Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Representation of Markov chains by random
maps: existence and regularity conditions

by

Jürgen Jost, Martin Kell, and Christian S. Rodrigues

REPRESENTATION OF MARKOV CHAINS BY RANDOM MAPS: EXISTENCE AND REGULARITY CONDITIONS

JÜRGEN JOST, MARTIN KELL, AND CHRISTIAN S. RODRIGUES

ABSTRACT. We systematically investigate the problem of representing Markov chains by families of random maps, and what regularity of these maps can be achieved depending on the properties of the probability measures. Our key idea is to use techniques from optimal transport to select optimal such maps. Optimal transport theory also tells us how convexity properties of the supports of the measures translate into regularity properties of the maps via Legendre transforms. Thus, from this scheme, we cannot only deduce the representation by measurable random maps, but we can also obtain conditions for the representation by continuous random maps. Finally, we show how to construct random diffeomorphisms from a given Markov chain.

1. INTRODUCTION

Amongst the main concerns of Dynamics, one usually wants to decide whether asymptotic states of a given class of systems are robust under small random fluctuations. Such randomness, corresponding to natural fluctuations in physical processes, are represented by either a *Markov chain model* with localised transition or by a sequence of *random maps*. To see how they arise, consider a discrete-time system f from a given topological space M into itself. Suppose at each iteration of f we allow a small mistake of size, say, at most $\varepsilon > 0$. Then a Markov chain is defined by a family $\{p_\varepsilon(\,\cdot\,|x)\}$ of Borel probability measures, such that every $p_\varepsilon(\,\cdot\,|x)$ is supported inside the ε-neighbourhood of $f(x)$. The orbit of our dynamics subject to such small errors is thus given by sequences of independent random variables $\{x_j\}$, where each x_{j+1} has distribution $p_\varepsilon(\,\cdot\,|x_j)$. Alternatively, one could think of the orbit as being made by the iteration $x_j = g_j \circ \cdots \circ g_1(x_0)$, where each measurable g_j is picked at random ε-close, in a sense to be made more precise, from the original map f. Endowing the collection of maps $\{g_j\}$ with a probability distribution ν_ε, we say that the sequence of random maps is a *representation of that Markov chain* if for every Borel subset U

$$p_\varepsilon(U|x) = \nu_\varepsilon(\{g : g(x) \in U\}). \tag{1}$$

In fact, given any sequence of random maps, one can always find a Markov chain which is represented by this scheme [BDV05, D.4], see also [ZaH07]; the Markov chain is simply given by (1), and one only needs to check that this satisfies the Markov chain criterion. The converse problem, however, is much more subtle, as we shall see. This is exactly the subject of this paper. In other words, we investigate under which conditions imposed on the Markov chain one can obtain a representation by a random map scheme and how its regularity properties are reflected.

The study of realisations of Markov chains via random maps goes back to Blumenthal and Corson [BlC70]. They considered the case where M is a connected and locally connected compact metric space under some strong requirements on the probability measures. Let us denote by $\mathcal{P}(M)$ the space of all probability measures on M. In addition, suppose each $x \mapsto p(\,\cdot\,|x)$, acting from M to $\mathcal{P}(M)$, is continuous relative to the weak* topology on $\mathcal{P}(M)$. Then, if *for each x the support of $p(\,\cdot\,|x)$ is all of M*, they showed that it is possible to obtain a probability measure ν on the space of continuous transformations of M such that condition (1) is fulfilled. Their proof is based upon the existence of a continuous projection of the subset of $\mathcal{P}(M)$ whose support is all of M onto the space of probability measures on an interval whose support is the whole interval itself, and its continuous inverse. The assumption of full support on the probability measures is essential to assure the continuity of the maps. See for example [Kif86].

Weakening this condition, Kifer showed that if M is a Borel subset of a complete separable metric space (Polish), then any Markov chain on M can be represented by a sequence of measurable random maps [Kif86]. His idea was to use Borel measurable isomorphisms of M to Borel subsets of the unity interval, as it had previously been shown by Kuratowski. Later, Quas [Qua91] tackled the case where M is a smooth compact orientable Riemannian manifold. He showed that probability families which are absolutely continuous with respect to the normalised Riemannian measure whose density is smooth can be represented by C^∞-random maps.

Afterwards, Araújo [Ara00] showed how to construct families of C^r-diffeomorphisms on the n-torus near an unperturbed C^r-diffeomorphism. He took advantage of the parallelizability of this manifold and of its quotient by integers. Then he used natural projections to identify orthonormal vector fields from which he could build these maps C^r-close to the original one; see [Ara00, Example 1]. Using a parametrised geodesic flow, he also showed the existence of parametrised families of diffeomorphisms, around an unperturbed one, of any compact boundaryless manifold; see [Ara00, Example 2]. Nevertheless, his procedures focus on *uniformly continuous perturbations*, requiring the small noise to uniformly cover a ball of positive radius around the unperturbed diffeomorphism. Furthermore, they do not yield a family of diffeomorphisms from a given probability distribution.

More recently, Benedicks and Viana [BeV06, Example 1.7], and [BDV05, D.4] constructed random maps for small non-uniform noise. They lift implicitly the measure to the tangent space at $f(z)$, then try to transform the distribution to a fixed measure on $[0, 1]^n$. If this transformation is invertible, its inverse regular enough, and varies smoothly with respect to z, then it is possible to select a random continuous map representing the perturbation. For topological reasons these constructions may fail on manifolds with non-trivial tangent bundle. Namely, their constructions implicitly assume the existence of global cross-section of the frame bundle.

