An optimal transportation approach to the decay of correlations for non-uniformly expanding maps

Benoît Kloeckner

An optimal transportation approach to the decay of correlations for non-uniformly expanding maps

Benoît R. Kloeckner *

July 12, 2022

We consider the transfer operators of non-uniformly expanding maps for potentials of various regularity, and show that a specific property of potentials ("flatness") implies a Ruelle-Perron-Frobenius Theorem and a decay of the transfer operator of the same speed than entailed by the constant potential. The method relies neither on Markov partitions nor on inducing, but on functional analysis and duality, through the simplest principles of optimal transportation.

As an application, we notably show that for any map of the circle which is expanding outside an arbitrarily flat neutral point, the set of Hölder potentials exhibiting a spectral gap is dense in the uniform topology. The method applies in a variety of situation, including Pomeau-Manneville maps with regular enough potentials, or uniformly expanding maps of low regularity with their natural potential; we also recover in a united fashion variants of several previous results.

Foreword

The published article (*Ergodic Theory Dynam. Systems* 40, 2020) contained a significant error in Lemma 2.14, used in the core Theorem 4.1. The present text is a consolidated version of the article, with the error corrected (and a few other minor points improved along the way). In order to fix the error, the assumption on coupling in 4.1 needs to be slightly modified (see also Definition 2.12) and Lemma 2.14 (now numbered 2.15) has been adjusted. Section 5.1 provides a criterion to ensure this new hypothesis in our cases of interest, so that all other results are unaffected. I apologize to readers of the

*Université Paris-Est, Laboratoire d'Analyse et de Matématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-94010, Créteil, France

previous version for this embarrassing mistake, and warmly thank Manuel Stadlbauer for pointing out this error to me.

1 Introduction

Chaotic dynamical systems are characterized by the unpredictability of individual orbits; their study has thus largely focused on the statistical behavior of typical orbits. The ergodic theorem, for example, is simply a law of large numbers for observables evaluated along an orbit. In many cases, it was realized that strong chaotic properties provide nice statistical properties just as much as they prevent the prediction of individual orbits.

An important example of this phenomenon is provided by expanding maps $T : \Omega \to \Omega$ (Ω being e.g. a torus or a symbolic space). Close points being pulled apart by a factor, imprecision grows exponentially fast along an orbit; but if one wants to understand the statistical properties of an orbit, the order in which the points are considered does not matter. One can thus look at the orbit backward: each point x_{t+1} is then followed by one of its inverse image $x_t \in T^{-1}(x_{t+1})$. This time reversal transforms the expanding property into a contraction property, but a choice between several inverse images is added. By making this choice randomly, one gets a Markov chain with a contraction property that can be used to prove nice statistical properties, which in turn can be translated for the original map. The *thermodynamical formalism* can be seen as a general framework in which to deal with such arguments; Markov chains are extended to more general averaging operators acting on a space of observables, called *transfer operators*. These operators $\mathscr{L}_{T,A}$ are constructed using the map T and a weight function A, called a potential, each (good enough) potential leading to an invariant measure μ_A. This measure is called here a Ruelle-Perron-Frobenius (RPF) measure[1] because it is obtained by proving a Ruelle-Perron-Frobenius for the transfer operator. The game then consists in finding suitable hypotheses on the potential and observables ensuring fine properties of the transfer operator, which can be translated into the construction of μ_A and statistical properties for μ_A-almost all orbits.

The most favorable case is when the transfer operator has a spectral gap: then one gets exponential decay of correlations, most usual limit theorems, analyticity of the "pressure", etc. More generally, one is interested in the decay of correlations for the measure μ_A, which can be related to decay of the action of the transfer operator on observables of vanishing μ_A-average.

As is long-known, a spectral gap occurs when the map is uniformly expanding and the potential and observables are Hölder continuous. The case of non-uniformly expanding maps or of less regular potential has then become the object of intense scrutiny. The main goal of the present article is to generalize a method first used in the uniformly expanding case in [KLS15], to prove decay results for transfer operator for non-uniformly expanding maps and potentials and observables of various moduli of continuity (the most classical ones being α-Hölder continuity). Decay of transfer operators in particular implies decay of correlations, but is stronger.

[1] The common name *Gibbs measure* conflicts with other, related but different, notions.

Compared to [KLS15], our main contributions are to introduce a simple yet effective framework to deal with various rates of decay in a uniform fashion (Section 2.4), to give a more central place to Markov chains, and to identify a simple condition on potentials, which we call flatness (Section 2.7), under which the methods of [KLS15] can be expanded to prove a Ruelle-Perron-Frobenius theorem and decay properties.

We use an approach introduced by Denker and Urbansky [DU91] to prove the Ruelle-Perron-Frobenius Theorem, which enables us to "normalize" the potential and to relate the dual operator $\mathscr{L}^*_{T,A}$ to a Markov transition kernel. Then we use coupling arguments to prove a general convergence results for a weighted transition kernel in term of the convergence of the unweighted kernel. This is done using optimal transportation as a powerful tool to implicitly combine couplings. We obtain a spectral gap (and thus exponential decay of correlations) in many cases, but our methods also enable us to obtain polynomial decay in cases where a spectral gap is unlikely to occur. The same methods can be used to obtain intermediate or slower decay rates in suitable cases.

1.1 Results with a spectral gap

While the method is relatively general, we will largely focus in this introduction on an emblematic case, the Pomeau-Manneville family, and to similar maps. Given an exponent $q > 0$, we consider on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ parametrized by $[0, 1)$ the map

$$\begin{aligned} T_q : \mathbb{T} &\to \mathbb{T} \\ x &\mapsto \begin{cases} \big(1 + (2x)^q\big)x & \text{if } x \in [0, \frac{1}{2}] \\ 2x - 1 & \text{if } x \in [\frac{1}{2}, 1) \end{cases} \end{aligned}$$

We denote by $\mathscr{C}^\alpha(\mathbb{T})$ the Banach space of α-Hölder real functions, with the convention that $\mathscr{C}^1$ means Lipschitz rather than continuously differentiable. Our first result is a slight variant of a theorem of Li and Rivera-Letelier (see below).

Theorem A (Spectral gaps in the Pomeau-Manneville family)**.** *If $q < 1$, for any $\alpha, \gamma \in (0, 1)$ such that $\gamma - \alpha \geqslant q$ we let $V = \mathscr{C}^\gamma(\mathbb{T})$; if $q \geqslant 1$, for any $\gamma > q$ and $\alpha \in (0, 1]$ such that $\gamma + 1 - \alpha \geqslant q$, we let V be the linear subspace of $\mathscr{C}^\alpha(\mathbb{T})$ made of potentials A that are continuously differentiable near 0 and such that $A'(r) = O_{r\to 0}(r^\gamma)$.*

In both cases, for any potential $A \in V$ the transfer operator $\mathscr{L}_{T_q,A}$ acting on $\mathscr{C}^\alpha(\mathbb{T})$ satisfies a Ruelle-Perron-Frobenius theorem with a spectral gap. In particular the RPF measure μ_A has exponential decay of correlations for all Hölder observables.[2]

Our method is different from the method of Li and Rivera-Letelier who proved this spectral gap (in the Keller space rather than the Hölder space) in the case $q < 1$, with the same assumption $\gamma > q$ on the Hölder exponent [LRL14b, LRL14a].

[2] Of *any* Hölder exponent: indeed as α decreases the space $\mathscr{C}^\alpha(\mathbb{T})$ becomes *larger*; the loss in taking α small in not in the observables for which one gets exponential decay of correlation (on the contrary), but in the norm for which the transfer operator decays.

Remark 1.1. As Li and Rivera-Letelier observe, the bound $\gamma > q$ is optimal in the case $q < 1$. The first reason is obvious: the potential $\log 1/|T_q'|$ is q-Hölder and its RPF measure, which is the absolutely continuous invariant measure (Acim), only has polynomial decay of correlations. Below we will obtain some result with lower rate of decay, but they will still not be applicable to Acims of intermittent maps: we always use a Ruelle-Perron-Frobenius Theorem on a space of continuous functions, while the density of the Acim is unbounded at 0 in this case.

