Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Stability of the global attractor under Markov-Wasserstein noise

by

Martin Kell

Preprint no.: 8 2011

STABILITY OF THE GLOBAL ATTRACTOR UNDER MARKOV-WASSERSTEIN NOISE

MARTIN KELL

Abstract. We develop a "weak Ważewski principle" for discrete and continuous time dynamical systems on metric spaces having a weaker topology to show that attractors can be continued in a weak sense. After showing that the Wasserstein space of a proper metric space is weakly proper we give a sufficient and necessary condition such that a continuous map (or semiflow) induces a continuous map (or semiflow) on the Wasserstein space. In particular, if these conditions hold then the global attractor, viewed as invariant measures, can be continued under Markov-type random perturbations which are sufficiently small w.r.t. the Wasserstein distance, e.g. any small bounded Markov-type noise and Gaussian noise with small variance will satisfy the assumption.

In this paper we are going to show that the invariant measures of a dynamical system having a global attractor (either discrete and continuous time) can be "continued" under small (not necessarily bounded) noise. Instead of just showing that there is a stationary measure of the perturbed system "weakly" close to the original ones (see e.g. [Kif88, 1.7]) we show that it is close w.r.t. the Wasserstein metric w_p (the order p depends on regularity of the noise). Previous research mainly focused on Gaussian type noise, "absolute continuous" noise or assumed implicitly bounded noise, e.g. Kifer [Kif88, p.103] and L.-S. Young [You86] considered noise on a positive invariant (bounded) neighborhood U of a local attractor which is zero on ∂U and thus is bounded. If X is compact than any noise will be bounded. Thus we are in particular interested in non-compact X, although we restrict our attention to proper metric spaces which includes all locally compact geodesic spaces.

All our results apply equally to discrete and continuous time dynamical system. We will mainly focus on discrete time because there is a better intuition behind these. In the first section we extend Rybakowski's continuation of a positive invariant isolating neighborhood to the discrete time setting which will be the key step to treat continuous and discrete time systems on equal footing. Using a kind of "weak Ważewski principle" we show that attractors can be continued in a weak sense without assuming admissibility of the perturbed system (theorem 5).

Then we introduce the Wasserstein space and show that the Wasserstein space of a proper metric space is weakly proper, i.e. closed δ-neighborhoods of compact sets are weakly compact (theorem 6). We give a necessary and sufficient condition such that a dynamical system f (resp. a semiflow π) on a proper metric space makes the transfer map $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$ (resp. transfer semiflow π_*), which is always

Date: March 17, 2011.

2000 *Mathematics Subject Classification.* 34D23, 37B35, 60B05, 60B10.

Key words and phrases. global attractor, random perturbation, Wasserstein space.

The author would like to thank the IMPRS "Mathematics in the Sciences" for financial support and his advisor, Prof. Jürgen Jost, and the MPI MiS for providing an inspiring research atmosphere.

(weakly) continuous, (strongly) continuous on any Wasserstein space $(\mathcal{P}_p(X), w_p)$ of order $1 \leq p < \infty$.

Finally, we look at Markov-type perturbations of a dynamical system f (resp. semiflow π) having a global attractor. These are perturbation of f_* in the Wasserstein space. Under quite general assumptions on f we can show using Conley theory that if the perturbation is sufficiently small then the perturbed system has an isolated (weak) attractor and there is at least one stationary measure (theorem 15). The perturbed attractor is strongly close to the unperturbed attractor and contains all invariant measures and its positive invariant isolating neighborhood is convex. If the noise is of order p then the mass of the invariant measures decays at least as R^{-p} where R is the distance from the global attractor of f (resp. π). Furthermore, a standard result from Conley theory shows that if the noise level converges to zero then the set of invariant measures converges in the Wasserstein space to the set of invariant measures of the deterministic system (as usual without further assumptions only upper semicontinuity holds).

Because of the special form of the Kantorovich-Rubinstein duality in $\mathcal{P}_1(X)$ we can show that the (local) weak attractor of the perturbed system $P : \mathcal{P}_1(X) \to \mathcal{P}_1(X)$ is actually the global weak attractor of P.

The framework of a metric space with a weaker topology used here is similar to the framework in [AGS08, section 2.1] used to construct general gradient flows.

Motivation. Consider a dynamical system $f : X \to X$ on a proper metric space X having a global attractor, i.e. a compact invariant set A that attracts all its (bounded) neighborhoods. We will not consider the map f itself, but the map $f_* : \mathcal{P}_1(X) \to \mathcal{P}_1(X)$ defined via push-forward map on the space of probability measures. If f is "nice" then f_* is continuous and has the global attractor $K = \mathcal{P}(A)$, i.e. the probability measures supported on the global attractor of f.

Markov-type noise can be considered as a perturbation $\tilde{F}$ of f_*, i.e. instead of $\delta_x \mapsto \delta_{f(x)}$ we have $\delta_x \mapsto p(dy|x)$ where $p(dy|x)$ and $\delta_{f(x)}$ uniformly close w.r.t. the Wasserstein distance w_1 for all x. This can be seen as a smearing of the image $f(x)$ or some uncertainty about the actual image. For example, if f is the time-1 map of a flow generated by the ODE $\dot{x} = g(x)$ then $\tilde{F}$ could be the time-one map of the flow of distributions of the SDE $dx = g(x)dt + \epsilon dW_t$, i.e. additive Gaussian noise with small variance.

If $\tilde{F}$ and f_* are sufficiently close then $\tilde{F}$ has a (weak) global attractor (in $\mathcal{P}_1(X)$) which is close to K w.r.t. w_1. Hence stability of the global attractor holds in the Wasserstein space $\mathcal{P}_1(X)$.

The following example is inspired by Crauel, Flandoli - "Additive Noise Destroys a Pitchfork Bifurcation" [CF98] and could be decribed as "Additive Noise Destroys Attractors". The noise will be worse than white noise used by Crauel and Flandoli, but can still be considered as small.

Example (Generic collapse under "small" noise)**.** (1) Suppose $f : X \to X$ has a global attractor and at least one fixed point x_0 (the argument works equally well with general attractors). Take any noise level $\epsilon > 0$ and let $P_\epsilon : \mathcal{P}(X) \to \mathcal{P}(X)$ be the Markov map induced by

$$x \mapsto (1 - \epsilon)\delta_{f(x)} + \epsilon\delta_{x_0}.$$

This map is (weakly) close to the unperturbed system $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$. Namely, if d_{LP} is the Levy-Prokhorov distance (which metrizes $\mathcal{P}(X)$) then

$$\sup_{\mu \in P(X)} d_{LP}(P_\epsilon(\mu), f_*(\mu)) \leq \epsilon.$$

But P has exactly one invariant measure, namely δ_{x_0}, and all others converge to this measure weakly.

(2) Now we want to show that this can also happen in any Wasserstein space $\mathcal{P}_p(X)$ for $1 \leq p < \infty$ ($\mathcal{P}_\infty(X)$ only allows bounded noise which when sufficiently small cannot destroy local attractors and thus a global attractors with at least two sinks never collapses, see [Kel11]). Suppose $a : X \to [0,1]$ is a continuous function. Define $q_a(dy|\cdot) : X \to \mathcal{P}_p(X)$ by

$$x \mapsto q_a(dy|x) = (1 - a(x))\delta_{f(x)} + a(x)\delta_{x_0},$$

which is obviously continuous. Thus

$$w_p(q_a(dy|x), \delta_{f(x)})^p = a(x)d(f(x), x_0)^p.$$

So if we define

$$a(x) = \frac{\epsilon^p}{1 + d(f(x), x_0)^p}$$

then

$$w_p(q_a(dy|x), \delta_{f(x)})^p = \frac{\epsilon^p \cdot d(f(x), x_0)^p}{1 + d(f(x), x_0)^p} \leq \epsilon^p.$$

So, in particular, the induced MW-map $Q_a : \mathcal{P}_p(X) \to \mathcal{P}_p(X)$ of order p relative to f has noise level ϵ. Furthermore, the only invariant measure of Q_a is δ_{x_0} and all other measures converge to it.

The example above should make clear that using arbitrary unbounded noise even when it is small can have strange effects on the global attractor. Although we have some "attracting" invariant measures of the perturbed system the attractor might look very different from the original one, in our case it might be just one fixed point and this one can even be the "most" unstable one of the original attractor. Therefore, stochastic stability of attractors under arbitrary "small" noise should not be referred to a single invariant measure but to all of them, even though we can speak of stochastic stability if the type of noise is more restricted, besides of being sufficiently "small".