In fact, it is not clear how to choose the random maps representing a Markov chain, and there might be many possible such choices. In such a situation, a basic strategy of geometric analysis is to select the maps according to some optimization principle. This typically has the advantage that an object selected by an optimization principle is not just some solution of the problem at hand; it typically enjoys additional properties derived from the optimization, and these properties can typically be usefully exploited. This is also the strategy we adopt in the present paper. Since the maps should relate different measures, it is natural to select them by optimizing a transportation problem between those measures. Thus, in this paper we introduce techniques from optimal transport in order to tackle the representation of Markov chains under different levels of regularity of the maps. The paper is organised as follows. After presenting the main definitions in Section 2, we review the main ingredients from optimal transport theory, in Section 3, to be used in the remaining part of the paper. In Section 4, we use optimal coupling to prove Theorem A, which shows how Markov chains can be represented by measurable maps. Then, in Section 5, we use Moser's coupling to show how representation by continuous random maps arise; the content of Theorem B. In the following Section 6, we discuss the regularity of densities and how they affect the properties of the transport maps. Then in the next Section 7, we use tangent bundle lifts of the measures and certain transformations to a fixed measure to construct continuous families of probabilities in the bundles. To use similar ideas of Araújo we rely on the fact that the tangent bundle of a manifold is always contained in a (smooth) trivial bundle, what can be seen by taking an isometric (Nash) embedding $M \hookrightarrow \mathbb{R}^n$. Then we lift (in a nice way) the measures from the tangent bundle to this trivial bundle to get a measure family $\{\mu_x\}_{x \in M}$ on $\mathbb{R}^n$. Using optimal transport theory and its regularity theory we get transformations to a fixed measure varying continuously with respect to $x \in M$ so that we can select sections of this bundle varying continuously, such that the distribution at a point x represents the measure $\tilde{\mu}_x$. (Smooth) Projections to the tangent bundle and the exponential map then give us the random continuous maps. Assuming further regularity, these maps are differentiable and we obtain random maps C^1-close to f, and the unperturbed map f is a diffeomorphism, we thus obtain a family of random diffeomorphisms. Therefore, we

give a geometric condition for the representation of Markov chains by random diffeomorphisms, summarised in Theorem C. In a following up work, the methods developed here are being extended in order to study statistical and stochastic stability.

2. Some notation and definitions

In this section we set up the notation and collect some main definitions to be used throughout this paper. For a comprehensive presentation on random perturbations of dynamics, see [Kif86, Kif88]. We consider here an arbitrary manifold, say M, supposing it to be finite dimensional, equipped with some Riemannian structure, fixed once and for all, which induces a distance $d_M : M \times M \to \mathbb{R}$. We call m its normalised Riemannian volume form on M, i.e. $m(M) = 1$, and unless otherwise stated, we take absolute continuity with respect to m. As before, let us denote by $\mathcal{P}(M)$ the space of all Borel probability measures on M. As usual, $\mathcal{P}(M)$ is endowed with the weak* topology.

Regarding measurability and continuity, we recall Lusin's theorem to be used in our proofs

Theorem 2.1. [Fed69, Theorem 2.3.5]. *Let M be a locally compact metric space, N a separable metric space, and μ a Borel measure on M. Let $f : M \to N$ be a measurable map, and $A \subset M$ a measurable set with finite measure, then for each $\delta > 0$ there is a closed set $K \subset A$, such that $\mu(A\backslash K) < \delta$ and the restriction of f to K is continuous.*

2.1. **Markov chains and random maps.** Let N be another arbitrary manifold or a separable complete metric space. We shall consider families of probability measures $(\mu_x)_{x \in M}$ in $\mathcal{P}(N)$ given by measurable maps $x \mapsto \mu_x$. We speak of a continuous family of probability measures if the maps $x \mapsto \mu_x$ vary continuously from M to $\mathcal{P}(N)$ relative to the weak* topology. In many cases one has $N = M$ or $N = \mathbb{R}^n$. Such families are sometimes called *continuous Markov kernels*. Therefore, Markov chains are special Markov kernels obeying some conditional probability with localised distribution.

Similarly, we can have a more general definition for our random maps. For a probability space $(\Omega, \mathcal{A}, \nu)$, consider a measurable collection of maps $\mathcal{F} : \Omega \times M \to N$, $(\omega, x) \mapsto f_\omega(x)$. Then, we call the family $(f_\omega : M \to N)_{\omega \in \Omega}$ random measurable maps. If in addition each map in $(f_\omega : M \to N)_{\omega \in \Omega}$ is continuous, or a diffeomorphism, then we say that it is a family of *random continuous maps*, or *random diffeomorphisms*, respectively. They are also known as *random fields*. Then, condition (1) says that $(f_\omega)_{\omega \in \Omega}$ represents $(\mu_x)_{x \in M}$ if the distribution of $\omega \mapsto f_\omega(x)$ equals μ_x for all $x \in M$.

3. On optimal transport

The remaining part of this paper is based upon techniques from *optimal transport*. Our main reference is the book by Villani [Vil09]. For the sake of completeness, we sample and collect in this section the concepts to be used

along the way. Readers familiar with optimal transport may wish to skip this section and only refer back to it when needed.

The basic problem in *optimal transport*, as introduced by Monge, consisted in moving a given distribution like a pile of sand from one place to another with a minimal *cost*. There are several possible ways to generalise and tackle this problem. For example, the given mass to be transported can be thought of as a distribution in an appropriate probability space. In other words, given measurable spaces M, N, and probability measures μ in $\mathcal{P}(M)$ and ν in $\mathcal{P}(N)$, we seek for a *coupling*, or a way to connect these two measures. More generally, one has the following definition.

Definition 1. *Let (M, μ) and (N, ν) be two probability spaces. We couple μ and ν by constructing two random variables X, Y on some probability space $(\Omega, \mathbb{P})$, such that $law(X) = \mu$, $law(Y) = \nu$. The law or distribution of (X, Y) is called coupling of (μ, ν).*

In our context, μ and ν are the only laws we shall be interested in, so we choose $\Omega = M \times N$. There are several examples of couplings arising in different contexts.

The first generalisation of Monge's original problem we can think of is given in terms of *transport maps*. That is, given measurable spaces M, N, probability measures μ in $\mathcal{P}(M)$ and ν in $\mathcal{P}(N)$, we seek for measurable maps $T : M \to N$, such that for all Borel $E \subset N$ one has $\mu(T^{-1}(E)) = \nu(E)$. This is an example of a so-called *deterministic coupling*. The requirement of a transport map, however, is a strong condition, and this problem may not have a solution unless more restrictions are made. The canonical example is when μ is a Dirac measure and ν is not.