The present method can be applied to other types of neutral points; for example when the graph of T is less tangent to the diagonal at the neutral fixed point, the loss of regularity between the potential and the space its transfer operator acts upon is lower. To state a precise form of a sample result in this direction, let us introduce a particular family of Banach spaces refining the Hölder family. Given $\alpha \in [0,1)$ and $\beta \in \mathbb{R}$, let $\mathscr{C}^{\alpha+\beta\log}(\Omega)$ be the Banach space of functions $\Omega \to \mathbb{R}$ with modulus of continuity at most a multiple of

$$\omega_{\alpha+\beta\log}(r) = \frac{r^\alpha}{\left(\log \frac{r_0}{r}\right)^\beta}$$

(see Lemma 2.4 for the definition of r_0 and Section 2.2 for the definition of the norm). When β is positive, this imposes slightly more regularity than α-Hölder, and when β is negative slightly less. When $\alpha = 0$, we impose $\beta > 0$ and write simply $\mathscr{C}^{\beta\log}(\Omega)$ (which is a *very loose* modulus of continuity).

Theorem B. *For some $q > 0$, consider the map*

$$T_{q\log} : \mathbb{T} \to \mathbb{T}$$
$$x \mapsto \begin{cases} 0 & \text{if } x = 0 \\ \left(1 + (1 - \log 2x)^{-q}\right)x & \text{if } x \in (0, \frac{1}{2}] \\ 2x - 1 & \text{if } x \in [\frac{1}{2}, 1). \end{cases}$$

For any $\alpha \in (0,1)$ and any $A \in V := \mathscr{C}^{\alpha+q\log}(\mathbb{T})$, the transfer operator $\mathscr{L}_{T_{q\log},A}$ acting on $\mathscr{C}^\alpha(\mathbb{T})$ satisfies a Ruelle-Perron-Frobenius theorem with a spectral gap (in particular as soon as A is Hölder of any exponent, μ_A enjoys exponential decay of correlations for all Hölder observables).

For very neutral points (when the graph of T is very tangent to the diagonal) one cannot merely adjust the regularity requirements without eliminating non-constant potentials; but one can instead ask for pointwise regularity at the neutral fixed point, as in the case $q \geqslant 1$ of Theorem A. To deal with general neutral points, one can simply ask for the maximal possible regularity there.

Theorem C (Density of spectral gap potentials)**.** *Let T be a degree k self-covering of the circle with a neutral fixed point 0, uniformly expanding outside each neighborhood of 0. For any $\alpha \in (0,1]$, let V be the linear space V of $\mathscr{C}^\alpha$ potentials which are constant near the neutral point. Then for all $A \in V$, the transfer operator $\mathscr{L}_{T,A}$ acting on $\mathscr{C}^\alpha(\mathbb{T})$ has a spectral gap.*

For all $\gamma \in (0,\alpha)$, V is dense in $\mathscr{C}^\alpha(\mathbb{T})$ for the γ-Hölder norm.

Note that the set of potentials the transfer operator of whose has a spectral gap is well-known to be open; in particular, this set is thus both open in the $\mathscr{C}^\alpha$ topology and dense in the uniform topology. We do not need to impose any particular behavior at the neutral fixed point, which can be arbitrarily flat. We can for example apply this to examples of Holland [Hol05] (but not to the natural potential yileding the Acim, as Remark 1.1 applies again).

Note that from the conclusions of each of Theorem A or B we have, as in Theorem C, that the set $\mathrm{SG}(T, \mathscr{C}^\alpha)$ of potentials whose transfer operator has a spectral gap contains an open set containing a linear subspace which is dense in the uniform norm.

As a consequence of Theorems A-C we can apply [GKLMF15] under each of their sets of assumptions to easily obtain various results of classical flavor, e.g. providing formulas for successive derivatives of $\int \varphi \mathrm{d}\mu_A$ with respect to A and expressions for the modulus of convexity of the pressure function; and one can also argue as in [KLS15] to show that the maximum entropy measure (or other RPF measures) depends on the map T in a locally Lipschitz way, with respect to a Wasserstein metric W_α (see Corollary 1.5 of [KLS15]; the differentiability assumption is unnecessary in the context of e.g. Theorem A, as one can see in the proof). We refer to these previous articles for these applications, but mention the following result as it needs some adaptation to an argument of [GKLMF15].

Corollary D. *In the context of Theorem A, B or C, for all $A \in V$ the RPF measure μ_A is the unique equilibrium state of the potential A, i.e. it uniquely maximizes*

$$h(\mu) + \int A \, \mathrm{d}\mu$$

among T-invariant measures μ, where h is the Kolmogorov-Sinai entropy.

1.2 Results with polynomial decay

Our method applies readily to uniformly expanding maps and lower-regularity potentials, for example yielding the following.

Theorem E (Polynomial decay in low regularity)**.** *Let $T : \Omega \to \Omega$ be a k-to-1 uniformly expanding map of a compact metric space, and let $\beta \in (1, +\infty)$. For any potential $A \in \mathscr{C}^{\beta \log}(\Omega)$, the transfer operator $\mathscr{L}_{T,A}$ acting on $\mathscr{C}^{(\beta-1)\log}(\Omega)$ satisfies a Ruelle-Perron-Frobenius theorem and the RPF measure has at least polynomial decay of correlations of degree $\beta - 1$.*

This can be generalized to other pairs of modulus of continuity (instead of $\omega_{\beta \log}$ for the potential and $\omega_{(\beta-1)\log}$ for the observables in the statement above).

Theorem E enables us to recover and slightly strengthen a result of Fan and Jiang [FJ01b, FJ01a].

Corollary F (Acim for maps with mildly regular derivative)**.** *Assume Ω is a connected manifold and $T : \Omega \to \Omega$ is continuously differentiable with $JT := |\det DT| : \Omega \to \mathbb{R}$ in $\mathscr{C}^{\beta \log}(\Omega)$ for some $\beta > 1$.*

Then T has an absolutely continuous invariant measure in $\mathscr{C}^{(\beta-1)\log}(\Omega)$, with at least polynomial decay of correlations of degree $\beta - 1$.

Note that compared to Corollary 1 in [FJ01a] we gain a logarithmic factor in the decay of correlations. In fact (as in all our other results) we obtain more than decay of correlations: if $\mathscr{L} = \mathscr{L}_{T,-\log JT}$ is the transfer operator of the natural potential (whose RPF measure is the Acim of T), for all $f \in \mathscr{C}^{(\beta-1)\log}(\Omega)$ with zero average we get

$$\|\mathscr{L}^t f\|_\infty \leqslant \frac{C(f)}{t^{\beta-1}}$$

In particular, if $\beta > 3/2$, we have a polynomial decay of degree above $\frac{1}{2}$ *in the uniform norm*, and the Acim will satisfy a (Functional) Central Limit Theorem, see [TK05].

We can also generalize Theorem A as follows.

Theorem G (Polynomial decay in the Pommeau-Manneville family)**.** *If $q < 1$, for any $\beta > 1$ and any $A \in \mathscr{C}^{q+\beta\log}(\mathbb{T})$, the transfer operator $\mathscr{L}_{T_q,A}$ acting on $\mathscr{C}^{(\beta-1)\log}(\mathbb{T})$ satisfies a Ruelle-Perron-Frobenius theorem, and the RPF measure has a polynomial decay of correlations of degree $\beta - 1$.*

We could also obtain with the same method a similar result for $q \geqslant 1$.

Note that the $\mathscr{C}^{(\beta-1)\log}$ regularity is very mild, which is a strength of Theorems E and G as they apply to many observables, much less regular than Hölder. Note that for example, in the case of a Pommeau-Manneville map T_q with $q < 1$, when A is γ-Hölder for some $\gamma > q$ Theorem A is not applicable to observables in $\mathscr{C}^{\beta\log}$. But then one can use Theorem G since in particular $A \in \mathscr{C}^{q+(\beta+1)\log}$, and then one obtains polynomial decay of correlations for such very weakly regular observables.