1. Discrete-time Conley theory for stable invariant sets

In this section we will use Conley theory, that is continuation methods from Conley index theory without using the topological (or (co)homological) Conley index. We will prove a continuation for a positive invariant neighborhood of a stable isolated invariant set of a time discrete dynamical system. The result will not require a compactness assumption (called admissibility) of the perturbed system and is a different type of continuation than [MR91]. Our proof will follow the proof of [Ryb87, Theorem 12.3] which is the continuation for semiflows. In particular, the results stated here and in the next sections also hold for semiflows if we assume that they do not explode on a given neighborhood.

In both cases the Ważewski principle for the index pair of the perturbed system does not apply. But we can use other assumptions to show that attractors continue, e.g. the map is weakly continuous and closed δ-neighborhoods of compact set are weakly compact, which is the case for the Wasserstein space on proper metric spaces.

We will now give the definitions used in [MR91] and [Ryb87] to prove the existence of an index pair for certain isolated invariant sets. Our setting will be a complete separable metric space Y and a dynamical system, i.e. a continuous map $f : Y \to Y$. A full left solution of f in N is a sequence $\{x_{-n}\}_{n\in\mathbb{N}} \subset N$ such that $f(x_{n-1}) = x_n$ for $n \leq 0$. Define the following sets

$$\begin{aligned} A^+(N) &= \{x \in N \,|\, f^k(x) \in N \text{ for all } k \geq 0\} \\ A^-(N) &= \{x \in N \,|\, \exists \text{ full left solution } \{x_{-n}\}_{n\in\mathbb{Z}} \text{ in } N \text{ through } x_0 = x\} \\ A(N) &= A^+(N) \cap A^-(N). \end{aligned}$$

These are called the maximal positive invariant (resp. negative invariant, resp. invariant) set in N. If N is unbounded then $A(N)$ usually denotes only the bounded invariant orbits instead of all of them. A set K is called invariant if $A(K) = K$. If there is a closed neighborhood N of an invariant set K with $A(N) = K$ then K is called isolated with isolating neighborhood N.

For $l, m \in \mathbb{N}$ and $l \leq k$ define

$$f^{[l,m]}(x) = \{y \,|\, f^k(x) = y \text{ for some } k \in \mathbb{N} \cap [l, m]\}$$

If $f_n : Y \to Y$ is a sequence of continuous maps such that $f_n \to f$, i.e. $f_n(x_n) \to f(x)$ whenever $x_n \to x$ as $n \to \infty$, then we say that a closed bounded set N is $\{f_n\}$-admissible if for any sequence $\{x_n\}_{n\in\mathbb{N}}$ with $f_n^{[0,m_n]}(x_n) \subset N$ and $m_n \to \infty$ the sequence of endpoints $\{f_n^{m_n}(x_n)\}_{n\in\mathbb{N}}$ is precompact. In case this property holds for $f_n \equiv f$ then we just say N is f-admissible.

Remark. Later on, we deal with dynamical systems on the space of probability measures $Y = \mathcal{P}(X)$ for some metric space X. An invariant measure for that system is invariant w.r.t. the definition above. In particular, periodic measures will be called invariant. A fixed point for these systems will be called stationary measure.

In the following we will use several ideas from [MR91]: Let N, N' be two f-admissible isolating neighborhood for some isolated invariant set K with

$$N \subset \operatorname{int} N' \cap f^{-1}(\operatorname{int} N').$$

The authors in [MR91, 4.4] used a so called Lyapunov pair (ϕ, γ) which is continuous on a small neighborhood $W \subset N$ of K and has the following properties: $K \subset \gamma^{-1}(0)$, ϕ (resp. γ) is decreasing (resp. increasing) along orbits and $\phi(x) = 0$ with $x \in W$ implies

$$x \in A^-(N) \cup \partial N'.$$

Because K is compact we can choose $d(W, \partial N') > 0$ and assume $x \in A^-(N)$ whenever $\phi(x) = 0$. Furthermore, it is shown that if $\phi(x_n) \to 0$ then x_n admits a convergent subsequence.

Theorem 1. *Suppose N is an isolating neighborhood for K such that the assumptions above hold and*

$$A^-(N) = A(N) = K \neq \varnothing.$$

Then there exists an admissible isolating neighborhood $B \subset N$ which is positive invariant, i.e. no trajectories exit B.

Remark. This is the discrete time version of [Ryb87, I-5.5] using the theory of [MR91]. The proof is essentially copied from Rybakowski using the Lyapunov pair above.

Proof. Define

$$\begin{aligned} P_1^\epsilon &= N \cap \operatorname{cl}\{x \in \operatorname{int} N' \mid \phi(x) < \epsilon\} \\ P_2^\epsilon &= P_1^\epsilon \backslash \{x \in \operatorname{int} N' \mid \gamma(x) < \epsilon\}. \end{aligned}$$

it was shown [MR91, 4.4] that $P_1^\epsilon \subset W$ is a neighborhood of K for sufficiently small $\epsilon > 0$ and whenever $x \in P_i^\epsilon$ and $f(x) \in N$ then $x \in P_i^\epsilon$ and if $x \in P_1^\epsilon$ and $f(x) \notin N$ then $x \in P_2^\epsilon$, i.e. P_2^ϵ is the exit ramp for P_1^ϵ.

Now fix a sufficiently small $\epsilon > 0$ and let $0 < \delta \le \epsilon$ then $P_1^\delta \subset P_1^\epsilon$. Define $\tilde{P}_2^\delta := P_1^\delta \cap P_2^\epsilon$ then because ϕ is decreasing along orbits $\tilde{P}_2^\delta$ is still an exit ramp for P_2^δ.

If $A^-(N) = A(N) = K$ then we claim that there is a $\delta > 0$ such that $\tilde{P}_2^\delta = \varnothing$ which implies that P_1^δ is positive invariant. If this does not hold then there is a sequence $x_n \in P_1^{\delta_n} \cap P_2^\epsilon$ with $\delta_n \to 0$. Thus $\phi(x_n) \to 0$ and $\gamma(x_n) \ge \epsilon$ which implies that there is a subsequence $x_{n'} \to x \in N$ such that $\phi(x) = 0$. Hence

$$x \in A^-(N) = K$$

and $\gamma(x) = 0$ by assumption. But γ is continuous and $x_{n'} \to x$ implies $\epsilon \le \gamma(x_{n'}) \to \gamma(x) = 0$ which is a contradiction. This proofs our claim and thus the theorem. □

In the following we will assume that K satisfies the assumption of the theorem and that $B := P_1^{\delta_0}$, ϕ and γ are given as in the proof. It is obvious that P_1^δ is positive invariant w.r.t. f for any $0 < \delta \le \delta_0$. Furthermore, suppose $f_n \to f$.

Theorem 2. *Assume N' (see above) is $\{f_{n_m}\}$-admissible for each subsequence of $\{f_n\}_{n\in\mathbb{N}}$. Set $\tilde{U} = \operatorname{int} B$ and define*

$$V(a) = \{x \in \tilde{U} \mid \phi(x) < a\}.$$

Then for some $a_0 > 0$, $N := \operatorname{cl} V(a_0) \subset \tilde{U}$. Furthermore, for some sufficiently small $\epsilon_0 > 0$ and all $0 < \epsilon \le \epsilon_0$ there is an $n_0 = n_0(\epsilon)$ such that for all $n \ge n_0$ there is a positive f_n-invariant closed $N_n(\epsilon)$ and

$$K_n \subset V(\epsilon) \subset N_n(\epsilon) \subset N.$$

Remark. The complete continuation theorem for index pairs does not hold for discrete time dynamical systems in general. A proof would require that there is a neighborhood such that the exit time is continuous in $\tilde{U}$, i.e. $\omega_n^+(x_n) \to \omega^+(x_0)$ whenever $x_n \to x_0$ in $\tilde{U}$, which holds for semiflows only for so called isolating blocks. These blocks do not necessarily exist for continuous maps.