In order to avoid ill-posed problems, one alternatively should look for weak solutions of the transport problem as it has been proposed by Kantorovich. In this case, we focus on probability measures γ in $\mathcal{P}(M \times N)$, whose projections (or marginals) are μ and ν. In other words, let $\Gamma(\mu, \nu) := \mathcal{P}(M \times N)$ to shorten the notation, and consider the canonical projections $\pi_{\mathcal{P}(M)} : \Gamma(\mu, \nu) \to \mathcal{P}(M)$ and $\pi_{\mathcal{P}(N)} : \Gamma(\mu, \nu) \to \mathcal{P}(N)$, then the marginals are given by the push-forward $\pi_{\mathcal{P}(M)*}\gamma = \mu$ and $\pi_{\mathcal{P}(N)*}\gamma = \nu$. The *Kantorovich minimisation problem* consists in obtaining

$$C(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{M \times N} c(x, y) d\gamma(x, y), \qquad (2)$$

where, for a given *cost* function $c : M \times N \to [0, +\infty]$, the infimum runs over the joint probabilities γ in $\Gamma(\mu, \nu)$. The joint probability measures are called *transport plans*, the ones achieving the minimum, *optimal transport plans*, and $C(\mu, \nu)$ the optimal transport cost. Thus, this coupling is called optimal transport coupling. Obviously, the solution of the Kantorovich minimisation problem depends on the choice of the cost function. The following theorem guarantees the existence of optimal coupling.

Theorem 3.1 (Existence of optimal coupling [Vil09, Theorem 4.1]). *Given two Polish spaces (M, μ) and (N, ν), and a lower semicontinuous cost function bounded from below, then there always exist optimal couplings of (μ, ν).*

Notice that any transport map $T : M \rightarrow N$ induces a transfer plan γ defined by $(Id \times T)_* \mu$. In fact, we can also impose conditions on the cost function such that the optimal coupling is obtained by a deterministic coupling. The search for a deterministic coupling (transport map) T which minimises Eq. 2 for a given cost function is called the *Monge problem*. The following proposition ensures the existence of a unique transport map solving the Monge problem.

Proposition 3.2 (Solution of the Monge problem). *Let M be a Riemannian manifold, $\mathcal{X}$ a closed subset of M, with $\dim(\partial \mathcal{X}) \leq n-1$ and $\mathcal{Y}$ an arbitrary Polish space. Let $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a continuous cost function, bounded from below and assume that for the probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, the optimal cost $C(\mu, \nu)$ is finite. If the following conditions are fulfilled*

i) c is differentiable everywhere;
ii) μ is absolutely continuous;
iii) $\nabla_x c(x, \cdot)$ is injective where defined, i.e., if x, y, y' are such that $\nabla_x c(x, y) = \nabla_x c(x, y')$, then $y = y'$,

then there exists a unique (in law) optimal coupling (x, y) of (μ, ν), and it is deterministic.

Proof. The proof follows from Theorem 10.28, Proposition 10.7, and Remark 10.33 of [Vil09]. $\square$

Corollary 3.3. *Suppose that for each $k \in \mathbb{N}$ we have a sequence of continuous cost functions $c_k : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ converging uniformly to $c : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$, where c is defined as above. Let $(\nu_k)_{k \in \mathbb{N}}$ be a sequence of probabilities on $\mathcal{Y}$ converging weakly to $\nu \in \mathcal{P}(\mathcal{Y})$, and assume that for each k there exist measurable maps $T_k : \mathcal{X} \rightarrow \mathcal{Y}$, such that each T_k is an optimal transport map between μ and ν_k. Then T_k converges to T in probability, i.e.,*

$$\forall \varepsilon > 0 \quad \mu \left[\{ x \in M; d\left(T_k(x), T(x)\right) \geq \varepsilon \} \right] \xrightarrow[k \to \infty]{} 0.$$

Proof. The proof follows from Proposition 3.2 above and [Vil09, Corollary 5.23]. $\square$

Corollary 3.4. *Let $M = \mathbb{R}^n$ and $c(x, y) = -x \cdot y$. Consider two probability measures μ, ν on M, such that μ is absolutely continuous, then the solution of Monge's problem can be written as*

$$y = T(x) = x + \nabla \psi(x),$$

where ψ is some convex, lower semicontinuous function.

Proof. The proof follows from Theorem 10.44, Particular case 10.45, and Particular case 5.3 of [Vil09]. See also Section 6 below. $\square$

Remark 1. *Proposition 3.2 is a slight variation of a more general theorem about the solution of the Monge problem. Its conditions can be weakened or replaced in a number of ways; see* [Vil09, Theorem 10.28].

When the cost function is given in terms of distances in a metric space, we can use (2) in order to define a distance between measures.

Definition 2 (Wasserstein distances). *Let (M, d) be a Polish metric space, and $p \in [0, \infty)$. The Wasserstein distance of order p between any two probability measures μ, ν on M is given by*

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int_M d(x, T(x))^p d\gamma(x, T(x)) \right)^{1/p}. \qquad (3)$$

Using the Wasserstein distances we can define a space given by the restriction on $\mathcal{P}(M) \times \mathcal{P}(M)$ on which W_p takes finite values, or the space of probability measures with *finite moment of order p*.

Definition 3 (Wasserstein spaces). *The Wasserstein space of order p is given by*

$$\mathcal{P}_p(M) = \left\{ \mu \in \mathcal{P}(M) : \int_M d(x_0, x)^p \mu(dx) < \infty \right\}. \qquad (4)$$

The choice of x_0 is arbitrary and the space does not depend on this.

As a last important example we shall mention the powerful Moser coupling [Mos65, Vil09].