1.3 Short account of some previous works

The works on decay of correlations for non-uniformly expanding maps are too numerous to all be cited; let us only mention a few of them in addition the the ones already discussed above. Manneville and Pomeau [PM80] introduced the $q = 1$ case of the family $(T_q)_{q>0}$ as a model for intermittent phenomena observed in the Lorentz model; the thermodynamical formalism has been studied for such intermittent maps at least since the work of Prellberg and Slawny [PS92]. Liverani, Saussol and Vaienti [LSV99] obtained good estimates of the decay of correlation for the Acims of the Pomeau-Manneville family with a simple approach. Young introduced the now called "Young towers" [You98, You99], giving an upper bound on decay of correlation for a wealth of non-uniformly hyperbolic systems. Her method was for example used by Holland [Hol05] to study maps with strongly neutral point, proving in some cases sub-polynomial decay of correlations. Fine statistical properties have notably been established by Gouëzel [Gou05] and Melbourne and Nicol [MN05]. Hu [Hu04] and Sarig [Sar02] proved *lower* bounds on decay of correlations for intermittent maps, refined by Gouëzel [Gou04].

All the above works deals with the Acim, and their results are therefore formally disjoint from our results A-D, G. It is often argued that the Acim or, more generally, the physical measures, are the invariant measures that matter most since Lebesgue-negligible events seem too elusive to be ever witnessed. However, we would like to

stress that Guihéneuf (Theorem 56 in [Gui15]) showed that in some cases (for generic conservative homeomorphisms), *all* invariant measures will in fact be "observed" in numerical simulations. The fact that RPF measures are in many cases equilibrium states for their potential is another reason to study them in general.

Equilibrium states of intermittent maps have notably been studied by Hofbauer and Keller [HK82] and Bruin and Todd [BT08], where they deal with potential that are sufficiently close to being constant. By contrast, our results need no such assumption but rely instead on regularity hypotheses. Note that, as will be seen below and is visible in Theorem A when $q \geqslant 1$ and in Theorem C, we in fact mostly need regularity of the potential near the neutral point; away from the neutral point, the potential can be merely Hölder (and observables are allowed to be arbitrary in the suitable regularity class, without need for a special behavior near the neutral point).

Previously to the aforementioned work by Li and Rivera-Letelier, uniqueness of equilibrium states and exponential decay of correlations where obtained in some cases by Liverani, Saussol and Vaienti [LSV98], using the Hilbert metric on cones. These authors do get a decay of the transfer operator, but in uniform norm with the BV norm of the observable as a factor – this inhomogeneity makes their result intermediate between "naked" exponential decay of correlations and a spectral gap. More recently, Castro and Varandas [CV13] obtained interesting results for a large family of non-uniformly expanding map, but they need the potential to be very close to a constant in Hölder norm (in this direction, see also [Klo17]). Compared to these works, the main features of the present approach thus are: to allow potentials with large variations; to provide spectral gaps in many cases; to be also applicable in situation where decay of correlations are likely not to be more than polynomial. We also expect the method to be flexible enough to be used in a wide array of examples; in any case, the method feels sufficiently different from the ones currently available (such as inducing) to be potentially useful beyond our main results.

Cyr and Sarig [CS09] proved that the spectral gap property for the transfer operator is dense for countable Markov shifts, but in a sense that enables any rich enough Banach space. This differs quite a bit from Theorem C where we get uniform density of the spectral gap property for transfer operators acting on the fixed, natural $\mathscr{C}^\alpha$ Banach space.

The Hölder moduli of continuity are quite ubiquitous in the literature, and we finish by mentioning the work of Lynch [Lyn06], who used Young towers to study the decay of correlations for the Acim of (possibly non-uniformly) expanding maps and observables of various weak regularities, including $\mathscr{C}^{\beta\,\log}$. In the uniformly expanding case, for $\mathscr{C}^{\beta\,\log}$ observables he obtained polynomial decay of correlation but only with a loosely controlled degree, and Theorem E is much more precise in this case. Our method could be adjusted to work with other modulus of continuity as well.

1.4 Structure of the article

In order to both obtain clean and easily stated results and make our method easily applicable in other situations, this article is constructed in layers. Depending on the

case one wants to apply our method to, one may use results stated in a high layer, or may have to start from one of the first layers and specialize it to one's precise case.

The "zeroth" layer, Section 2, sets up notation and definitions of general scope: optimal transportation, moduli of continuity and associated generalized Hölder spaces, general considerations on convergence speed to the fixed point for iterations of weakly contracting maps, couplings of a transition kernel, flatness of a potential. We also explain how Corollary D follows from Theorems A-C and [GKLMF15].

The first layer is centered on transition kernels: in Section 3 we prove the Ruelle-Perron-Frobenius, and in Section 4 we prove our core contraction result, Theorem 4.1.

The second layer, Section 5, specializes this contraction result to transition kernels that are backward random walk for k-to-1 maps. Several Lemmas aiming at proving flatness for various potential in various contexts are proved, and the first layer is condensed into Theorem 5.12 for k-to-1 maps and $\mathscr{C}^{\alpha+\beta\log}$ potentials. The previous Section should be useful in more generality, see Remark 5.13

The third and last layer, Section 6, relies on the previous one to finish the proofs of the results stated in this introduction.

2 General setting and preliminary results

In all the article Ω denotes a compact metric space, with distance function d, and $T : \Omega \to \Omega$ is a map. We denote by $\mathcal{P}(\Omega)$ the set of probability measures on Ω, endowed with the weak-* topology. Given $\mu \in \mathcal{P}(\Omega)$ and a Borel-measurable $f : \Omega \to \mathbb{R}$, we denote the integral of f with respect to μ either by $\int_\Omega f \, \mathrm{d}\mu$ or $\mu(f)$.

If x, y are quantities depending on some parameters, when writing $x \leqslant Cy$ we may dispense from introducing the constant C, which may change from paragraph to paragraph. We will only introduce C more explicitly when we feel there is a risk of confusion, notably on the dependance of C on some of the parameters. Sometimes we will indicate a change in the constant more explicitly, for example writing $x_2 \leqslant C' y_2$. When we prefer to let the constant C completely implicit, we write $x \lesssim y$.

We denote by $\mathbf{1}$ the constant function on Ω with value 1.

2.1 Maps, transfer operators and transition kernels

Following the ideas of thermodynamical formalism, to construct and study invariant measures one considers a Banach algebra $\mathcal{X}$ (i.e. $\mathcal{X}$ is stable by product and $\|fg\| \leqslant \|f\|\|g\|$) of "potentials" $A : \Omega \to \mathbb{R}$ and a Banach space $\mathcal{Y}$ of "observables" $A : \Omega \to \mathbb{R}$ such that for any potential the following "transfer operator" is bounded on $\mathcal{Y}$:

$$\mathscr{L}_{T,A} f(x) = \sum_{T(y)=x} e^{A(y)} f(y) \qquad \forall f \in \mathcal{Y} \tag{1}$$

(we shall let either or both subscripts implicit when T or, more rarely A, is clear from the context). The hypothesis that $\mathcal{X}$ is an algebra is only meant to ensure that $e^A \in \mathcal{X}$

so that we speak indifferently of the regularity of A or e^A; and in most cases $\mathcal{X}$ will be equal to or a subspace of $\mathcal{Y}$.

The above formula is suitable when T is k-to-one, but poses problems otherwise. It could be ill-defined if the number of inverse images is infinite, or could map continuous functions to non-continuous ones if the number of inverse images is not locally constant. These issues can often be dealt with, but solution sometimes feel ad-hoc. It is thus natural to replace the sum with an integral with respect to a Markov transition kernel (all measurability properties shall be considered with respect to the Borel algebra).

Definition 2.1. By a *transition kernel* on Ω we mean a family $\mathsf{M} = (m_x)_{x\in\Omega}$ of probability measures m_x on Ω (we ask $x \mapsto m_x$ to be furthermore Borel measurable).