Proof. By [MR91, 3.9] there is an $a_0 > 0$ such that $N = \operatorname{cl} V(a_0) \subset \tilde{U}$. And similar to [Ryb87, I-4.5] we can show that for $0 < \epsilon \le a_0$ and all $n \ge n_0(\epsilon)$

$$K_n \subset V(\epsilon).$$

Define

$$N_n(\epsilon) = N \cap \operatorname{cl}\{y \mid \exists x \in V(\epsilon),\, m \ge 0 \text{ s.t. } f_n^{[0,m]}(x) \subset \tilde{U} \text{ and } f_n^m(x) = y\}.$$

Following the proof of [Ryb87, I-12.5] we can show that $N_n(\epsilon)$ satisfies the following properties for $n \ge n_0(\epsilon)$

- $x \in N_n(\epsilon)$ and $f_n(x) \in N$ implies $f_n(x) \in N_n(\epsilon)$
- $K_n \subset V(\epsilon) \subset N_n(\epsilon)$

We claim that for small $\epsilon_0 > 0$ whenever $\epsilon \le \epsilon_0$ and $n \ge n_0(\epsilon)$ then $N_n(\epsilon)$ is positive invariant w.r.t. f_n. If this is not true then there is a sequence $\epsilon_m \to 0$ and

$$y_m \in N_{n_m}(\epsilon_m)$$

with $f_{n_m}(y_m) \notin N$. By definition of $N_{n_m}(\epsilon_m)$ there is a sequence $\tilde{y}_m \in Y$, $x_m \in V(\epsilon_m)$ and $k_m \ge 0$ such that $d(y_m, \tilde{y}_m) < 2^{-m}$, $f_{n_m}^{[0,k_m]}(x_m) \subset \tilde{U}$ and $\tilde{y}_m = f_{n_m}^{k_m}(x_m)$. Because $\phi(x_m) \to 0$ and $A_f^-(B) = A_f(B)$ we can assume w.l.o.g. that $x_m \to x_0 \in A_f(B)$. Admissibility and $f_{n_m} \to f$ imply the sequence $\{f_{n_m}^{k_m}(x_m)\}_{m\in\mathbb{N}}$ has a convergent subsequence and w.l.o.g. $\tilde{y}_m = f_{n_m}^{k_m}(x_m) \to y_0 \in A_f^-(N') = A_f(N') \subset \operatorname{int} N$ and thus $y_m \to y_0$.

Since $f_{n_m}(y_m) \notin N \subset \operatorname{int} N' \cap f^{-1}(\operatorname{int} N')$ and $f_n \to f$ we have $f_{n_m}(y_m) \to f(y_0) \in N' \backslash \operatorname{int} N$. But $y_0 \in A_f(N')$ implies $f(y_0) \in A_f(N')$ which is a contradiction because $A_f(N')$ and $N' \backslash \operatorname{int} N$ are disjoint. ☐

Corollary 3. *Under the assumption above for all $n \ge n_0$ we can find positive f_n-invariant $N_n, N_n^{'}$ such that*

$$N_n \subset U_\delta(K) \subset N_n^{'} \subset B$$

for some δ-neighborhood of K denoted by

$$U_\delta(K) = \{x \in Y \,|\, d(x,y) < \delta \, \text{for some} \, y \in K\}.$$

Furthermore, we have $K \subset \operatorname{int} N_n$ and there is an $\epsilon > 0$ such that

$$U_\epsilon(K_n) \subset N_n^{'}.$$

Proof. Applying the previous theorem we get

$$K \cup K_n \subset V(\tilde{\epsilon}) \subset N_n^{'} \subset B.$$

Recalling the definition of ϕ it is obvious that because $A_f^-(B) = A_f(B) = K$ for small $0 < \epsilon' < 1$ and $x \in P_1^{\epsilon'}$

$$d(x, K) \le \epsilon'.$$

Because K is compact and $V(\tilde{\epsilon})$ a neighborhood of K the δ-neighborhood $U_\delta(K)$ of K is contained in $V(\tilde{\epsilon})$ for δ sufficiently small. Furthermore, we can find an $\epsilon' > 0$ with $\epsilon' < \delta$ such that $P_1^{\epsilon'} \subset U_\delta(K) \subset V(\tilde{\epsilon})$. Applying the theorem again for $P_1^{\epsilon'}$ instead of B we get for $n \ge n_0$

$$N_n \subset P_1^{\epsilon'} \subset U_\delta(K) \subset N_n^{'} \subset B$$

and $K \subset V(\epsilon') \subset N_n$.

To show that $U_\epsilon(K_n) \subset N_n$ we need another positive f_n-invariant neighborhood $N_n^{''}$. First note that there is a $\delta' > 0$ such that

$$d(x, K) \ge \delta'$$

for all $x \in \partial P^{\epsilon'}$. So if we choose $0 < 2\epsilon < \delta$ then

$$U_\epsilon(P_1^\epsilon) \subset P_1^{\epsilon'}.$$

Applying the previous theorem again we get a positive f_n-invariant isolating $N_n^{''}$ of K_n inside of P_1^ϵ. Hence

$$U_\epsilon(K_n) \subset U_\epsilon(N_n^{''}) \subset U_\epsilon(P_1^\epsilon) \subset P_1^{\epsilon'} \subset N_n^{'}.$$

□

Now we are able to continue the attractor. Instead of an admissibility assumption for the perturbed map we will use weak compactness of close δ-neighborhoods of compact sets.

Definition 4 (weak attractor)**.** Suppose Y has a weaker (Hausdorff) topology (i.e. $x_n \to x$ strongly implies $x_n \rightharpoonup x$ weakly) and f is continuous and weakly continuous. An isolated invariant set K is called a weak attractor if it admits a positive f-invariant isolating neighborhood N such that $\omega^{\text{weak}}(N) \subset K$ where $\omega^{\text{weak}}(N)$ is defined as

$$\omega^{\text{weak}}(N) = \{y \in Y \,|\, \exists x_n \in N, m_n \to \infty \,\text{s.t.}\;\; f^{m_n}(x_n) \rightharpoonup y\}.$$

Remark. (1) Our definition of weakness of an attractor is w.r.t. the weaker topology and is different from one defined in [Hur01]. Even our definition of a (strong) attractor is weaker than the one used there because we only require the existence of a positive invariant isolating neighborhood of the invariant set. But there might be a connection to Ochs' weak random attractor [Och99].

(2) A Conley theory with weak-admissibility instead of admissibility might not make sense since the continuation proof requires continuity of the metric and usually the metric is only lower semicontinuous w.r.t. weak convergence.

(3) A weakly continuous function might not be continuous and vice versa (see counterexample in the proof of theorem 8)

Theorem 5. *Under the assumption of the previous theorem, suppose there is a weaker (Hausdorff) topology on Y and that (strongly) closed δ-neighborhoods of compact sets are weakly (sequentially) compact, i.e.* $\operatorname{cl} U_\delta(C)$ *is weakly compact for compact C. If f_n is weakly continuous then K_n is non-empty and a weakly compact weak attractor w.r.t. f_n for all $n \geq n_0$. Furthermore, $U_\epsilon(K_n) \subset N_n^{'}$ for some $\epsilon > 0$ and positive f_n-invariant $N_n^{'}$.*

Proof. Applying the previous corollary we get

$$N_n \subset U_\delta(K) \subset N_n^{'} \subset B$$

and

$$U_\epsilon(K_n) \subset N_n^{'}.$$

Because f_n is weakly continuous, N_n is positive f_n-invariant and $\operatorname{cl} U_\delta(K) \subset N_n^{'}$ is closed and thus weakly compact the set

$$\omega_n^{\text{weak}}(x) = \{y \in Y \,|\, f_n^{n_k}(x) \rightharpoonup y \,\text{for some}\, n_k \to \infty\}$$

for $x \in N_n$ is non-empty and weakly compact. This implies $K_n \neq \varnothing$ and in particular

$$\omega_n^{\text{weak}}(x) \subset K_n \subset N_n.$$

Similarly weak compactness of $\operatorname{cl} U_\delta(K)$ implies $\omega_n^{\text{weak}}(N_n) \subset A_n(N_n^{'}) = K_n \subset N_n$ and thus K_n is a weak attractor. Obviously K_n is weakly closed and contained in the weakly compact set $\operatorname{cl} U_\delta(K)$ and is therefore weakly compact as well. □

2. Wasserstein spaces

Now we will introduce some notation and results for Wasserstein spaces of a metric space, general references are [AGS08] and [Vil09].

Let (X,d) be a complete separable metric space, also called Polish space. We call it proper if every bounded closed set is compact. In particular, this implies that X is locally compact. The metric of a non-compact proper metric space is necessarily unbounded.

The space of probability measures on the Borel σ-algebra of X is denoted by $\mathcal{P}(X)$. This space is given the weak topology, i.e. $\mu_n \rightharpoonup \mu$ if $\int f d\mu_n \to \int f d\mu$ for all bounded continuous functions f . Let x_0 be an arbitrary point of X and define $\mathcal{P}_p(X)$, the Wasserstein space (of order p), by

$$\mathcal{P}_p(X) = \{\mu \in \mathcal{P}(X) \,|\, \int d(x_0,x)^p d\mu(x) < \infty\}.$$

Furthermore, define for $\mu,\nu \in \mathcal{P}_p(X)$

$$w_p(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int d(x,y)^p d\pi(x,y)\right)^{\frac{1}{p}}$$

where $\pi \in \Pi(\mu,\nu) \subset \mathcal{P}(X \times X)$ with $\pi(A \times X) = \mu$ and $\pi(X \times A) = \nu$ for all Borel sets A and B. Then $(\mathcal{P}_p(X), w_p)$ is a complete separable metric space. This topology is usually stronger than the induced subspace topology of $\mathcal{P}_p(X) \subset \mathcal{P}(X)$.