Theorem 3.5 (Moser coupling). *Consider a smooth compact Riemannian manifold M and its volume form m. In addition, consider Lipschitz continuous positive probability densities ρ_0 and ρ_1 on M. Then there is a deterministic coupling of $\mu_0 = \rho_0 m$ and $\mu_1 = \rho_1 m$. In other words, there exists a measurable map T such that for all Borel $E \subset M$, we have $\mu_1(E) = T_* \mu_0(E)$. Furthermore, if ρ_0, ρ_1 are $C^{k,\alpha}$ then T is $C^{k+1,\alpha}$.*

Remark 2. *The map T is explicitly given, for each $x \in M$, in terms of a solution of the elliptic equation*

$$\Delta u(x) = \rho_0 - \rho_1,$$

where Δ denotes the Laplace operator. The transport map is obtained by defining the locally Lipschitz vector field

$$\xi(t, x) = \frac{\nabla u(x)}{(1 - t)\rho_0(x) + t\rho_1(x)},$$

which integrates to the flow $(T_t(x))_{0 \le t \le 1}$ with an associated family of measures $(\mu_t)_{0 < t < 1}$. In particular, the time-1 map pushes forward μ_0 to μ_1. See [Vil09] *for more details.*

4. REPRESENTATION BY MEASURABLE RANDOM MAPS

From this section we begin to apply the techniques from optimal transport just presented. We shall start by tackling the problem of representing a Markov chain by measurable continuous maps. Our result implies Theorem 1.1 by Kifer [Kif86, Ch. 1]. We also treat the case of maps between different spaces. The main result of this section is

Theorem A. *Let N be a manifold and consider a measurable family of probability measures $(\mu_x)_{x \in M}$ with finite p-moments for $p \geq 1$, i.e. for some $y \in N$*

$$\sup_{x \in M} W_p(\delta_y, \mu_x) < \infty.$$

Then there exist random measurable maps $(f_\omega : M \to N)_{\omega \in \Omega}$ representing $(\mu_x)_{x \in X}$.

Proof. Let ν be any probability measure absolutely continuous with respect to some volume measure on N, such that $W_p(\nu, \delta_y) < \infty$. Then Proposition 3.2 shows that for each $x \in M$ there is a unique optimal coupling realised by a (measurable) transport map $T_x : N \to N$, i.e.

$$W_p(\nu, \mu_x)^p = \int d(y, T_x(y))^p d\mu_0(y).$$

Moreover, by Lusin's theorem and Corollary 3.3, these maps a.e. vary continuously with respect to $x \in M$ in the topology of convergence in probability with respect to ν. Setting $(\Omega, \tau) = (N, \nu)$, we can take for some random variable $X : \Omega \to N$ with $\mathrm{law}(X) = \nu$, the restriction,

$$f_\omega(x) = T_x(X(\omega)).$$

Because of continuity with respect to convergence in probability we can assume without loss of generality that $(f_\omega)_{\omega \in \Omega}$ is measurable and separable. $\square$

5. REPRESENTATION BY CONTINUOUS RANDOM MAPS

In this section we apply optimal transport and regularity theory to give conditions for the representation of Markov chains by continuous random maps. We start with the following result.

Proposition 5.1. *Let $(\mu_x)_{x \in M}$ be a continuous family of probability measures. Suppose for a fixed measure $\nu \in \mathcal{P}(N)$ with compact support there exists a family of continuous maps $(T_x : \mathrm{supp}\, \nu \to N)_{x \in M}$ varying continuously in the C^0-topology such that*

$$(T_x)\nu = \mu_x.$$

Then $(\mu_x)_{x \in M}$ can be represented by random continuous maps $(f_\omega)_{\omega \in \Omega}$ such that $(\Omega, \tau) = (\mathrm{supp}\, \nu, \nu)$.

If, in addition, T_x varies Hölder or Lipschitz continuously with respect to $x \in M$ then so does f_ω with the same constants (resp. exponents).

Proof. Set $(\Omega, \tau) = (\operatorname{supp} \nu, \nu)$ and let $X : \Omega \to \Omega$ be any random variable such that law $X = \nu$. Define

$$f_\omega(x) = T_x(X(\omega)).$$

By construction, $(f_\omega)_{\omega \in \Omega}$ represents $(\mu_x)_{x \in M}$. We need to show that $f_\omega : M \to N$ is continuous. Since T_x varies continuously with respect to the C^0-topology and its domain is compact we have for a fixed T_x

$$d(T_x(X(\omega)), T_y(X(\omega)) \le d_{C^0}(T_x, T_y) < \epsilon$$

whenever $d(x, y) < \delta$ for sufficiently small $\delta = \delta(\epsilon, x) > 0$. $\qquad \square$

Remark 3. *Proposition 5.1 above generalises Example 1.7 by Benedicks-Viana* [BeV06], *where ν is the Lebesgue measure on $[0, 1]^n$ and T_x is the inverse of a rearrangement S_x of the positive measure $\mu_x \in \mathcal{P}([-\epsilon, \epsilon]^n)$ to $[0, 1]^n$ assuming S_x is a homeomorphism which varies continuously in C^0.*

Remark 4. *To get random continuous maps it is actually enough to assume that $x \mapsto T_x$ is pointwise continuous ν-a.e, i.e. if $x \to y$ then for ν-a.e. ω we have $T_x(\omega) \to T_y(\omega)$.*

Remark 5. *Notice that the assumption on the continuity of each T_x could be relaxed by using Lusin's theorem. Furthermore, using a general version of the Kolmogorov-Chentsov continuity lemma for random fields obtained in* [Pot09], *a similar result holds if we only assume that each T_x is Borel measurable converging fast enough as $x \to y$ in the topology of convergence in probability with respect to (Ω, τ).*

For the following main theorem we apply Moser's coupling. Our theorem directly implies the main result of Quas [Qua91].

Theorem B. *Assume N is a compact Riemannian manifold without boundary with normalised volume measure m. Let $(\mu_x)_{x \in M}$ be a continuous family of probability measures such that each μ_x, absolutely continuous with respect to m, has a positive Hölder continuous (for some exponent $\alpha > 0$) probability density varying continuously with respect to $x \in M$.*

Then $(\mu_x)_{x \in M}$ can be represented by random continuous maps $(f_\omega)_{\omega \in \Omega}$.