The transition kernel M is said to be a *backward walk* of the map T if for all[3] $x \in \Omega$, the measure m_x is supported on $T^{-1}(x)$. Given a potential $A \in \mathcal{X}$, we define the *transfer operator* of M with respect to A by

$$\mathscr{L}_{\mathsf{M},A} f(x) = \int_\Omega e^{A(y)} f(y)\, \mathrm{d}m_x(y)$$

(as above we may keep either or both subscripts implicit.) A transition kernel is said to be *compatible* with $\mathcal{X}$ if for all $A \in \mathcal{X}$, the above formula defines a continuous operator $\mathcal{X} \to \mathcal{X}$

As soon as $\mathcal{X}$ is large enough to separate probability measures, i.e.

$$\Big(\forall f \in \mathcal{X} : \mu(f) = \nu(f)\Big) \implies \mu = \nu,$$

one can define by duality the operator $\mathscr{L}^*_{\mathsf{M},A}$ acting on $\mathcal{P}(\Omega)$:

$$\int_\Omega f\, \mathrm{d}\Big(\mathscr{L}^*_{\mathsf{M},A}\mu\Big) = \int_\Omega \mathscr{L}_{\mathsf{M},A} f\, \mathrm{d}\mu \qquad \forall f \in \mathcal{X}.$$

Equivalently,

$$\mathrm{d}\Big(\mathscr{L}^*_{\mathsf{M},A}\mu\Big)(x) = \int_\Omega (e^A\, \mathrm{d}m_x)\, \mathrm{d}\mu(x).$$

A classical computation shows that $\mathscr{L}^t_A$ can be expressed as

$$\mathscr{L}^t_A f(x) = \int e^{A^t(\bar{x})} f(x_t)\, \mathrm{d}m^t_x(\bar{x})$$

where A^t denote the Birkhoff sum:

$$\begin{aligned} A^t : \quad & \Omega^t \to \mathbb{R} \\ & \bar{x} = (x_1, \dots, x_t) \mapsto A(x_1) + \cdots + A(x_t). \end{aligned}$$

[3] This could be replaced by "almost all x" with respect to the RPF measure for most purposes, the problem being that at this point the RPF measure is not known.

Example 2.2. The most classical example is when T is a k-to-one local homeomorphism, m_x is the uniform distribution over $T^{-1}(x)$, and $\mathcal{X}$ is the algebra of continuous, or α-Hölder functions. Then the transfer operator of M is equal to the classical transfer operator of T up to a constant. When T is an unimodal map $[0,1] \to [0,1]$ with critical point c and such T maps each of $[0,c]$ and $[c,1]$ onto $[0,1]$, since $x = 1$ has only one inverse image while the $x < 1$ have two the formula (1) is somewhat inappropriate: the transfer operator would not preserve the space of continuous functions. Taking m_x to be the uniform distribution over $T^{-1}(x)$ solves this inconvenience.

Another classical example is with T a finite-to-one, piecewise continuous map, m_x the uniform distribution over $T^{-1}(x)$, and $\mathcal{X}$ the algebra of BV functions.

While we will be primarily interested in transition kernels which are backward walks for maps with expanding properties, the question of the spectral gap for $\mathscr{L}_{\mathsf{M},A}$ is relevant in all generality.

It is common to single out the potentials A such that $\mathscr{L}_A \mathbf{1} = \mathbf{1}$; these potentials are said to be *normalized*, and can be characterized in several manners. One particularly relevant one is to observe that A is normalized exactly when $e^A \, \mathrm{d}m_x$ is a probability measure for all x, i.e. when $(e^A \, \mathrm{d}m_x)_{x\in\Omega}$ is a transition kernel.

Let $\mathcal{X}$ be a Banach space of functions defined on Ω, whose norm is denoted by $\|\cdot\|$, and assume that $\mathscr{L}_{\mathsf{M},A}$ acts continuously on $\mathcal{X}$

Definition 2.3. We shall say that $\mathscr{L}_{\mathsf{M},A}$ satisfies a Ruelle-Perron-Frobenius (RPF) Theorem on $\mathcal{X}$ if there exist a positive function $h \in \mathcal{X}$ and a positive constant ρ such that $\mathscr{L}_{\mathsf{M},A} h = \rho h$, and there exist a positive, finite measure ν such that $\mathscr{L}^*_{\mathsf{M},A} \nu = \rho \nu$.

Then the positive measure μ_A defined by $\mathrm{d}\mu_A = h \, \mathrm{d}\nu$ (choosing the eigenfunction h so as to make it a probability) is called the *RPF measure* of the potential A.

We will say that $\mathscr{L}_A$ has a spectral gap if it satisfies the RPF Theorem and for some $C \geqslant 1$ and $\delta \in (0,1)$ and all $t \in \mathbb{N}$, all $f \in \mathcal{X}$ such that $\nu(f) = 0$ we have

$$\|\mathscr{L}^t_{\mathsf{M},A} f\| \leqslant C\rho^t (1-\delta)^t \|f\|$$

The definition of spectral gap above may seem formally stronger than more usual definitions, but we will get it in this form and it is the definition needed to apply [GKLMF15]; moreover it can be shown to be equivalent to more standard definitions, see e.g. [Klo17].

A classical computation shows that whenever $\mathcal{X}$ is large enough and M is a backward walk for a map T, the RPF measure μ_A is T-invariant. Since the framework of transition kernels is not completely standard, let us give some details.

First one observes that for all $f \in \mathcal{X}$, we have

$$\mathscr{L}_{\mathsf{M},A}(g \cdot f \circ T)(x) = \int_\Omega e^{A(x_1)} g(x_1) f(\underbrace{T(x_1)}_{=x}) \, \mathrm{d}m_x(x_1) = f(x) \cdot \mathscr{L}_{\mathsf{M},A}(g)(x)$$

Then setting

$$\tilde{A} = A + \log h_A - \log h_A \circ T - \log \rho_A$$

we get a new potential which is normalized and such that $\mathscr{L}_{\mathsf{M},A}$ and $\mathscr{L}_{\mathsf{M},\tilde{A}}$ are conjugated up to a constant. Indeed:

$$\mathscr{L}_{\mathsf{M},\tilde{A}} f(x) = \int e^{A(x_1)} \frac{h(x_1)}{\rho h(T(x_1))} f(x_1)\, \mathrm{d}m_x(x_1) = \frac{1}{\rho h} \mathscr{L}_{\mathsf{M},A}(hf)(x)$$

and in particular

$$\mathscr{L}_{\mathsf{M},\tilde{A}} \mathbf{1} = \frac{1}{\rho h} \mathscr{L}_{\mathsf{M},A} h = \mathbf{1}$$

i.e. $\tilde{A}$ is normalized, which can also be interpreted as having 1 as eigenvalue and $\mathbf{1}$ as eigenfunction. Similarly, one shows that $\mathscr{L}^*_{\mathsf{M},A}$ has eigenprobability μ_A. Then for all $f \in \mathcal{X}$:

$$\int f \circ T \,\mathrm{d}\mu_A = \int f \circ T \,\mathrm{d}\big(\mathscr{L}_{\mathsf{M},\tilde{A}} \mu_A\big) = \int \mathscr{L}_{\mathsf{M},\tilde{A}}(\mathbf{1} \cdot f \circ T)\,\mathrm{d}\mu_A = \int f \,\mathrm{d}\mu_A$$

As soon as $\mathcal{X}$ separates measures, we can deduce that μ_A is T-invariant. To proceed as above, we have implicitly assumed that $\mathscr{L}_{\mathsf{M},\tilde{A}}$ also acts on $\mathcal{X}$; this and the separation property will be easily seen to hold in all our cases of interest.

2.2 Generalized Hölder spaces

In this Section we describe a class of Banach algebras of functions providing a good compromise between generality and simplicity.

It has become customary in dynamical systems to use Hölder spaces, which are nothing else than spaces of functions having a modulus of continuity of power type. It makes about as much sense to consider arbitrary modulus of continuity, and this leads to generalized Hölder spaces.