If X is compact so is $\mathcal{P}_p(X)$. And $\mathcal{P}_p(X)$ is local compact only if X is compact. A counterexample for non-proper metric spaces is given in [AGS08, 7.1.9]. We will adjust their example to non-compact proper metric spaces by showing that the closed ϵ-ball $B_\epsilon^{w_p}(\delta_{x_0})$ around δ_{x_0} in $\mathcal{P}_p(X)$ cannot be compact for any $\epsilon > 0$ and thus $\mathcal{P}_p(X)$ cannot be locally compact.

Example. Assume X is non-compact and proper and define

$$\mu_n = m_n\delta_{x_n} + (1-m_n)\delta_{x_0}$$

for some sequence $\{x_n\}_{x\in\mathbb{N}} \subset X$. Then

$$w_p(\mu_n, \delta_{x_0})^p = m_n d(x_n, x_0)^p.$$

Suppose $d(x_n,x_0) \geq \epsilon > 0$ and set $m_n = \epsilon^p \cdot d(x_n,x_0)^{-p}$ then $\mu_n \in \partial B_\epsilon^{w_p}(\delta_{x_0})$. If $\{m_n\}_{n\in\mathbb{N}}$ stays bounded away from 0 then $\{d(x_n,x_0)\}_{n\in\mathbb{N}}$ is bounded and thus $\{x_n\}_{n\in\mathbb{N}}$ and $\{m_n\}_{n\in\mathbb{N}}$ have convergent subsequences $x_{n'} \to x_\infty$ and $m_{n'} \to m \in (0,1]$ and $\mu_n \to m\delta_{x_\infty} + (1-m)\delta_{x_0}$ strongly in $\mathcal{P}_p(X)$. But if we assume $d(x_n,x_0) \to \infty$ then $m_n \to 0$ and thus $\mu_n \rightharpoonup \delta_{x_0}$ weakly. Because strong convergence requires that $w_p(\mu_n,\delta_{x_0}) = \epsilon$ converges to 0 the sequence cannot converge strongly in $\mathcal{P}_p(X)$.

Even though Wasserstein spaces are in general not locally compact we can still show that the following holds for proper metric spaces. The result is probably known or at least implicitly used in case $X = \mathbb{R}^n$. Because it will be our main reason why the "weak" Conley theory is applicable and because we couldn't find any reference, we will prove it completely.

Theorem 6. *If (X,d) is a proper metric space then all closed δ-neighborhoods of compact sets in $(\mathcal{P}_1(X), w_1)$ are weakly compact, where the weak topology of $\mathcal{P}_1(X)$ is the induced subspace topology $\mathcal{P}_1(X) \subset \mathcal{P}(X)$. A space, e.g. $\mathcal{P}_1(X)$, having this property may be called weakly proper.*

Corollary 7. *For $1 \leq p < q$ closed δ-neighborhoods of compact sets in $(\mathcal{P}_q(X), w_q)$ are compact in $(\mathcal{P}_p(X), w_p)$, i.e. $\mathcal{P}_q(X)$ is weakly proper w.r.t. the subspace topology induced by $\mathcal{P}_q(X) \subset \mathcal{P}_p(X)$.*

Remark. This is stronger then a compact embedding $i : (\mathcal{P}_q(X), w_q) \to (\mathcal{P}_p(X), w_p)$ because if $\mu_n \in \mathcal{P}_q(X)$ is bounded then w.l.o.g. $i(\mu_n) \to \mu_*$ in $\mathcal{P}_p(X)$ and necessarily $\mu_* \in \mathcal{P}_q(X)$, i.e. bounded sequences never "leave" the space.

Proof of theorem. We will show that

$$B_r^w := \{\nu \in \mathcal{P}_1(X) \,|\, w_1(\nu, \delta_{x_0}) \leq r\}$$

is weakly compact for all $r \geq 0$. Since B_r^w is closed and $\mu_n \rightharpoonup \mu$ implies $w_1(\mu_n, \delta_{x_0}) \leq \liminf_{n\to\infty} w_1(\mu_n, \delta_{x_0}) \leq r$ we only need to show that B_r^w is tight.

Tightness of a subset $\mathcal{K} \subset \mathcal{P}_1(X)$ means for all $\epsilon > 0$ there is a compact K_ϵ such that for all $\mu \in \mathcal{K}$

$$\mu(X \backslash K_\epsilon) \leq \epsilon.$$

For $\mu \in B_r^w$ we have

$$\int_X d(x, x_0) d\mu = w_1(\mu, \delta_{x_0}) \leq r.$$

Now choose $K_\epsilon = B_{\frac{r}{\epsilon}}(x_0)$, the closed ball around x_0 with radius $\frac{r}{\epsilon}$, which is compact because X is proper. Then we have

$$\mu(X \backslash K_\epsilon) = \int_{X \backslash K_\epsilon} d\mu(x) \leq \frac{\epsilon}{r} \int_{X \backslash K_\epsilon} d(x, x_0) d\mu(x) \leq \epsilon.$$

Thus the closed ball in $\mathcal{P}_1(X)$ around δ_{x_0} is weakly compact.

Let $K \subset \mathcal{P}_1(X)$ be a compact set , e.g. $K = \{\mu\}$, then the closed R-neighborhood around K is defined as

$$N_R^w(K) = \{\nu \in \mathcal{P}_1(X) \,|\, w_1(\mu, \nu) \leq R \text{ for some } \mu \in K\}.$$

This set is closed and bounded and for some $\tilde{R}$ we have

$$N_R^w(K) \subset B_{\tilde{R}}^w.$$

Let $\nu_n \in N_R^w(K)$ be an arbitrary sequence. Then there are $\mu_n \in K$ with $w_1(\mu_n, \nu_n) \leq R$. Because $B_{\tilde{R}}^w$ is weakly compact and K is compact there are $\nu_\infty \in B_{\tilde{R}}^w$ and $\mu_\infty \in K$ such that for some subsequence (also denoted by μ_n, resp. ν_n)

$$\begin{aligned} \mu_n &\to \mu_\infty \\ \nu_n &\rightharpoonup \nu_\infty. \end{aligned}$$

Since $w_1(\cdot, \cdot)$ is weakly lower semicontinuity we have

$$w_1(\mu, \nu) \leq \liminf_{n\to\infty} w_1(\mu_n, \nu_n) \leq R,$$

i.e. $\nu \in N_R^w(K)$ which implies weak compactness. □

Proof of corollary. We only show that $B_r^{w_q}(\delta_{x_0})$ is weakly compact w.r.t. the induced subspace topology $\mathcal{P}_q(X) \subset \mathcal{P}_p(X)$ for $1 \leq p < q$. The rest will follow by the same arguments used above.

Assume $\{\mu_n\}_{n\in\mathbb{N}} \subset B_r^{w_q}(\delta_{x_0})$. Since $w_q \leq w_1$ the previous theorem implies w.l.o.g. $\mu_n \rightharpoonup \mu_\infty$ for some $\mu_\infty \in \mathcal{P}_1(X)$. Because

$$w_q(\mu_\infty, \delta_{x_0}) \leq \liminf_{n\to\infty} w_q(\mu_n, \delta_{x_0}) \leq r$$

we actually have $\mu_\infty \in B_r^{w_q}(\delta_{x_0}) \subset \mathcal{P}_q(X)$.

Because $1 \le p < q$

$$\int_{X\backslash B_R} d(x,x_0)^p d\mu_n(x) \le \frac{1}{R^{q-p}} \int_{X\backslash B_R} d(x,x_0)^q d\mu_n(x) \le \frac{r}{R^{q-p}}.$$

Hence

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int_{X\backslash B_R} d(x,x_0)^p d\mu_n(x) \le \lim_{R\to\infty} \frac{r}{R^{q-p}} = 0.$$

This and the $\mu_n \rightharpoonup \mu_\infty$ weakly show that $\mu_n \to \mu_\infty$ in $\mathcal{P}_p(X)$ (see [Vil09, 6.8]). □

In the following assume that X is proper. Suppose now $f : X \to X$ is a continuous map having a global (set) attractor, i.e. there is a compact f-invariant $A \subset X$ such that for all bounded sets B

$$\lim_{n\to\infty} \tilde{d}(f^n(B), A) = 0,$$

where $\tilde{d}$ is the semi-Hausdorff metric induced by d such that $\tilde{d}(A,B) = 0$ iff $A \subset \operatorname{cl} B$. The map f induces a continuous map $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$ with $f_*(\mu)(B) = \mu(f^{-1}(B))$ for all Borel set B. Furthermore, under slightly stronger assumptions f_* has a global attractor

$$\mathcal{P}(A) = \{\mu \in \mathcal{P}(X) \,|\, \mu(A) = 1\}.$$

Remark. For compact X the global attractor is always X itself. In particular, since $\mathcal{P}_p(X) = \mathcal{P}(X)$ for $1 \le p < \infty$ is compact the global attractor of $\mathcal{P}(X)$ is the space itself and we don't get new information. The whole theory is only interesting for non-compact proper metric spaces X.

