

$$\mathbb{E}[g(\delta\|Y\|)] = \delta^\alpha \mathbb{E}\left[\|Y\|^\alpha e^{c\delta\|Y\|}\right] \leq \delta^\alpha \mathbb{E}\left[\|Y\|^\alpha e^{c c_t \|Y\|}\right] = \tilde{c}\delta^\alpha$$

□

We recall Proposition 1.13 from [9].

Proposition B.1. *Let* $Y = N(0, Q)$ *and*

$$\lambda_1 = \max_{\|x\| \le 1} \langle x, Qx \rangle_H$$

be the highest eigenvalue of Q*. Then for* $0 < \varepsilon < 1/\lambda_1$

$$\int_H e^{\varepsilon \|x\|^2/2} \, \mathrm{d}\gamma(x) = \frac{\exp\left(-\frac{1}{2}\langle a, (1-\varepsilon Q)^{-1} a\rangle_H\right)}{\sqrt{\det(1-\varepsilon Q)}} \tag{B.1}$$

whereas for $\varepsilon \ge 1/\lambda_1$ *the integral is infinite.*

Proof of Proposition 3.8. Set $c_t = 1$ for simplicity. By the previous proposition, for any $\lambda > 0$,

$$\mathbb{E}[e^{\lambda \|Y\|_H}] < \infty \quad .$$

For $k \in \mathbb{N}$ and $a > 0$ we have $a^k s^k \le k! e^{as}$ hence choosing $k = \lceil \alpha \rceil$,

$$s^\alpha e^{sc} \le \frac{k!}{a^k} e^{s(a+c)}$$

so again we define

$$\tilde{c} = \mathbb{E}[g(\|Y\|)] < \infty \quad ;$$

and we proceed as in the above proof of Proposition 3.6. □

Proof of Proposition 3.15. By contradiction, suppose there is; up to substituting $f(n)$ with $\bigcup_{j=0}^n f(j)$ we can suppose that f is monotonic. Let $c_t = 1$ for simplicity. We build iteratively $\tau \in \mathfrak{T}$ such that $\forall n, \tau \not\subseteq f(n)$, in this way. We will build a (non decreasing) sequence $n_m \in \mathbb{N}$ such that $n_m \to_m \infty$, and a sequence $t_0 = 0 < t_1 < \ldots \in \tau$ satisfying the requisites in Definition 3.9. Let $t_0 = 0, t_1 = 1$, $n_0 = n_1 = 0$; for $m \ge 1$ having chosen $t_m \in \tau$ and n_m, we look for $k > n_m$ such that there is a $t \in f(k) \setminus f(n_m) \wedge t \ge t_m + 1/2$;

- if there is no such k, we stop the iterative process by adding to τ an arbitrary sequence $t_{m+1} < t_{m+2} < \ldots$ with $t_{m+j} \notin \bigcup_k f(k)$ and $1/2 < t_{m+j+1} - t_{m+j} < 1$; we set $n_{m+j} = n_m + j$; all that for $j \ge 0$.
- If there is such k, t, we add to τ an arbitrary sequence

  $$t_{m+1} < t_{m+2} < \ldots < t_{m+l} = t$$

  such that

  $$1/2 < t_{m+j+1} - t_{m+j} < 1 \ \text{ for } j = 0, \ldots l-1 \ ;$$

  then we set $n_{m+1} = \ldots = n_{m+l} = k$; then we repeat the iteration using $m + l$ as the new m.

In any case we obtain that for infinitely many m there is a l such that $t_{m+l} \notin f(n_m)$. □

Acknowledgments. The author thanks Prof. Thomas G. Kurtz for the help in finding some references, and Prof. Luciano Mari for helpful discussions.

This project was supported by the `RICBASE_2022_MENNUCCI` funding.

Contents

References

[1] R. Azencott. Behavior of diffusion semi-groups at infinity. *Bull. Soc. Math. (France)*, 102:193–240, 1974.

[2] Eleonora Bardelli and Andrea C. G. Mennucci. Probability measures on infinite dimensional Stiefel manifolds. *Journal of Geometric Mechanics*, 2017.

[3] C. Bessaga. Every infinite-dimensional hilbert space is diffeomorphic with its unit sphere. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 14:27–31, 1966.

[4] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1995.

[5] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley, 2013.

[6] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[7] Leo Breiman. *Probability*. Addison-Wesley, 1968.

[8] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2014.

[9] Giuseppe Da Prato. *An introduction to infinite-dimensional analysis*. Springer, 2006.

[10] Stefano Deiala. Una metrica riemanniana sullo spazio delle curve e applicazioni. Master's thesis, Università di Pisa, 2010.

[11] N. Dunford, J.T. Schwartz, W.G. Bade, and R.G. Bartle. *Linear Operators: General theory*. Linear Operators. Interscience Publishers, 1958.

[12] A. Edelman, T. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J Matrix Analy Appl*, 20:303–353, 1998. arXiv:physics/9806030v1 (1998).

[13] N. Etemadi. On some classical results in probability theory. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 47(2):215–221, 1985.

[14] Alexander Grigor'yan. Analytic and geometric background of recurrence and non-explosion of the brownian motion on riemannian manifolds. *Bulletin (New Series) of the American Mathematical Society*, 36(2):135–249, 1999.

[15] Philipp Harms and Andrea Mennucci. Geodesics in infinite dimensional Stiefel and Grassmann manifolds. *Comptes rendus - Mathématique*, 350:773–776, 2012.

[16] E.P. Hsu. *Stochastic Analysis on Manifolds*. Graduate studies in mathematics. American Mathematical Society, 2002.

[17] Andrea Mennucci. Metrics of curves in shape optimization and analysis. In Osher and Burger [19], page 114. Lectures given at the C.I.M.E. Summer School held in Cetraro, 8-13 September 2008.

[18] Andrea C. G. Mennucci. Compendium for: Tightness of Random Walks in Infinite Dimensional Spaces and Manifolds. https://cvgmt.sns.it/paper/5750/, 2022.

[19] S. Osher and M. Burger, editors. *Level Set and PDE based Reconstruction Methods: Applications to Inverse Problems and Image Processing*. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2013. Lectures given at the C.I.M.E. Summer School held in Cetraro, 8-13 September 2008.

[20] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1415–1428, July 2011.

[21] Ganesh Sundaramoorthi, Andrea Mennucci, Stefano Soatto, and Anthony Yezzi. A new geometric metric in the space of curves, and applications to tracking deforming objects by prediction and filtering. *SIAM Journal on Imaging Sciences*, 4:109–145, 2011.

[22] N. Vakhania, V. Tarieladze, and S. Chobanyan. *Probability Distributions on Banach Spaces*. Mathematics and its Applications. Springer Netherlands, 2012.

[23] Wikipedia contributors. Law of the iterated logarithm — Wikipedia, the free encyclopedia, 2022. [Online; accessed 5-ottobre-2022].

[24] Laurent Younes. Computable elastic distances between shapes. *SIAM Journal of Applied Mathematics*, 58(2):565–586, 1998.

[25] Laurent Younes, Peter W. Michor, Jayant Shah, and David Mumford. A metric on shape space with explicit geodesics. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 19(1):25–57, 2008.