We will call *modulus of continuity* a continuous, increasing, concave function $\omega : [0, +\infty) \to [0, +\infty)$ such that $\omega(0) = 0$, and say that a function $f : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) is ω-Hölder if for some constant C we have

$$|f(x) - f(y)| \leqslant C\omega(d(x, y)) \qquad \forall x, y \in \Omega.$$

The least such C is denoted by $\mathrm{Hol}_\omega(f)$ and we denote by $\mathscr{C}^\omega(\Omega)$ the space of all such functions, endowed with the norm

$$\|f\|_\omega := \|f\|_\infty + \mathrm{Hol}_\omega(f).$$

This makes $\mathscr{C}^\omega(\Omega)$ a Banach algebra, i.e. it is complete and $\|fg\|_\omega \leqslant \|f\|_\omega \|g\|_\omega$. This is elementary and can be seen directly, or it can be deduced from the analogous result for Lipschitz functions (i.e. the case $\omega(t) = t$): indeed ω-Hölder functions are nothing else than Lipschitz functions with respect to the metric $\omega \circ d$ obtained by composing the original metric d with the modulus ω (this is indeed a metric since ω is concave).

This observation could make one think that we did not gain any generality; but in many cases the modified distance $\omega \circ d$ is less natural than d and, more importantly, we will rely on expansion properties of the map $T : \Omega \to \Omega$, which will be expressed in the metric d; in general these properties are not invariant under such a change of distance. Moreover we will sometimes ask potentials and observable to have different regularities, and will thus consider several moduli of continuity simultaneously.

Note that, Ω having finite diameter, only the germ at zero of ω truly matters: changing ω away from zero only changes $\|\cdot\|_\omega$ into an equivalent norm.

The Hölder moduli of continuity, $\omega_\alpha(t) = t^\alpha$ for $\alpha \in (0, 1]$, are of particular importance. In subscripts, we shall replace ω_α by α. We will use a generalization including powers of logarithm, which defined by the following.

Lemma 2.4. *For all $\alpha \in (0, 1]$ and $\beta \in \mathbb{R}$, there is a modulus of continuity $\omega_{\alpha+\beta\log}$ such that*

$$\omega_{\alpha+\beta\log}(r) \sim \frac{r^\alpha}{|\log r|^\beta} \text{ as } r \to 0$$

and such that for some $\theta \in (0, 1)$ and all $r \in [0, 1]$

$$\omega(r/2) \leqslant \theta\omega(r).$$

Proof. For $r \in (0, 1]$ we write

$$\omega_{\alpha+\beta\log}(r) := \frac{r^\alpha}{(\log \frac{r_0}{r})^\beta}$$

where r_0 is to be determined. Note that for each r_0, the asymptotic behavior at 0 is as desired. If r_0 is large enough, the above formula defines an increasing and concave function on $[0, 1]$, that can thus be extended into a modulus of continuity.

If $\beta \geqslant 0$, we have $\omega(r/2) \leqslant 2^{-\alpha}\omega(r)$ so that we can take $\theta = 2^{-\alpha}$, and we turn to the case $\beta < 0$. Then

$$\omega(r/2) \leqslant \theta\omega(r) \quad \text{with} \quad \theta = 2^{-\alpha}\Big(1 + \frac{\log 2}{\log r_0}\Big)^{-\beta}$$

and it suffices to take r_0 large enough to ensure $\theta < 1$. □

When $\alpha = 0$, we impose $\beta > 0$ and take

$$\omega_{\beta\log}(r) := \Big(\log \frac{r_0}{r}\Big)^{-\beta} \qquad \forall r \in [0, 1]$$

where r_0 is large enough (in function of β) to ensure $\omega_{\beta\log}$ is increasing and concave on $[0, 1]$. Then we extend it arbitrarily (e.g. linearly) to $[0, +\infty)$. Lemma 2.4 does not stand for $\alpha = 0$.

2.3 Proof of Corollary D

Corollary D states that, in the context of any of Theorems A-C, the RPF measure of a potential $A \in V$ (obtained as $\mathrm{d}\mu_A = h_A \, \mathrm{d}\nu_A$ where h_A is positive eigenfunction of $\mathscr{L}_{T,A}$ an ν_A is a probability eigenmeasure of the dual operator) is the unique equilibrium state of A. This is of a very classical flavor, but given the diversity of assumptions one finds in the literature it deserves a proof in our precise situation.

We use [GKLMF15], Section 7, but some adaptation is needed: we don't have a spectral gap for *all* potentials in $\mathscr{C}^\alpha(\mathbb{T})$. However we have a spectral gap for an open set $\mathcal{U}$ containing the linear subspace V, which is $\|\cdot\|_\infty$-dense in $\mathscr{C}^\alpha(\mathbb{T})$, and this will be sufficient.

First, observe that for all $A \in \mathcal{U}$ we have $P(A) = \log \lambda_A$, where P is the topological pressure and λ_A is the leading eigenvalue of $\mathscr{L}_{T,A}$. To see this, one only has to prove that $P(A) = 0$ whenever A is normalized: indeed each potential in $\mathcal{U}$ differs from a normalized potential by a coboundary $\log h_A - \log h_A \circ T$ and a constant $\log \lambda_A$, and both λ_A and $P(A)$ change in the same way when A is added a coboundary and a constant. To see that $P(A) = 0$ whenever A is normalized, one considers the two definitions of pressure based on (n, ε)-separated sets and (n, ε)-spanning sets respectively, and applies them to a set obtained as follows. Let E_0 be a set of $1/\varepsilon$ points regularly spaced on the circle, then let E be the set of all inverse images by T^n of elements of E_0. Since T is non-contracting, E is both (n, ε)-separated and (n, ε)-spanning. The normalization property enables one to group the terms of $\sum_{x \in E} e^{\sum_{i=0}^{n-1} A(T^i(x))}$ into $1/\varepsilon$ sums over each set of inverse images of each element $x_0 \in E_0$, which by normalization each sum to 1. It follows that $P(A) = 0$ whenever A is normalized, and thus $P(A) = \log \lambda_A$.

By Theorem 9.12 in [Wal82] (note that h is upper semi-continuous since T is expansive), one can write

$$h(\mu) = \inf_{\varphi \in \mathscr{C}^0(\mathbb{T})} \left\{ P(\varphi) - \int \varphi \, \mathrm{d}\mu \right\}.$$

Since V is $\|\cdot\|_\infty$-dense in $\mathscr{C}^\alpha(\mathbb{T})$, and therefore dense in $\mathscr{C}^0(\mathbb{T})$, and since P is continuous in the uniform norm ([Wal82] Theorem 9.7) we can rewrite this as

$$h(\mu) = \inf_{B \in V} \left\{ P(B) - \int B \, \mathrm{d}\mu \right\} = \inf_{B \in V} \left\{ \log \lambda_B - \int B \, \mathrm{d}\mu \right\},$$

or yet $h(\mu) = \inf_V H(\mu, B)$ where $H(\mu, B) := \log \lambda_B - \int B \, \mathrm{d}\mu$.

We are trying to maximize the functional $P_A(\mu) := h(\mu) + \int A \, \mathrm{d}\mu$, which amounts to maximizing

$$P_A(\mu) - P(A) = \inf_{B \in V} H(\mu, B) + \int A \, \mathrm{d}\mu - \log \lambda_A = \inf_{B \in V} \big(H(\mu, B) - H(\mu, A) \big)$$

For $\mu \neq \mu_A$, the formula for the derivative of $H(\mu, B)$ with respect to B ([GKLMF15] Proposition 7.2) and the density of V show that there is a $B \in V$ near A such that $H(\mu, B) < H(\mu, A)$, so that $P_A(\mu) < P(A)$. We only have left to prove that $H(\mu_A, B) \geqslant H(\mu_A, A)$ for all $B \in V$; but A is a critical point of $B \mapsto H(\mu_A, B)$ which is a convex functional on V, and we are done.