Proof. Let $\mu_x = \rho^{(x)} dm$. Then, by Theorem 3.5, there exists a coupling of (m, μ_x) induced by the time-1 map of a Lipschitz continuous vector field varying continuously with respect to $x \in M$. It is given by the solution of the following elliptic equation on N

$$\Delta u^{(x)} = 1 - \rho^{(x)}.$$

By elliptic regularity theory, $u^{(x)}$ is $C^{2,\alpha}$ so we can define the vector field on N

$$\xi_x(t, y) = \frac{\nabla u^{(x)}(y)}{(1 - t) + t\rho^{(x)}(y)}.$$

According to our assumptions it is well-defined and Lipschitz continuous
and thus integrates to a flow $T_{x,t} : N \to N$. Furthermore, we have

$$(T_{x,1})_* m = \mu_x.$$

Since the densities vary continuously, so does the vector field and thus the
map $T_{x,1}$ which implies by Proposition 5.1 that $(\mu_x)_{x \in M}$ can be represented
by random continuous maps $(f_\omega)_{\omega \in \Omega}$. $\square$

6. Regularity of densities

The next step is to establish conditions on representations of Markov
chains by random diffeomorphisms. Before that, we shall again take up
the discussion on optimal transport and its regularity properties applied to
the regularity of the densities of Markov chains. We will focus on families
of probabilities on $\mathbb{R}^n$. After that, we show how to use the results of this
section on Riemannian manifolds via lifting and embedding techniques. We
start by showing how convexity of the support of measures is related to the
regularity of the transport maps.

Consider a *potential* $\phi : U \subset \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, and define at $y \in U' \subset \mathbb{R}^n$

$$\phi^c = \sup_{x \in U}(-c(x,y) - \phi(x)).$$

Furthermore, let us define

$$G_\phi(x) = \{y \in U' : \phi(x) + \phi^c(y) = -c(x,y)\},$$

then we can state the following general result, taken from [Loe09, Theorem
2.7].

Theorem 6.1. *Let $U, U' \subset \mathbb{R}^n$ be bounded domains of $\mathbb{R}^n$, and $c(x,y) = -x \cdot y$. let μ, and ν be probability measures on U, and U'. Assume that μ does not give mass to sets of Hausdorff dimension less than or equal to $n - 1$. Then there exist a μ-a.e. unique T solving the Monge problem for this cost function c. Moreover, there is a convex potential ϕ on U, such that $T = G_\phi$. Finally, if ψ is convex and satisfies $(G_\psi)_*\mu = \nu$, then $\nabla\psi = \nabla\phi$, μ-a.e.*

In fact, for the cost function $c(x,y) = -x \cdot y$, one can show that $\phi = \psi^c$,
and that they are Legendre transforms of each other. See Particular case
5.3, and Definition 5.7 in [Vil09].

Regarding the regularityf of ϕ, Loeper proves the following result that we
shall use in the sequel.

Theorem 6.2 ([Loe09, Theorem 3.4]). *Let $U, U' \subset \mathbb{R}^n$ be bounded, strictly convex with respect to each other. Suppose μ and ν are probability measures on U and $V' \subset U'$, respectively, with V' being convex with respect to U.*

$$\mu(B_\varepsilon(x)) \leq C_\mu \varepsilon^{n(1-\frac{1}{p})} \quad \text{for all } \varepsilon > 0 \text{ and } x \in U$$

and $\nu \geq \kappa m^n$ on V' for some $\kappa > 0$. Assume that for the convex potential ϕ and cost function $c(x, y) = -x \cdot y$ we have $(G_\phi)_* \mu = \nu$.

Then ϕ is continuously differentiable on $U_\delta = \{x \in U \,|\, d(x, \partial U) > \delta\}$ with Hölder continuous derivatives for $\delta > 0$. In particular, for some $\beta \in (0, 1)$ and $\mathcal{C}$ depending only on U, U', $m > 0$, $\delta > 0$, p and C_{μ_0}

$$\|\phi\|_{C^{1,\beta}(U_\delta)} \leq \mathcal{C}.$$

If, furthermore, μ is supported on some $\bar{V}$ compactly contained in U then G_ϕ is a Hölder continuous map with Hölder norm bounded by $\mathcal{C}$.

Consider some compact $K \subset \mathbb{R}^n$, and assume $(\mu_x)_{x \in K}$ to be a family of measures, such that each μ_x satisfies the assumption of μ in Theorem 6.2. Then it follows that we can define optimal transport maps S_x between μ_x and $\nu = m_{|[0,1]^n}$. These maps are Hölder continuous with Hölder norm bounded by a constant only depending on some compact convex neighbourhood Ω_x of the support of μ_x. In particular, by uniqueness of the optimal transport map, we have

Lemma 6.3. *Let $(\mu_x)_{x \in K}$ be a continuous family of probability measures on some compact set $K \subset \mathbb{R}^n$. Suppose that each μ_x satisfies the assumptions of Theorem 6.2 with $C = C_{\mu_x}$ independent of μ_x and all supports contained in some convex $U \subset \mathbb{R}^n$. If we assume that the supports of μ_x are contained in $D = \mathrm{cl}(U_\delta)$ for some $\delta > 0$ then the optimal transport maps S_x between μ_x and ν vary uniformly, that is if $x_n \to x$ then $S_x : D \to U'$ varies in the uniform topology of $C^0(D, U')$.*

Proof. Because the $\| \cdot \|_{C^{1,\beta}}$-norm of $\{\phi_x\}_{x \in K}$ is uniformly bounded, the set is pre-compact in $C^1(D)$. Since up to a constant the potentials are unique we can assume $\phi_x(y_0) = 0$ for all $x \in K$ and some $y_0 \in D$. The limit of $\lim_{n \to \infty} \phi_{x_n} \to \tilde{\phi}_{x_0}$ for some $x_n \to x_0$ is also a convex potential solving the optimal transport problem with $\tilde{\phi}_{x_0}(y_0) = 0$ which implies $\tilde{\phi}_{x_0} = \phi_{x_0}$. Therefore $\{\phi_x\}_{x \in K}$ is already closed and thus compact and $S_x : D \to U'$ varies continuously in the C^0-topology as x varies in K. $\square$

Proposition 6.4. *Let $(\mu_x)_{x \in K} \subset \mathcal{P}(\mathbb{R}^n)$ be as above. In addition, assume each μ_x is supported on a convex set and has strictly positive Lebesgue density on its support, i.e. $\mu_x \geq \kappa \cdot m$ for some $\kappa > 0$ independent of x, then*

$$S_{x | \mathrm{supp}\, \mu_x} : \mathrm{supp}\, \mu_x \to [0, 1]^n$$

is continuously invertible. Furthermore, if the supports vary continuously with respect to the Hausdorff metric on (compact) subsets of $\mathbb{R}^n$ then the inverse $T_x = S_x^{-1} : [0, 1]^n \to \mathrm{supp}\, \mu_x \subset U$ varies continuously in the $C^0([0, 1]^n, U)$. In addition, such a family $(\mu_x)_{x \in X}$ can be represented by random continuous maps.