Since the Wasserstein space includes distance the behavior of f at infinity becomes important.

Theorem 8. *The map f_* induces a continuous map $(\mathcal{P}_p(X), w_p) \to (\mathcal{P}_p(X), w_p)$ (also denoted by f_*) if and only if for some $x_0 \in X$*

$$\sup_{x\in X} \frac{d(f(x),x_0)}{1+d(x,x_0)} < \infty.$$

Remark. For a semiflow π we need that $\sup_{x\in X} d(x\pi t,x_0)/1+d(x,x_0) = M_t < \infty$ for $t \in [0,T_0]$. Which means, in particular, that there has to be a global lower bound on the blow-up time and thus there cannot be blow-ups at all, i.e. π has to be a global semiflow. Which implies that the induced semiflow π_* on $\mathcal{P}_p(X)$ is also a global semiflow. A necessary requirement for the existence of a global attractor is an upper bound on M_t for all $t \ge 0$. The requirements in [AGS08, Chapter 8] are sometimes too strong. A sufficient condition is that a one-sided Lipschitz condition holds globally (e.g. $v(x) = x - x^3$ is an unbounded vector field and only locally Lipschitz, but satisfies a one-sided Lipschitz condition).

Proof. Suppose first that

$$M = \sup_{x\in X} \frac{d(f(x),x_0)}{1+d(x,x_0)} < \infty.$$

We can assume w.l.o.g. $M > 0$, otherwise $f|_X \equiv x_0$ and the result is obvious. For $\mu \in \mathcal{P}_p(X)$ we have

$$\begin{aligned}\int d(x,x_0)^p df_*\mu(x) &= \int d(f(x),x_0)^p d\mu(x)\\ &\le M^p \int (1+d(x,x_0))^p d\mu(x) < \infty,\end{aligned}$$

i.e. $f_*(\mu) \in \mathcal{P}_p(X)$. So we only need to show continuity.

Suppose $\mu_n \to \mu$ in $\mathcal{P}_p(X)$ then $f_*\mu_n \rightharpoonup f_*\mu$. Since $d(f(\cdot),x_0)^p$ is continuous and grows at most like $d(\cdot,x_0)^p$ it follows that

$$\begin{aligned}\int d(x,x_0)^p df_*\mu_n(x) &= \int d(f(x),x_0)^p d\mu_n(x)\\ &\longrightarrow \int d(f(x),x_0)^p d\mu(x) = \int d(x,x_0)^p df_*\mu(x).\end{aligned}$$

Thus $f_*\mu_n \to f_*\mu$ (see [Vil09, 6.8]) which shows that $f_* : (\mathcal{P}_p(X), w_p) \to (\mathcal{P}_p(X), w_p)$ is (strongly) continuous.

It remains to show that f_* is not continuous if there is a sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ such that

$$d_n = \frac{d(f(x_n),x_0)}{1+d(x_n,x_0)} \to \infty.$$

Because X is proper and f continuous we must have $d(x_n,x_0) \to \infty$. For large n we can assume $0 < \frac{1}{d_n} < d(x_n,x_0)$. Set $c_n = \frac{1}{d_n}$ then

$$\mu_n = c_n^p \frac{1}{d(x_n,x_0)^p}\delta_{x_n} + (1 - c_n^p \frac{1}{d(x_n,x_0)^p})\delta_{x_0} \in \mathcal{P}_1(X)$$

and $w_p(\mu_n, \delta_{x_0})^p = c_n^p \to 0$, i.e. $\mu_n \to \delta_{x_0}$ strongly in $\mathcal{P}_p(X)$. We have

$$f_*\mu_n = c_n^p \frac{1}{d(x_n,x_0)^p}\delta_{f(x_n)} + (1 - c_n^p \frac{1}{d(x_n,x_0)^p})\delta_{f(x_0)}$$

and therefore

$$w_p(f_*\mu_n, f_*\delta_{x_0})^p = c_n^p \frac{d(f(x_n),x_0)^p}{d(x_n,x_0)^p} = \frac{1+d(x_n,x_0)^p}{d(x_n,x_0)^p} \to 1$$

which implies that f_* cannot be (strongly) continuous in δ_{x_0}. □

The following results will hold for any $\mathcal{P}_p(X)$. To simplify the notation and some of the proofs we will just state them for $\mathcal{P}_1(X)$. Furthermore, we assume from now on that $f_* : \mathcal{P}_1(X) \to \mathcal{P}_1(X)$ is strongly continuous (for short just continuous) and whenever we speak about f_* we mean the map $f_* : \mathcal{P}_1(X) \to \mathcal{P}_1(X)$. Since $f_*(\mathcal{P}_1(X)) \subset \mathcal{P}_1(X)$ (for $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$) this also implies that f_* is weakly continuous in $\mathcal{P}_1(X)$. Similarly we could say that $f_* : \mathcal{P}_q(X) \to \mathcal{P}_q(X)$ is continuous and "weakly" continuous in $\mathcal{P}_q(X)$ w.r.t. the induced subspace topology of $\mathcal{P}_q(X) \subset \mathcal{P}_p(X)$ for any $1 \le p < q$.

Example. Having a global attractor does not imply that f_* is strongly continuous, even finite time compactness is not sufficient: Let X be $\mathbb{R}$ with the Euclidean metric $|\cdot|$. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & x \ge 0 \\ x^2 & x < 0. \end{cases}$$

Then f is continuous and $f^2 \equiv 0$ but for $x_n = -n$

$$\frac{d(f(x_n),0)}{1+d(x_n,0)} = \frac{n^2}{1+n} \to \infty,$$

i.e. f_* is not continuous on $(\mathcal{P}_p(X), w_p)$.

If $K \subset \mathcal{P}_1(X)$ is invariant w.r.t. f_* then it is invariant w.r.t. $f_* : \mathcal{P}(X) \to \mathcal{P}(X)$. Which implies that all measures in K are supported on the global attractor A of f. Since $\mathcal{P}(A) \subset \mathcal{P}_1(X)$ the maximal invariant set of $\mathcal{P}_1(X)$ is $\mathcal{P}(A) = \mathcal{P}_1(A)$.

Suppose f is finite time compact, i.e. there is an m such that $f^m(X) \subset B_R$ for some compact set B_R. It should be obvious that this implies $K = \mathcal{P}_1(A)$ is the global attractor of f_*. Furthermore, we have the following:

Proposition 9. *Suppose for some $m > 0$,*

$$f^m(X) \subset B_R(x_0).$$

Then f_* is finite time compact and thus any closed set $B^w \subset \mathcal{P}_1(X)$ is f_*-admissible.

Proof. Let $\{\nu_n\}_{n\in\mathbb{N}}$ be any sequence in $f_*^m(\mathcal{P}_1(X))$. Then there is a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ such that $\nu_n = f_*^m(\mu_n)$. $f^m(X) \subset B_R(x_0)$ implies $\operatorname{supp}\nu_n \subset B_R(x_0)$. Thus ν_n is tight and

$$\int_{d(x,x_0)\geq R+\epsilon} d(x,x_0)d\nu_n(x) = 0,$$

i.e. $\{\nu_n\}_{n\in\mathbb{N}}$ has uniformly integrable first moments. Which means that $\{\nu_n\}_{n\in\mathbb{N}}$ has a convergent subsequence. Therefore, $f_*^m(\mathcal{P}_1(X))$ is compact which easily implies admissibility for any closed B^w. □

Proposition 10. *Suppose there is an R_0, $0 \leq c < 1$ and $m > 0$ such that for all $R \geq R_0$*

$$f^m(B_R) \subset B_{cR}.$$

Then any bounded closed set $B^w \subset \mathcal{P}_1(X)$ is f_*-admissible.

Proof. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence in B^w such that $f_*^{[0,m_n]}(\mu_n) \subset B^w$ for some $m_n \to \infty$. Because B^w is bounded we have

$$\int_X d(x,x_0)d\mu_n(x) \leq M.$$

First assume $m_n = k_n \cdot m$ for an unbounded sequence $k_n \in \mathbb{N}$. Then for $R \geq R_0$

$$\begin{aligned}\int \chi_{X\backslash B_R}(x)\cdot d(x,x_0)df_*^{m_n}\mu_n(x) &= \int \chi_{X\backslash B_R}(f^{m_n}(x))\cdot d(f^{m_n}(x),x_0)d\mu_n(x)\\ &\leq \int \chi_{X\backslash B_{c^{-k_n}R}}(x)\cdot c^{k_n}d(x,x_0)d\mu_n(x)\\ &\leq c^{k_n}M \to 0,\end{aligned}$$

which shows that $\{f_*^{m_n}(\mu_n)\}_{n\in\mathbb{N}}$ has uniformly integrable first moments which implies that the sequence has a convergent subsequence.