2.4 Rates of decay

In this section, we establish some notation and a few elementary results to study the rate of decay of the iteration of a map. Let X be a metric space, whose metric shall be denoted by ρ; in the sequel, X will be either the phase space Ω, or a Banach space of potential or observables, or a space of measures.

Let $P : X \to X$ be a map (which will be either an inverse branch of the dynamical system under study, or a transfer operator, or the dual of a transfer operator). To motivate the next definition, consider

$$F(t,r) := \sup_{\substack{\rho(x,y)\leqslant r \\ n\geqslant t}} \rho(P^n(x), P^n(y))$$

and observe that F is non-increasing in t, non-decreasing in r, and satisfies

$$F(t_1+t_2, r) \leqslant F(t_1, F(t_2, r)).$$

This property encompasses a lot of information, as short time information at some scales reflects on longer time information at some other scales. This is what we shall harness to get effective bounds.

Definition 2.5. A non-negative function $F : \mathbb{N} \times (0,R) \to (0,+\infty)$ (where R is a positive number or $+\infty$) is said to be a *decay function* if

i. $F(t,r)$ is non-increasing in t, non-decreasing and concave in r,

ii. $F(t,r) \to 0$ as either $t \to \infty$ or $r \to 0$, the other variable being fixed,

iii. for some $C > 0$ and all t, r: $F(t,r) \leqslant Cr$,

iv. for all t_1, t_2 and r: $F(t_1+t_2, r) \leqslant F(t_1, F(t_2, r))$.

The concavity in r will be important when using optimal transport, as it will enable us to bound above an integral of the decay function by the decay function of an integral. The third condition, corresponding to a uniform Lipschitz condition on the maps $(P^t)_{t\in\mathbb{N}}$, will ensure some uniformity of the behavior of F (it is used implicitly in Lemmas 2.6 and 2.7 below).

We shall say that a map P has decay rate F (meaning implicitly: "at least F"; sometimes we specify "in the metric ρ") if for all t, x, y,

$$\rho(P^t(x), P^t(y)) \leqslant F(t, \rho(x,y)). \tag{2}$$

It will be convenient to introduce for all $\theta \in (0,1)$ the *decay times* of F as

$$\tau_\theta(r) = \min\{t \in \mathbb{N} : F(t,r) \leqslant \theta r\} = \min\{t \in \mathbb{N} : \forall s \geqslant r,\ F(t,s) \leqslant \theta s\}$$

Where the second equality comes from concavity of $F(t,\cdot)$.

The standard notion of half-life corresponds to $\tau_{1/2}$, and is constant in case of exponential decay. More generally, we have the following elementary result.

Lemma 2.6. *The following are equivalent:*

i. there exist $C \geqslant 1$, $\delta \in (0,1)$ *such that* $F(t,r) \leqslant C(1-\delta)^t r$ *for all* t, r,

ii. for some $\theta \in (0,1)$, *there exist* $D > 0$ *such that for all* r: $\tau_\theta(r) \leqslant D$,

iii. for all $\theta \in (0,1)$, *there exist* $D > 0$ *such that for all* r: $\tau_\theta(r) \leqslant D$.

When P is Lipschitz, Lemma 2.6 provides a (uniform) exponential decay for P as soon as for some $\theta \in (0,1)$, some $t_0 \in \mathbb{N}$ and all $x, y \in X$:

$$\rho(P^{t_0}(x), P^{t_0}(y)) \leqslant \theta \rho(x,y).$$

Concerning polynomial decay, we have the following analogue.

Lemma 2.7. *Let* α *be a positive real number. The following are equivalent:*

i. there exist $B \geqslant 1, b \in (0,1)$ *such that* $F(t,r) \leqslant \frac{Br}{(tr^\alpha + b)^{\frac{1}{\alpha}}}$ *for all* t, r,

ii. for some $\theta \in (0,1)$, *there exist* $D > 0$ *such that* $\tau_\theta(r) \leqslant \frac{D}{r^\alpha}$ *for all* r,

iii. for all $\theta \in (0,1)$, *there exist* $D > 0$ *such that* $\tau_\theta(r) \leqslant \frac{D}{r^\alpha}$ *for all* r.

When these conditions are realized, we will say that F is polynomial (with degree $1/\alpha$).

Proof. It is clear that *iii* implies *ii* and that *i* implies *iii*.

Assume *ii* and let us prove *i*. Let $t \in \mathbb{N}$, $r \in [0, \infty)$. Let $k \in \mathbb{N}$ be the largest integer such that

$$t \geqslant \frac{D}{r^\alpha} + \frac{D}{\theta^\alpha r^\alpha} + \cdots + \frac{D}{\theta^{\alpha k} r^\alpha}$$

(taking B large enough, b small enough we can easily deal with the range $t \leqslant D/r^\alpha$, which we thus ignore from now on). We have $F(t,r) \leqslant \theta^k r$ and:

$$\begin{aligned}
t &\leqslant \frac{D}{r^\alpha} + \frac{D}{\theta^\alpha r^\alpha} + \cdots + \frac{D}{\theta^{\alpha(k+1)} r^\alpha} \\
t &\leqslant \frac{D}{r^\alpha} \frac{\theta^{-\alpha(k+2)} - 1}{\theta^{-\alpha} - 1} \\
\theta^{-\alpha(k+2)} &\geqslant t r^\alpha \frac{\theta^{-\alpha} - 1}{D} + 1 \\
\theta^k &\leqslant \frac{1}{\left(t r^\alpha (\theta^\alpha - \theta^{2\alpha})/D + 1\right)^{\frac{1}{\alpha}}} \\
F(t,r) &\leqslant \frac{Br}{(tr^\alpha + b)^{\frac{1}{\alpha}}} \qquad \text{for some } B, b.
\end{aligned}$$

□

We shall only consider the two families of decay functions given in the first items of Lemmas 2.6 and 2.7, but more general decay rates and decay times can be considered.

2.5 Optimal transportation

Optimal transportation is a vast subject, from which we will only use the Wasserstein metric with respect to various distances of the form $\omega \circ d$ where ω is a modulus of continuity. One defines it for any $\mu, \nu \in \mathcal{P}(\Omega)$ as the least cost needed to move μ to ν, when the cost of moving an amount of mass from x to y is $\omega \circ d(x, y)$. Formally:

$$\mathrm{W}_\omega(\mu, \nu) = \inf_{\Pi \in \Gamma(\mu,\nu)} \int_{\Omega \times \Omega} \omega \circ d(x, y) \, \mathrm{d}\Pi(x, y)$$

where $\Gamma(\mu, \nu)$ is the set of probability measures on $\Omega \times \Omega$ with marginals μ and ν; one calls elements of $\Gamma(\mu, \nu)$ *transport plans*, and a transport plan is said to be optimal (with respect to ω) if it realizes the infimum in $\mathrm{W}_\omega(\mu, \nu)$. Optimal transport plans always exist, and W_ω metrizes the weak-$*$ convergence of measures (here the compactness of Ω is used). The space $\mathcal{P}(\Omega)$ endowed with W_ω is compact (and in particular complete).

The Wasserstein metric has the advantage of being flexible: from a upper bound, we can deduce the existence of a transport plan with small cost, and to obtain a upper bound we only have to produce a transport plan with small cost. It makes it quite easy to prove that good transport plans exist by a sequence of small arguments (notably using the fact that W_ω is indeed a metric: it satisfies the triangular inequality) that would otherwise need careful explicit coupling arguments.

To prove lower bounds or to use upper bounds on the Wasserstein metric, an important tool is the Kantorovich duality ensuring that we can rewrite the Wasserstein metric as follows:

$$\mathrm{W}_\omega(\mu, \nu) = \sup_{\mathrm{Hol}_\omega(f) \leqslant 1} |\mu(f) - \nu(f)|.$$

The inequality $\geqslant$ is the easy one, obtained by a direct computation, and is in fact the only one we shall use.