Remark 6. *If we do not assume that the supports vary continuously then it is still possible to show that the maps T_x converge pointwise on $[0, 1]^n$ to T_y as x converges to y in K.*

Proof. By the previous result S_x varies continuously in $C^0(D, \mathbb{R}^n)$. Then Theorem 6.1 implies that the optimal transport problem from ν to μ_x has a unique continuous optimal transport map $T_x : [0,1]^n \to \operatorname{supp} \mu_x$ which, by Corollary 3.4, is the derivative of a differentiable convex potential ψ_x. Because ψ_x is (up to a constant) the Legendre transform of the potential ϕ_x, whose derivative is S_x, we necessarily have $T_x = S_x^{-1}$ on $[0,1]^n$.

To prove that $x \mapsto T_x$ is continuous it is sufficient to show that the graphs converge with respect to the Hausdorff metric. Because $x \mapsto S_x$ is continuous and the supports of μ_x vary continuously with respect to the Hausdorff metric, the restricted graph

$$\widetilde{gr}(S_x) = \{(y, S_x(y)) \in \operatorname{supp} \mu_x \times [0,1]^n\}$$

is continuous with respect to the Hausdorff metric on subsets of $\mathbb{R}^n \times \mathbb{R}^n$. But this implies

$$x \mapsto gr(T_x) = (\widetilde{gr}(S_x))^{-1}$$

is continuous and thus $x \mapsto T_x$ is continuous as well.

In particular, Proposition 5.1 implies that any such family $(\mu_x)_{x \in K}$ can be represented by random continuous maps.

$\square$

Proposition 6.5. *Suppose that μ is absolutely continuous with respect to Lebesgue measure m, and $\nu = m_{|[0,1]^n}$. Suppose further that both satisfy the assumptions of μ and ν, respectively, as in Proposition 5.1. Then S is continuously invertible on U, where U is the set of point z which admit a (convex) neighbourhood where the density of μ is strictly positive. In particular, this holds in $\{\rho > 0\} = \operatorname{int}(\operatorname{supp} \mu)$ if the density ρ of μ is continuous.*

Proof. Let S be the (Hölder) continuous optimal transport map from Remark 3 of Proposition 5.1. Since S and $T = S^{-1}$ (as multi-valued map) are the gradients of convex potentials ϕ and ψ, which are the Legendre transforms of each other, it suffices to show that T is single-valued on $S(U)$. By definition we have

$$S(x) = \{y \mid \phi(x) + \psi(y) = x \cdot y\}$$

and

$$\psi(y) = \sup\{xy - \phi(x)\},$$

which implies that $S^{-1}(S(x))$ is convex. So if we show that T is one-to-one on $S(V_x)$ for some small neighbourhood $x \in U$ then it implies that $S^{-1}(S(x)) \cap V_x = x$, i.e. S and T are both single-valued and thus continuous on resp. U and $S(U)$.

Let π be the optimal transport plan between ν and μ, in particular we have

$$(\operatorname{id} \times T)_* \nu = \pi.$$

Let x be a point in U and V_x be a closed neighbourhood with the prescribed property. By the restriction property for optimal transport plans [Vil09, Theorem 4.6] the plan $\tilde{\pi}_{|\mathbb{R}^n \times V_x}$ is an optimal transport plan between

its marginal, i.e. between some $\tilde{\nu} \leq \nu$ and $\tilde{\mu} = \mu_{|V_x}$. Obviously this plan is also induced by T, i.e.

$$(\mathrm{id} \times T_*)\tilde{\nu} = \tilde{\pi}.$$

The measures $\tilde{\nu}$ and $\tilde{\mu}$ satisfy the assumptions of Proposition 5.1, which implies that T is (Hölder) continuous. Therefore T is one-to-one on $\operatorname{supp} \tilde{\nu} = S(V_x)$. $\qquad \square$

Corollary 6.6. *Let μ and ν be as above. Assume the support of μ is strictly convex and*

$$U = \{\rho > 0\} = \mathrm{int}(\operatorname{supp} \mu)$$

where ρ is the continuous density of μ. Then the optimal transport map T from ν to μ is a continuous map from $\operatorname{supp} \nu$ to $\operatorname{supp} \mu$.

Proof. Restrict S to $\operatorname{supp} \mu$. By the previous theorem S is injective in the interior of its domain. Furthermore, $S^{-1}(S(x))$ is convex. Because the support of μ is strictly convex this also implies S is injective on the boundary, i.e. if $x' \in S^{-1}(S(x))$ for $x \in \partial \operatorname{supp} \mu$ then $\lambda x' + (1 - \lambda)x \in \partial \operatorname{supp} \mu$ which implies $x = x'$.

Because S is one-to-one on its (convex) domain $\operatorname{supp} \mu$ and $S(\operatorname{supp} \mu) = \operatorname{supp} \nu$, so is its inverse T on $\operatorname{supp} \nu$. Thus T is continuous. $\qquad \square$

The two previous results show that problems happen if the supports are not strictly convex or not positive everywhere. Furthermore, if the support of μ is not simply connected (e.g. has a hole) then this hole is an isolated singularity of T in the interior of the support of ν. Other phenomena like disjoint support might look like complete cut of the support of ν.