If $m_n \not\equiv 0 (\operatorname{mod} m)$ then for some $0 \leq l_n < m$ we have $m_n - l_n \equiv 0(\operatorname{mod} m)$. Therefore, if we set $\nu_n = f_*^{l_n}(\mu)$ then the argument above applies to ν_n and the sequence of endpoints (which is equal to $\{f_*^{m_n}(\mu_n)\}_{n\in\mathbb{N}}$) has a convergent subsequence. □

Remark. Kifer used in [Kif88, Theorem 1.7] linear attraction instead of exponential. This might not be sufficient for admissibility. Nevertheless, later we will assume that a Markov-type perturbation of f_* is small in the Wasserstein distance which is stronger than Kifer's assumption and thus an invariant (probability) measure exists for the perturbation by the same theorem. But that theorem does not imply that the perturbed and unperturbed invariant measures are close w.r.t. the Wasserstein distance, the perturbed invariant measures might not even be in the Wasserstein space. So our result improves this sufficiently.

Before we show how to use Conley theory for small Markov-type noise applied to f we give a sufficient condition such that bounded sets in $\mathcal{P}_1(X)$ are $\{F_n\}$-admissible for $F_n \to f_*$.

Proposition 11. *Let B^w be closed and bounded and U be a δ-neighborhood of B^w with f_*-admissible closure. Suppose $F_n \to f_*$ uniformly on some U, i.e.*

$$\sup_{\mu \in U} w_1(f_*(\mu), F_n(\mu)) = \epsilon_n \to 0.$$

If f_* is uniformly continuous in U then B^w is $\{F_n\}$-admissible.

Remark. (1) The idea is to use the uniform convergence and uniform continuity to construct longer and longer orbits of f_* close the the last part of the orbits of F_n, i.e. the orbit $f^{[0,m_n-k_n]}(y_n)$ and $F_n^{[k_n,m_n]}(x_n)$ should be closer and closer and $m_n - k_n \to \infty$ for $y_n = F_n^{k_n}(x_n)$.

(2) Uniform continuity of f_* and uniform convergence of $F_n \to f_*$ are the assumptions Benci [Ben91] used to prove his continuation theorem for the Conley index. Besides having continuous time dynamical systems he also needs invertibility.

Proof. Let $\mu_n \in N$ and $m_n \to \infty$ be sequences with $F_n^{[0,m_n]}(\mu_n)$. Uniform convergence of $F_n \to f_*$ and uniform continuity of f imply that for some $\epsilon(\epsilon_n) \to 0$ as $\epsilon_n \to 0$

$$\begin{aligned} w_1(F_n^2(\mu), f_*^2(\mu)) &\leq w_1(F_n^2(\mu), f_*(F_n(\mu)) + w_1(f_*(F_n(\mu)), f_*^2(\mu)) \\ &\leq \epsilon_n + \epsilon(\epsilon_n) =: \epsilon_{n,2} \to 0 \text{ as } n \to \infty. \end{aligned}$$

Similarly we can show that there are $\epsilon_{n,k} \to 0$ as $n \to \infty$ such that

$$w_1(F_n^k(\mu), f_*^k(\mu)) \leq \epsilon_{n,k} \to 0.$$

Therefore, there is a sequence $k_n \geq 0$ with $m_n - k_n \to \infty$ such that $F_n^{[0,m_n]}(\mu_n) \subset B^w$ implies that

$$f_*^{[0,m_n-k_n]}(\nu_n) \subset U = U_\delta(B^w)$$

for $\nu_n = F_n^{k_n}(\mu_n)$. Furthermore, we can choose k_n such that

$$\delta_n = \max_{k \in [0,m_n-k_n]} \epsilon_{n,k} \to 0$$

and therefore

$$w_1(F_n^{m_n}(\mu_n), f_*^{m_n-k_n}(\nu_n)) \leq \delta_n.$$

Because the closure of U is f_*-admissible, the sequence of endpoints $\{f_*^{m_n-k_n}(\nu_n)\}_{n\in\mathbb{N}} \subset U$ has a convergent subsequence which implies that $\{F_n^{m_n}(\mu_n)\}_{n\in\mathbb{N}}$ has a convergent subsequence, in particular the limit point is in B^w. □

This proposition applies in particular to Lipschitz continuous functions $f : X \to X$ because the induce map $f_* : \mathcal{P}_1(X) \to \mathcal{P}_1(X)$ is Lipschitz continuous as well.

E.g. suppose $X = \mathbb{R}^n$ and f is the time h map of a flow generated by an ODE $\dot{x} = g(x)$ such that g satisfies the one-sided Lipschitz condition for some $M \in \mathbb{R}$

$$\langle x - y, g(x) - g(y) \rangle \le M \|x - y\|^2$$

then f is Lipschitz continuous with constant e^{Mh}.

3. Markov-Wasserstein maps

Definition 12. A Markov-Wasserstein map (MW-map) of order p is a continuous map $P : \mathcal{P}_p(X) \to \mathcal{P}_p(X)$ which is convex linear, i.e. for $\mu, \nu \in \mathcal{P}_p(X)$ and $a \in [0, 1]$

$$P(a\mu + (1 - a)\nu) = aP(\mu) + (1 - a)P(\nu).$$

Suppose P is induced by a kernel $p(dy|x) : X \to \mathcal{P}_p(X)$ (necessarily continuous), i.e.

$$P : d\mu(y) \mapsto \int p(dy|x) d\mu(x).$$

The map $P = P_f^M$ is called an MW-map of order p relative to f with noise level (at most) M if

$$\sup_{x \in X} w_p(p(dy|x), \delta_{f(x)}) \le M.$$

This implies by [Vil09, 4.8]

$$w_p(P(\mu), f_*(\mu))^p \le \int w_p(p(dy|x), \delta_{f(x)})^p d\mu(x) \le M^p.$$

Remark. As in the sections before, the results also hold for semiflows and a suitable definition for MW-semiflows, i.e. a continuous semigroups $(P_t)_{t \ge 0}$ on $\mathcal{P}_p(X)$. The noise level model can be stated similarly, but we only require that it is uniformly small for all "small" t. We will focus here only on maps, resp. MW-maps, because the intuition behind these is easier, but all results also hold for MW-semiflows if the noise level is sufficiently small.

MW-maps (resp. MW-semiflows) appear naturally in the theory of Markov chains (resp. processes). The map $p(dy|\cdot) : X \to \mathcal{P}_p(X)$ is the Markov transition probability function, whereas the map $P : \mathcal{P}_p(X) \to \mathcal{P}_p(X)$ almost never has a name. If $P = f_*$ for some dynamical system $f : X \to X$ then P is sometimes called transfer map. The following will show that we only need continuity of $p(dy|\cdot)$ to ensure that P is continuous in $\mathcal{P}_p(X)$ (and thus for any $\mathcal{P}_q(X)$, $1 \le q < p$ and for $\mathcal{P}(X)$).

Theorem 13. *Let $p(dy|\cdot) : X \to \mathcal{P}_p(X)$ be a Markov kernel, i.e. a measure-valued map, continuously depending on x. If $M = \sup_{x \in X} w_p(p(dy|x), \delta_x) < \infty$ then P defined by*

$$P : d\mu(y) \mapsto \int p(dy|x) d\mu(x)$$

is an MW-map of order p relative to $\mathrm{id} : X \to X$ *with noise level M.*

Remark. For MW-semiflows weak continuity of $p_t(dy|\cdot)$ corresponds to Feller continuity of the corresponding stochastic process. In fact, if the initial distribution of

$(X_t^n)_{t\geq 0}$ is δ_{x_n}, i.e. $X_0^n = x_n$, then by our continuity requirement if $x_n \to x_0$ then $\mu_t^n = P_t\delta_{x_t} \to P_t\delta_{x_0} = \mu_t^0$ which implies that

$$u(x_n) = \mathbb{E}g(X_t^n) = \int g(x)d\mu_t^n(x) \to \int g(x)d\mu_t^0(x) = \mathbb{E}g(X_t^0) = u(x_0)$$

for all bounded continuous function $g(x)$, i.e. the stochastic process generated by $(P_t)_{t\geq 0}$ is Feller continuous. The continuity of the moments implies that, in addition, the moments are also continuous. This condition could be called p-Feller continuous. This type of continuity does not need $t \in \mathbb{R}$ and thus applies equally to Markov chains, i.e. discrete time stochastic processes.