The following generalization of Corollary 5.2 from [KLS15] is simple but central to our methods. Let $\mathscr{L}$ be a positive bounded operator on $\mathscr{C}^\omega(\Omega)$ such that $\mathscr{L}\mathbf{1} = \mathbf{1}$ and let $\mathscr{L}^*$ be its dual, acting on probability measures of Ω. Let F be a decay function and let ω' be a second modulus of continuity.

Proposition 2.8. *If $\mathscr{L}^*$ has decay rate at least F in the distance W_ω (recall (2)), then it has a unique invariant probability measure μ and for all $f \in \mathscr{C}^\omega(\Omega)$ we have*

$$\|\mathscr{L}^t f - \mu(f)\|_\infty \leqslant \mathrm{Hol}_\omega(f) F(t, \omega(\operatorname{diam} \Omega)) \quad \text{and} \quad \mathrm{Hol}_{\omega'}(\mathscr{L}^t f) \leqslant \mathrm{Hol}_\omega(f) F^\omega_{\omega'}(t)$$

where

$$F^\omega_{\omega'}(t) := \sup_{s \in (0, \operatorname{diam} \Omega)} \frac{F(t, \omega(s))}{\omega'(s)}.$$

In particular, when $F(t, r) = C(1 - \delta)^t r$ (where $C \geqslant 1$ and $\delta \in (0, 1)$), we have

$$\|\mathscr{L}^t f - \mu(f)\|_\omega \leqslant C'(1 - \delta)^t \, \mathrm{Hol}_\omega(f)$$

Proof. The existence of an invariant measure is standard, and uniqueness follows from the decay hypothesis.

For all $x \in \Omega$ we can write $\mathscr{L}^t f(x) = \int \mathscr{L}^t f \, \mathrm{d}\delta_x = \int f \, \mathrm{d}\mathscr{L}^{*t}\delta_x$ so that

$$\begin{aligned}\left|\mathscr{L}^t f(x) - \int f \, \mathrm{d}\mu\right| &= \left|\int f \, \mathrm{d}\mathscr{L}^{*t}\delta_x - \int f \, \mathrm{d}\mathscr{L}^{*t}\mu\right| \\ &\leqslant \mathrm{Hol}_\omega(f)\, \mathrm{W}_\omega(\mathscr{L}^{*t}\delta_x, \mathscr{L}^{*t}\mu) \\ &\leqslant \mathrm{Hol}_\omega(f) F(t, \mathrm{W}_\omega(\delta_x, \mu)) \\ &\leqslant \mathrm{Hol}_\omega(f) F(t, \omega(\operatorname{diam}\Omega)).\end{aligned}$$

Similarly, for all $x, y \in \Omega$ we get

$$\begin{aligned}\left|\mathscr{L}^t f(x) - \mathscr{L}^t f(y)\right| &= \left|\int f \, \mathrm{d}\mathscr{L}^{*t}\delta_x - \int f \, \mathrm{d}\mathscr{L}^{*t}\delta_y\right| \\ &\leqslant \mathrm{Hol}_\omega(f)\, \mathrm{W}_\omega(\mathscr{L}^{*t}\delta_x, \mathscr{L}^{*t}\delta_y) \\ &\leqslant \mathrm{Hol}_\omega(f) F(t, \omega \circ d(x, y)) \\ &\leqslant \mathrm{Hol}_\omega(f) F^\omega_{\omega'}(t)\omega' \circ d(x, y).\end{aligned}$$

□

It will prove convenient to first consider Dirac measure to prove decay, and the linearity of Wasserstein metric will enable us to extended our conclusions to all measures.

Lemma 2.9. *If $\mathscr{L}^*$ is a linear operator on finite signed measures which preserves the set of probability measures and ω is a modulus of continuity, then for all $\mu, \nu \in \mathcal{P}(\Omega)$ and all $\Pi \in \Gamma(\mu, \nu)$,*

$$\mathrm{W}_\omega(\mathscr{L}^*\mu, \mathscr{L}^*\nu) \leqslant \int \mathrm{W}_\omega(\mathscr{L}^*\delta_x, \mathscr{L}^*\delta_y) \, \mathrm{d}\Pi(x, y).$$

Proof. For each $x, y \in \Omega$, let $\tilde{\pi}_{x,y}$ be an optimal plan between $\mathscr{L}^*\delta_x$ and $\mathscr{L}^*\delta_y$. Assume the map $(x, y) \mapsto \tilde{\pi}_{x,y}$ is measurable (this can be done, by a standard selection theorem, using that, Ω being compact, the set of optimal plans is always compact). By linearity of $\mathscr{L}^*$ the measure

$$\tilde{\Pi} = \int \tilde{\pi}_{x,y} \, \mathrm{d}\Pi(x, y)$$

is a coupling of $\mathscr{L}^*\mu$ and $\mathscr{L}^*\nu$, so that

$$\mathrm{W}_\omega(\mathscr{L}^*\mu, \mathscr{L}^*\nu) \leqslant \iint \omega \circ d(x_1, y_1) \, \mathrm{d}\tilde{\pi}_{x,y}(x_1, y_1) \, \mathrm{d}\Pi(x, y) = \int \mathrm{W}_\omega(\mathscr{L}^*\delta_x, \mathscr{L}^*\delta_y) \, \mathrm{d}\Pi(x, y)$$

□

2.6 Couplings of a transition kernel

One prominent reason why expanding maps are easily amenable to the thermodynamical formalism is that their inverse branches are contracting. When studying non-uniformly

expanding maps, one is led to relax this contracting condition on branches, which can be done in several ways. Since we are more generally interested in transition kernels, we will need a suitably defined contraction property for these objects, that should fit a given modulus of continuity ω.

Let us first fix some notation. We are given a transition kernel $\mathsf{M} = (m_x)_{x\in\Omega}$ on the compact metric space Ω; given $t \in \mathbb{N}$ and $x \in \Omega$ we denote by m_x^t the measure on Ω^t which is the law of a Markov chain $(X_1, ..., X_t)$ starting at $X_0 = x$ and following the transition kernel M. In other words, denoting by $\bar{x} = (x_1, \dots, x_t)$ points of Ω^t, m_x^t is defined by

$$\int_{\Omega^t} f(\bar{x})\, \mathrm{d}m_x^t(\bar{x}) = \int \cdots \iint f(\bar{x})\, \mathrm{d}m_{x_{t-1}}(x_t)\, \mathrm{d}m_{x_{t-2}}(x_{t-1}) \cdots \mathrm{d}m_x(x_1);$$

we also write

$$(m_x^t)_{x\in\Omega} = \underbrace{(m_x)_{x\in\Omega} \circ (m_x)_{x\in\Omega} \circ \cdots \circ (m_x)_{x\in\Omega}}_{t \text{ times}} \quad \text{or} \quad \mathsf{M}^t = \mathsf{M} \circ \cdots \circ \mathsf{M}.$$

Observe that m_x^t is a probability measure on Ω^t; the measure on Ω giving the law of the n-th element X_n of the Markov chain with transition kernel M started at $X_0 = x$ is then $(e_t)_* m_x^t$, where $e_i : \Omega^t \to \Omega$ is the projection to the i-th factor (aka evaluation at time i) and the $*$ index denotes push forward. It shall be observed that

$$(e_t)_* m_x^t = \mathscr{L}_{\mathsf{M},0}^{*t} \delta_x.$$

Given two probability measures $\bar{\mu}$, $\bar{\nu}$ on Ω^t, we denote by $\Gamma(\bar{\mu}, \bar{\nu})$ the set of probability measures Π on $\Omega^t \times \Omega^t$ whose marginals are $\bar{\mu}$ and $\bar{\nu}$, i.e. such that

$$p_{1*}\Pi = \mu \quad \text{and} \quad p_{2*}\Pi = \nu$$

where p_1, p_2 are the projection maps $\Omega^t \times \Omega^t \to \Omega^t$ on each factor. Then in particular, denoting by $(e_t, e_t) : \Omega_t \times \Omega_t \to \Omega \times \Omega$ the evaluation at time t, $(e_t, e_t)_*\Pi$ is a transport plan between $\mu_t := (e_t)_*\bar{\mu}$ and $\nu_t := (e_t)_*\bar{\nu}$. By abuse of notation, we will sometimes write $\mathrm{W}_\omega(\bar{\mu}, \bar{\nu})$ for $\mathrm{W}_\omega(\mu_t, \nu_t)$.