Proposition 6.7. *Let K be some (compact) set and B_1 be the closed unit ball in $\mathbb{R}^n$ and $(f_x : B_1 \to f_x(B_1) \subset \mathbb{R}^n)_{x \in K}$ be a family of diffeomorphisms (onto their images) varying continuously in C^1 w.r.t. x. Assume $(\mu_x)_{x \in K}$ is a continuously varying family of measures supported on the images of f_x, i.e.*

$$\operatorname{supp} \mu_x = f_x(B_1).$$

If each μ_x is Lebesgue regular with L^∞ (resp. continuous) density then there is a (unique) continuously varying family $(\nu_x)_{x \in K}$ of Lebesgue regular measures with L^∞ (resp. continuous) densities with $(f_x)_* \nu_x = \mu_x$.

Furthermore, if every point in the interior of the support of μ_x admits a neighbourhood such that the Lebesgue density of μ_x is strictly positive in the interior of the support then the same holds for ν_x.

Proof. Assume $d\mu_x = \rho_x dm$ and there is some (Lebesgue regular) family $d\nu_x = \vartheta_x dm$ such that $(f_x)_* \nu_x = \mu_x$. By the Jacobian equations we have

$$\vartheta_x(z) = \rho_x(f_x(z)) \cdot J_{f_x}(z).$$

So defining $\vartheta_x(z)$ as above gives us ν_x.

Because $(f_x)_{x \in K}$ are continuously varying diffeomorphisms the Jacobians $J_{f_x}(z)$ vary continuously with respect to $(x, z) \in K \times B_1$ which implies the statement of the proposition. $\qquad \square$

Corollary 6.8. *Assume $(\mu_x)_{x \in K}$ is a continuous family of measures and the supports admit diffeomorphisms as above. Then there exists a family of continuous maps $T_x : [0,1]^n \to \mathbb{R}^n$ varying in the C^0-topology and the following holds*

$$(T_x)_* m_{|[0,1]^n} = \mu_x.$$

In particular, all such families $(\mu_x)_{x \in K}$ can be represented by random continuous maps $(f_\omega)_{\omega \in \Omega}$.

Proof. Follows from Proposition 6.4, Corollary 6.6 and the previous proposition. Finally the last statement follows from Proposition 5.1.   □

Remark 7. *The corollary can be applied if the supports are star-shaped with differentiably varying centre and radial function, i.e. there are $z_x \in B_x = \operatorname{supp} \mu_x$ and differentiable maps*

$$r_x : \mathbb{S}^{n-1} \to (0, \infty)$$

such that $x \mapsto (z_x, r_x)$ is continuous from K to $\mathbb{R}^n \times C^1(\mathbb{S}^{n-1}, (0, \infty))$ (continuity of $x \mapsto z_x$ is enough to show that $J_{f_x}(z)$ is continuous in x).
The diffeomorphisms $f_x : B_1 \to \mathbb{R}^n$ are constructed via

$$f_x((\alpha, \rho)) = z_x + (\alpha, r_x(\alpha) \cdot \rho).$$

(For simplicity we mixed polar coordinates (α, ρ) with Cartesian z_x).

7. Lifting measures

In the previous section we constructed random maps for a family of measures on $\mathbb{R}^n$. In order to use this result also for a family of measures on a manifold we need to lift these measures to measures on $\mathbb{R}^n$. The idea is to use the local equivalence via the exponential map of a neighbourhood of a point $p \in N$ and a neighbourhood of 0 of the tangent space at that point [Jos11]. In other words, the map

$$x \mapsto \mu_x \in \mathcal{P}(N),$$

implicitly lifts to a map

$$x \mapsto \tilde{\mu}_x \in \mathcal{P}(T_{f(x)} N),$$

for some continuous map $f : M \to N$, e.g. the centre of mass. In general, f is given by the unperturbed system if one considers a (bounded) random perturbation such that

$$(\exp_{f(x)}^{-1})_* \mu_x = \tilde{\mu}_x.$$

In [Ara00, Example 1] Araújo used a similar construction for uniformly distributed perturbations and in case the tangent bundle is parallelizable, i.e.

$$TN \cong N \times \mathbb{R}^n.$$

Notice in that case $x \mapsto \tilde{\mu}_x$ can be considered as a pair of maps

$$x \mapsto (f(x), \tilde{\mu}_x) \in N \times \mathbb{R}^n.$$

For the second factor there is a random variable X such that the exponential map gives naturally the distribution of random maps around f on N, i.e.

$$f_\omega(x) = \exp_{f(x)}(X_\omega(x)).$$

More generally, if there is a trivial bundle $F \cong N \times \mathbb{R}^k$ with a local equivalence $\rho : F \to N$ of a neighbourhood of 0 of $F_{f(x)}$ to the manifold (or the tangent bundle at $T_{f(x)}N$) such that we can lift the measures to

$$x \mapsto \hat{\mu}_x \in \mathcal{P}(F_{f(x)})$$

with $(\hat{\mu}_x)_{x \in M}$ satisfying the assumptions of the previous section and

$$\rho_* \hat{\mu}_x = \mu_x$$

then there is a family of random maps representing $(\mu_x)_{x \in M}$.

Since the exponential map at a point $p \in N$ is a local diffeomorphism between a neighbourhood of that point and a neighbourhood of 0 in the tangent space at p, we get the following lemmata.

Lemma 7.1. *Let $(\mu_x)_{x \in M}$ be a continuous family of probability on a manifold N. Suppose that there is a C^r-map $f : M \to N$, for $r \geq 0$, such that for each x the support of μ_x is contained in a sufficiently small neighbourhood $U_{f(x)}$ of $f(x)$.*

Then $(\mu_x)_{x \in M}$ lifts to a continuous family of probability measures $(\tilde{\mu}_x)_{x \in M}$ on TN with $\tilde{\mu}_x$ supported on $T_{f(x)}N$ (considered as a subset of TN).

Lemma 7.2. *If, in addition to the assumptions above, μ_x is absolutely continuous with respect to a volume form on N then $\tilde{\mu}_x$ is absolutely continuous with respect to the Lebesgue measure on $T_{f(x)}N$ (and by equivalence to the standard Lebesgue measure on $\mathbb{R}^n$).*

Also, if N is compact then the the Lebesgue densities of μ_x and $\tilde{\mu}_x$ are comparable in the sense that they have the same growth conditions, Lipschitz- or Hölder-constants or positivity properties of the density in the interior of their support. And, in particular, (strict) convexity of the support of μ_x is preserved if the support is contained in a ball around $f(x)$ with radius less than the convexity radius of N.