Proof. Continuous dependency implies that $P : \mathcal{P}_p(X) \to \mathcal{P}(X)$ is continuous. So we only need to show that $\mu_n \to \mu$ in $\mathcal{P}_p(X)$ implies that $P\mu_n \to P\mu$ in $\mathcal{P}_p(X)$

Because $w_p(p(dy|x), \delta_x) \leq M$ for some $M < \infty$ we have

$$\begin{aligned}\int d(y,x_0)^p p(dy|x) &= w_p(p(dy|x), \delta_{x_0})^p \\ &\leq (w_p(p(dy|x), \delta_x) + w_p(\delta_x, \delta_{x_0}))^p \\ &\leq 2^{p-1}(M^p + d(x,x_0)^p)\end{aligned}$$

This implies that $g(x) = \int d(y,x_0)^p p(dy|x)$ grows at most like $d(x,x_0)^p$. Furthermore, g is continuous because $x \mapsto p(dy|x)$ and $\mu \mapsto w_p(\mu, \delta_{x_0})^p$ are. Thus by [Vil09, 6.8]

$$\int d(y,x_0)^p dP\mu_n(x) = \int g(x)d\mu_n(x) \to \int g(x)d\mu(x) = \int d(y,x_0)^p dP\mu(y),$$

i.e. $P\mu_n \to P\mu$ in $\mathcal{P}_p(X)$. ☐

Example. (1) Bounded noise can be modeled by Markov maps with

$$M = \sup_{x\in X} d(x, \operatorname{supp} p(dy|x)) < \infty.$$

Then $w_p(p(dy|x), \delta_x) \leq M$ and thus continuity of $p(dy|\cdot) : X \to \mathcal{P}_p(X)$ for some p implies that of $P : \mathcal{P}_p(X) \to \mathcal{P}_p(X)$. This could also be used to model multi-valued perturbations, i.e. maps $f_n : X \to 2^X$ with $\sup_{x\in X} d(f(x), f_n(x)) \leq M$.

(2) Let $X = \mathbb{R}^n$ with its Euclidean distance. If ν is the standard normal distribution then $\nu = \rho(x)dx$, where dx is the Lebesgue measure on $\mathbb{R}^n$. Any normal distribution with mean x and variance σ^2 can be modeled as follows

$$\mu_{x,\sigma^2} = \delta_x * \rho_\sigma$$

where $\rho_\epsilon(x) = \epsilon^{-n}\rho(x/\epsilon)$. For $\sigma = 0$ we set $\mu_{x,0} = \delta_x$.

Since $m_p = \int |x|^p \rho(x)dx < \infty$ for all p this implies (see [AGS08, 7.1.10]) that

$$w_p(\delta_x, \mu_{x,\sigma^2}) \leq \sigma m_p.$$

Thus Gaussian noise with uniformly small variance is uniformly small in all Wasserstein spaces (although the noise level diverges to ∞ as $p \to \infty$).

Corollary 14. *If $f : X \to X$ induces a continuous self map on $\mathcal{P}_p(X)$ and $p(dy|\cdot) : X \to \mathcal{P}_p(X)$ is continuous with*

$$M = \sup_{x\in X} w_p(p(dy|x), \delta_{f(x)}) < \infty$$

then $P : \mathcal{P}_p(X) \to \mathcal{P}_p(X)$ defined as above is an MW map of order p relative to f with noise level M.

A random perturbation can now be modeled as a composition of f followed by a smearing via $p(dy|\cdot)$, i.e. f_* followed by $P = P_{\mathrm{id}}^M$. This corresponds to additive noise depending only on the image, whereas a general MW-map relative to f might smear the image $f(x)$ and $f(y)$ for $x \neq y$ differently even if $f(x) = f(y)$.

If for some sequence $\tilde{p}_n(dy|\cdot)$ the noise level

$$\sup_{x \in X} w_p(\tilde{p}_n(dy|x), \delta_{f(x)}) = \epsilon_n$$

converges to zero then $\tilde{P}_n$ converges to f_* uniformly on $\mathcal{P}_p(X)$, i.e.

$$\sup_{\mu \in \mathcal{P}_p(X)} w_p(\tilde{P}_n(\mu), f_*(\mu)) \to 0.$$

An MW-chain relative to f satisfies the Markov property, i.e. future behavior only depends on the current state. Furthermore, this models only time-independent random perturbations. Time-dependent perturbations can be modeled with the result of [Kel11]. There it is shown that a local attractor can be continued if the non-autonomous perturbations is uniformly small. Translated into this framework this means

$$\sup_{\mu \in \mathcal{P}_p(X), k \in \mathbb{Z}} w_p(P(\mu, k), f_*(\mu)) < \epsilon$$

for the non-autonomous dynamical system ($\approx$ inhomogeneous Markov map)

$$(\mu, k) \mapsto (P(\mu, k), k+1).$$

Instead of using the semi-admissibility argument to show that the invariant set K_n is non-empty we can use a weak compactness argument to get the same result.

Example. Consider the ODE with $\dot{x} = x - x^3$. This satisfies the one-side Lipschitz condition with $M = 1$ and generates a global semiflow that attracts in finite time. Thus the time-one map induces a Lipschitz continuous map f_* on $\mathcal{P}_p(X)$ which attracts in finite time, too. Therefore, any bounded closed set in $\mathcal{P}_p(X)$ is $\{F_n\}$-admissible for $F_n \to f_*$ uniformly. In particular, the MW-map of order p for small noise level has an attractor close to the original w.r.t. the Wasserstein metric w_p.

Theorem 15. *Suppose $f : X \to X$ induces a dynamical system f_* on $\mathcal{P}_p(X)$ having a global attractor and that f_* is uniformly continuous in a neighborhood of the global attractor. If $P_n \to f_*$ is a sequence of MW-maps of order p relative to f with noise level $\epsilon_n \to 0$. Then for $n \geq n_0$ there is a positive P_n-invariant isolating neighborhood N_n such that $K_n = A_{P_n}(N_n)$ is non-empty and a weakly compact weak attractor which contains all bounded P_n-invariant measures, i.e. $A_{P_n}(P_p(X)) = K_n$. Furthermore, there is at least one stationary measure in K_n.*

Remark. Suppose $p > 1$. Whenever K is (strongly) compact in $\mathcal{P}_p(X)$ then $\operatorname{cl} U_\delta(K)$ is weakly compact w.r.t. the weaker subspace topology of $\mathcal{P}_p(X)$ induced by $\mathcal{P}_p(X) \subset \mathcal{P}_q(X)$ for any $1 \leq q < p$. Thus for all $\mu \in N_n$ there is a $\mu_K \in K_n$

$$P_n(\mu) \xrightarrow{w_q} \mu_K.$$

This means that, although the p-moment may not converge, any q-moment converges for $1 \leq q < p$, but the convergence may get worse the closer q comes to p.

Proof. Everything but the existence of a stationary measure and $A_{P_n}(\mathcal{P}_p(X)) = K_n$ follows from theorem 5. Since P_n is convex linear K_n must be convex. This implies that for any $\mu \in K_n$ the sequence

$$\left\{ \frac{1}{m} \sum_{k=0}^{m-1} P_n^k(\mu) \right\}_{m\in\mathbb{N}}$$

is in K_n and thus weakly converging to some $\nu \in K_n$ and by the Krylov-Bogolyubov theorem it must be a fixed point of P_n, i.e. ν is a stationary measure of P_n. Furthermore, if R is the distance from the global attractor of f then its mass must decay as R^{-p}.

The invariant measures must all be contained in the interior of N_n. Otherwise take $\mu \in A_{P_n}(\mathcal{P}_p(X))\backslash K_n$. If μ is P_n-stationary then the argument is as follows: For $t \in [0,1]$ and some $\mu_0 \in K_n$ the graph of $t \mapsto t\mu + (1-t)\mu_0$ is stationary and intersects ∂N_n, which implies $K_n = A_{P_n}(N_n)$ intersects ∂N_n. This contradicts the isolatedness of K_n.

For the general case assume $\sigma : \mathbb{Z} \to \mathcal{P}_p(X)$ is a bounded full solution through μ, i.e. $P_n(\sigma(k)) = \sigma(k+1)$ and $\sigma(0) = \mu$. Now define the function $g : \mathbb{Z} \to [0,1]$

$$g : k \mapsto \sup\{t \in [0,1] \,|\, s\sigma(k) + (1-s)\mu_0 \in N_n^{'} \,\text{for all}\, s \in [0,t]\}.$$

Because $N_n^{'}$ is a positive invariant neighborhood of K_n, μ_0 stationary, P_n convex linear and $\{\sigma(k)\}_{k\in\mathbb{Z}}$ bounded and not entirely in $N_n^{'}$ we have

$$0 < \delta \le g(k) \le g(k+1).$$

This implies that

$$T = \inf_{k\le 0} g(k) \ge \delta.$$

Because $\sigma(0) \notin N_n^{'}$ we have $T < 1$ and thus by definition of g

$$w_p(T\sigma(k) + (1-T)\mu_0, \partial N_n^{'}) \to 0 \quad \text{as}\, k \to -\infty$$

and thus there is a $T_1 \le T$ such that $\tilde{\sigma}(k) = T_1\sigma(k) + (1-T_1)\mu_0$ is a full solution in $N_n^{'}$ with

$$w_p(\tilde{\sigma}(k_1), \partial N_n^{'}) \le \frac{\epsilon}{2}$$

for some $k_1 \le 0$. But $\tilde{\sigma}(k_1) \in K_n$ and $U_\epsilon(K_n) \subset N_n^{'}$ which implies that

$$\epsilon \le w_p(\sigma(k_1), \partial N_n^{'}) \le \frac{\epsilon}{2}.$$

This is a contradiction and thus $A_{P_n}(\mathcal{P}_p(X))\backslash K_n = \varnothing$, i.e. K_n contains all bounded invariant measures in $\mathcal{P}_p(X)$. □

Corollary 16. *The positive f_*-invariant isolating neighborhood B and the positive P_n-invariant isolating neighborhood N_n can be chosen convex, i.e. if $\mu_i \in B$ (resp. $\mu_i \in N_n$) for $i = 0,1$ then $\mu_t \in B$ (resp. $\mu_t \in N_n$) for $\mu_t = t\mu_0 + (1-t)\mu_1$ and $t \in [0,1]$.*

Proof. Let $\nu_i, \mu_i \in \mathcal{P}_p(X)$ for $i = 0,1$ and define $\mu_t = t\mu_1 + (1-t)\mu_0$ and $\nu_t = t\nu_1 + (1-t)\nu_0$. Assume

$$w_p(\mu_i, \nu_i) < \epsilon.$$

Then there are optimal transference plans $\pi_i \in \Pi(\mu_i, \nu)$ such that

$$\int d(x,y)^p d\pi_i(x,y) < \epsilon^p.$$

The plan $\pi_t = t\pi_1 + (1-t)\pi_0$ is a transference plan for the pair (μ_t, ν_t) and thus

$$\begin{aligned} w_p(\mu_t,\nu_t)^p &\le \int d(x,y)^p d\pi_t(x,y) \\ &\le t\int d(x,y)^p d\pi_1(x,y) + (1-t)\int d(x,y)^p d\pi_0(x,y) \\ &< t\epsilon^p + (1-t)\epsilon^p = \epsilon^p. \end{aligned}$$

The construction of B is done via a Lyapunov pair (ϕ, γ) essentially measuring a weighted distance of the forward orbit of a point, i.e.

$$F_{N'} : \mu \mapsto \min\{1, w_p(\mu, A^-(N') \cup \partial N')\}$$

and

$$\phi : \mu \mapsto \sup\{(2n+1)F_{N'}(f_*^n(x))/(n+1) \mid n \in \mathbb{N}, n \le \omega_{N'}(x)\}$$

Let P_1^ϵ be defined as in the proof of theorem 1. We can assume that $w_p(P_1^\epsilon, \partial N') > 2\epsilon$ for some for sufficiently small $\epsilon > 0$. Because $A^-(N') = A(N')$ is convex, $\phi(\mu_i) < \epsilon$ for $i = 0,1$ implies

$$\phi(\mu_t) < \epsilon.$$

Hence P_1^ϵ is convex and we can choose $B = P_1^\delta$ for some small $\delta > 0$. Similarly $V(a)$ defined in theorem 2 is convex.

The set N_n was defined as

$$N_n(\epsilon) = N \cap \mathrm{cl}\{y \mid \exists x \in V(\epsilon),\, m \ge 0 \text{ s.t. } P_n^{[0,m]}(x) \subset \tilde{U} \text{ and } P_n^m(x) = y\}$$

where $\tilde{U} = \mathrm{int}\, B$ and $N = \mathrm{cl}\, V(\epsilon_0)$ are convex sets. Hence $N_n(\epsilon)$ is convex . ☐

Corollary 17. *Under the assumption of the previous theorem if $p = 1$ then all orbits of P_n are bounded for $n \ge n_0$, i.e. $P_n^{[0,\infty]}(\mu) \subset B_R^w(\mu_0)$ for some $R \ge 0$ and some fixed μ_0. In particular, K_n is the global weak attractor of P_n.*

Remark. The idea of the proof is to control the distance of μ_1 and μ_0 by the distance of μ_t and μ_1 where μ_0 will be some stationary measure and $t \in (0,1]$ is sufficiently small.

Proof. Using the Kantorovich-Rubinstein formula we have the following equality for $\mu, \nu \in \mathcal{P}_1(X)$

$$w_1(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y) d\pi(x,y) = \sup_{\|\phi\|_{\mathrm{Lip}} \le 1} \left\{ \int \phi d\mu - \int \phi d\nu \right\},$$

i.e. there is a sequence ϕ_k with $\|\phi_k\|_{\mathrm{Lip}} \le 1$ such that $\int \phi_k d\mu - \int \phi_k d\nu \nearrow w_1(\mu,\nu)$. Furthermore, there exist an optimal plan $\pi \in \Pi(\mu,\nu)$ such that the infimum is actually attained.

Choose $\mu_0 \in \mathcal{P}_1(X)$ and define $\mu_t = t\mu_1 + (1-t)\mu_0$ for $t \in [0,1]$ and $\mu_1 \in \mathcal{P}_1(X)$. We claim

$$w_1(\mu_t, \mu_0) = t w_1(\mu_t, \mu_0).$$

Suppose $\pi \in \Pi(\mu_1, \mu_0)$ is the optimal plan and ϕ_k the sequence of Lipschitz maps as above. Then $\tilde{\pi} = t\pi + (1-t)(\mathrm{id}, \mathrm{id})_* \mu_0$ is in $\Pi(\mu_t, \mu_0)$. Thus

$$w_1(\mu_t, \mu_0) \leq \int d(x,y) d\tilde{\pi}(x,y) = t w_1(\mu_1, \mu_0).$$

Furthermore, we have

$$w_1(\mu_t, \mu_0) \geq \int \phi_k d\mu_t - \int \phi_k d\mu_0 = t \left(\int \phi_k d\mu_1 - \int \phi_k d\mu_0 \right).$$

Because the left hand side converges monotonically to $t w_1(\mu_t, \mu_0)$ we have proved our claim.

Now fix some stationary measure $\mu_0 \in K_n$. Since N_n is a neighborhood of K_n there is a $t \in (0,1]$ for all $\mu_1 \in X$ such that μ_t as defined above is in N_n. Thus $P_n^{[0,\infty)}(\mu_t)$ is in N_n and bounded, i.e. $w_1(P_n^k(\mu_t), \mu_0) \leq R$ for some R. Because P_n is convex linear and μ_0 stationary we have $P_n^k(\mu_t) = t P_n^k(\mu_1) + (1-t)\mu_0$ and hence

$$w_1(P_n^k(\mu_1), \mu_0) \quad = \quad \frac{1}{t} w_1(P_n^k(\mu_t), \mu_0) \leq \frac{R}{t},$$

which implies that the orbit of μ_1 is bounded. □

References

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[Ben91] V. Benci, *A new approach to the Morse-Conley theory and some applications*, Annali di Matematica Pura ed Applicata **158** (1991), no. 1, 231–305.

[CF98] H. Crauel and F. Flandoli, *Additive Noise Destroys a Pitchfork Bifurcation*, Journal of Dynamics and Differential Equations **10** (1998), no. 2, 259–274.

[Hur01] M. Hurley, *Weak attractors from Lyapunov functions*, Topology and its Applications **109** (2001), no. 2, 201–210.

[Kel11] M. Kell, *Local attractor continuation of non-autonomously perturbed systems*, preprint (2011).

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

[MR91] M. Mrozek and K.P. Rybakowski, *A cohomological Conley index for maps on metric spaces*, Journal of Differential Equations **90** (1991), no. 1, 143–171.

[Och99] G. Ochs, *Weak Random Attractors*, preprint (1999).

[Ryb87] K.P. Rybakowski, *The Homotopy Index and Partial Differential Equations*, Springer, Berlin, 1987.

[Vil09] C. Villani, *Optimal transport: old and new*, Springer Verlag, 2009.

[You86] L.-S. Young, *Stochastic stability of hyperbolic attractors*, Ergodic Theory and Dynamical Systems **6** (1986), no. 2, 311–319.

Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22-26, D-04103 Leipzig, Germany

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