As a final step we need to construct a trivial bundle with a natural projection and show that the measures lift again. As it is well know, the tangent bundle TN is a sub-bundle of the trivial bundle $F = N \times \mathbb{R}^k$ for some $k \geq n$ with a natural projection

$$\pi : F \to TN,$$

which is a linear projection from $\mathbb{R}^k$ to $\mathbb{R}^n$ at each fibber at $p \in N$ [Jos11].

Suppose we have a fixed measure ν on $\mathbb{R}^k$ which has a smooth density with respect to the Lebesgue measure on $\mathbb{R}^k$, such that for any linear projection $r : \mathbb{R}^k \to \mathbb{R}^n$ the measure

$$\nu^{(r)} = r_* \nu$$

is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ with positive density inside the interior of its support. Then we claim that we

can lift any Lebesgue regular measure $\tilde{\mu}$ on $\mathbb{R}^n$ with support in the interior of the support of $r_*\nu$ to a Lebesgue regular measure $\hat{\mu}$ on $\mathbb{R}^k$.

To show this, notice that $\tilde{\mu}$ is absolutely continuous with respect to $\nu^{(r)}$, i.e.

$$d\tilde{\mu}(x) = g(x)d\nu^{(r)}(x)$$

then defining

$$d\hat{\mu}(y) = g(r(y))d\nu$$

gives the required measure $\hat{\mu}$. Obviously if the support of $\tilde{\mu}$ is convex so is the support of $\hat{\mu}$ and positivity in the interior also lifts in this way.

8. Representation by random diffeomorphisms

As a result of the conditions of representation of measures by continuous maps, and the lifting properties of measures, we have proved the following theorem giving conditions to represent Markov chains by a family of random diffeomorphisms.

Theorem C. *Let M and N be compact Riemannian C^r-manifolds without boundary, and let m be the normalised volume measure on N. Let $(\mu_x)_{x\in M}$ be a continuous family of probability measures on N such that each μ_x is absolutely continuous with respect to m, has positive density and convex support. Suppose that there is a C^r-diffeomorphism $f : M \to N$, for $r \geq 1$, such that for each x, the support of μ_x is contained in a sufficiently small neighbourhood $U_{f(x)}$ of $f(x)$. Then $(\mu_x)_{x\in M}$ can be represented by a family $(f_\omega)_{\omega\in\Omega}$ of C^r-random diffeomorphisms.*

Recall that the space $\mathrm{Diff}^r(M, N)$ of diffeomorphisms is open in $C^r(M, N)$, for $r \geq 1$ [Hir76]. Therefore, what we need to do is to perturb f in the C^r topology in a controlled way, according to our measure μ_x. Indeed, as discussed in Section 7, it is possible to lift the measures on N to its tangent bundle via the exponential map and we can identify TN with a subbundle of a trivial bundle $N \times \mathbb{R}^l$ [Jos11]. Then Lemma 7.1, and Lemma 7.2 provide a natural way of constructing a continuous family of probabilities on the bundles according to $(\mu_x)_{x\in M}$ on N. Furthermore, the construction of a trivial bundle shows a natural way of embedding, thus lifting again this measures to some $\mathbb{R}^k$, such that the results of Section 6 can be applied. In other words, we can continuously select sections of the bundles. Thus, we choose maps C^r-close to f according to μ_x. This completes our argument.

Acknowledgements

References

[Ara00]   V. Araújo, *Attractors and time averages for random maps.*, Annales de l'Institut Henri Poincaré Analyse Non Linéaire **17** (2000), no. 3, 307–369.

[BeV06]   M. Benedicks and M. Viana, *Random perturbations and statistical properties of Hénon-like maps.*, Annales de l'Institut Henri Poincaré Analyse Non Linéaire **23** (2006), no. 5, 713–752.

[BlC70]   R. M. Blumenthal and H. H. Corso, *On continuous collections of measures.*, Annales de l'Institut Fourier **20** (1970), no. 2, 193–199.

[BDV05]   C. Bonatti, L. J. Díaz, and M. Viana, *Dynamics beyond uniform hyperbolicity: a global geometric and probabilistic perspective*, Encyclopedia of Mathematical Sciences, vol 102, Springer, Berlin, 2005.

[Fed69]   H. Federer, *Geometric measure theory*, Grundlehren der Mathematischen Wissenschaften, vol. 153, Springer, New York, 1969.

[Hir76]   M. W. Hirsch, *Differential topology*, Springer-Verlag, Berlin, 1976.

[Jos11]   J. Jost, *Riemannian Geometry and Geometric Analysis*, Springer, Berlin, 2011.

[Kif86]   Yu. Kifer, *Ergodic theory of random transformations.*, Birkhäuser, Boston, 1986.

[Kif88]   Yu. Kifer, *Random perturbations of dynamical systems.*, Birkhäuser, Boston, 1988.

[Loe09]   G. Loeper, *On the regularity of solutions of optimal transportation problems*, Acta Mathematica **202** (2009), no. 2, 241–283.

[Mos65]   J. Moser, *On the volume elements on a manifold*, Trans. Amer. Math. Soc., **120** (1965), 286–294.

[Pot09]   J. Potthoff, *Sample properties o random fields II: continuity*, Communications in stochastic analysis, **3** (2009), no. 3, 331–348.

[Qua91]   A. N. Quas, *On representation of Markov chains by random smooth maps*, Bulletin of the London Mathematical Society **23** (1991), no. 5, 487–492.

[Vil09]   C. Villani, *Optimal transport: old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, Berlin, 2009.

[ZaH07]   H. Zmarrou and A.J. Homburg, *Bifurcations of stationary measures of random diffeomorphisms*, Ergodic Theory and Dynamical Systems **27** (2007), no. 5, 1651–1692.

Jürgen Jost, Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany, and Department of Mathematics, University of Leipzig, 04081 Leipzig, Germany

E-mail address: `jost@mis.mpg.de`

Martin Kell, Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany

E-mail address: `mkell@mis.mpg.de`

Christian S. Rodrigues, Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany

E-mail address: `christian.rodrigues@mis.mpg.de`