Definition 2.10. A *coupling* of a transition kernel $\mathsf{M} = (m_x)_{x\in\Omega}$ is a family P of probability measures $\Pi_{x,y}^t \in \Gamma(m_x^t, m_y^t)$ indexed by $t \in \mathbb{N}$ and $x, y \in \Omega$, such that for each t, the map $(x, y) \mapsto \Pi_{x,y}^t$ is Borel-measurable.

A *coupling at time i* of M is defined similarly, restricting to $t = i$. We sometimes call this a *restricted coupling*, and whenever we want to make clear we mean a coupling defined for all times t, we use the term *full coupling*.

In other words, a (full) coupling records a way of pairing trajectories of two Markov chains following the same kernel M but starting at possibly different points x, y. It would make sense to ask for a consistency condition, e.g. $(r_s)_*\Pi_{x,y}^t = \Pi_{x,y}^s$ whenever $s < t$, where $r_s : \Omega^t \to \Omega^s$ is the restriction to the first s coordinates. However we will not actually need such a condition.

Definition 2.11. A coupling $\mathsf{P} = (\Pi^t_{x,y})_{t,x,y}$ (possibly restricted to a time i) of the transition kernel $\mathsf{M} = (m_x)_{x\in\Omega}$ is said to be ω-*Hölder* when there is a constant C such that for all t (only $t = i$ in the restricted case) and all x, y,

$$\int \omega \circ d(x_t, y_t)\, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) \leqslant C\omega \circ d(x, y).$$

We also say that a full coupling is ω-Hölder at step i if its restriction to $t = i$ is ω-Hölder.

A full coupling P is said to have ω-*decay rate* F, where F is a decay function, when for all t and all x, y it holds

$$\int \omega \circ d(x_t, y_t)\, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) \leqslant F(t, \omega \circ d(x, y)).$$

Definition 2.12. Given a coupling P and a normalized potential $A : \Omega \to \mathbb{R}$, we define $\mathsf{P}_A = \left(e^{A^t(\bar{x})}\, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y})\right)_{t,x,y}$ and we extend the previous definition by saying that P_A has ω-*decay rate* F when for some constant $C = C(A)$, for all t and for all x, y

$$\int \omega \circ d(x_t, y_t) e^{A^t(\bar{x})}\, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) \leqslant CF(t, \omega \circ d(x, y)).$$

P is then said to have *stable* ω-*decay rate* F when the above holds for all bounded, normalized A.

The role of the constant C is to allow a factor depending on A, but the decay rate we will consider (exponential or polynomial) are all defined up to a constant anyway. Observe that since A is assumed to be normalized,

$$\int e^{A^t(\bar{x})}\, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) = \int e^{A^t(\bar{x})}\, \mathrm{d}m^t_x(\bar{x}) = 1 \quad \forall t, x, y$$

so that for each (t, x, y), $e^{A^t(\bar{x})}\, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y})$ is a probability measure; however Definition 2.12 really is an extension of Definition 2.11 since P_A is not a coupling: its first marginal is the Markov chain $\mathsf{M}_A = \left(e^{A(x_1)}\, \mathrm{d}m_x(x_1)\right)_x$ but its second marginal is different, and might not even be a Markov chain (in P_A the weight in the pairing of a $\bar{x}$ with a $\bar{y}$ is given by $A^t(\bar{x})$, independently of $\bar{y}$). A sufficient condition to have stable decay is given in Section 5.1.

Let $\mathsf{M} = (m_x)_{x\in\Omega}$ be a transition kernel, $\mathscr{L}_0 = \mathscr{L}_{\mathsf{M},0}$ be the unweighted transfer operator, ω be a continuity modulus and F be a decay function. Then concavity ensures that checking decay only for Dirac masses suffices to obtain decay for all probability measures.

Lemma 2.13. *Let* $\mathsf{P} = (\Pi^t_{x,y})_{x,y,t}$ *be a coupling of* M, $t \in \mathbb{N}$ *and* $\mu, \nu \in \mathcal{P}(\Omega)$. *If* P *is* ω-*Hölder with constant* C, *then*

$$\mathrm{W}_\omega(\mathscr{L}_0^{*t}\mu, \mathscr{L}_0^{*t}\nu) \leqslant C\, \mathrm{W}_\omega(\mu, \nu).$$

If P *has* ω-*decay rate* F, *then*

$$\mathrm{W}_\omega(\mathscr{L}_0^{*t}\mu, \mathscr{L}_0^{*t}\nu) \leqslant F(t, \mathrm{W}_\omega(\mu, \nu))$$

(in other words, $\mathscr{L}_0^*$ *also has decay rate* F *in the metric* W_ω.*)*

Proof. We prove the second part, the first being the same, only simpler. Let Π^0 be an optimal coupling of μ and ν for the cost $\omega \circ d$; then

$$\Pi := \int_{(x,y)\in\Omega\times\Omega} \left((p_t, p_t)_* \, \mathrm{d}\Pi^t_{x,y}\right) \mathrm{d}\Pi^0(x,y) \in \mathcal{P}(\Omega \times \Omega)$$

is a coupling of $\mathscr{L}_0^{*t}\mu$ and $\mathscr{L}_0^{*t}\nu$, so that we have

$$\begin{aligned}
\mathrm{W}_\omega(\mathscr{L}_0^{*t}\mu, \mathscr{L}_0^{*t}\nu) &\leqslant \int \omega \circ d(x_t, y_t) \, \mathrm{d}\Pi(x_t, y_t) \\
&= \iint \omega \circ d(x_t, y_t) \, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) \, \mathrm{d}\Pi^0(x,y) \\
&\leqslant \int F(t, \omega \circ d(x,y)) \, \mathrm{d}\Pi^0(x,y) \\
&\leqslant F\Big(t, \int \omega \circ d(x,y) \, \mathrm{d}\Pi^0(x,y)\Big) \\
&\leqslant F(t, \mathrm{W}_\omega(\mu,\nu))
\end{aligned}$$

where in the penultimate line we used that $F(t,\cdot)$ is concave. □

2.7 Flat potentials

Given a transition kernel $\mathsf{M} = (m_x)_{x\in\Omega}$, finding a coupling $\mathsf{P} = (\Pi^t_{x,y})_{t,x,y}$ with a good decay gives a bound of the form

$$\mathrm{W}_\omega(m^t_x, m^t_y) \leqslant F(t, \omega \circ d(x,y))$$

which will translate into a similar control for the unweighted operator $\mathscr{L}_{\mathsf{M},0}$ acting on $\mathscr{C}^\omega(\Omega)$. This is very simple to achieve, but to extend this to a weighted operator $\mathscr{L}_A = \mathscr{L}_{\mathsf{M},A}$, we will need to ask the potential A to have some special quality.

Definition 2.14. Assume a coupling P is fixed for the transition kernel M. We say that a potential $A \in \mathscr{C}^\omega(\Omega)$ is *flat* (with respect to P and ω) whenever for some constant $C > 0$, for all $t \in \mathbb{N}$, all $x, y \in \Omega$ and $\Pi^t_{x,y}$-almost all $(\bar{x}, \bar{y})$ it holds

$$\left|A^t(\bar{x}) - A^t(\bar{y})\right| \leqslant C\omega \circ d(x,y).$$

The name has been chosen with the case of a map with a neutral point in mind: flatness is related to small variations near the neutral point.

We will see that this condition enables us to prove the Ruelle-Perron-Frobenius theorem for $\mathscr{L}_A$ if the coupling is ω-Hölder at step 1, and to prove in addition a spectral gap if the coupling has exponential ω-decay.

Lemma 2.15. *For all flat normalized potential $A \in \mathscr{C}^\omega(\Omega)$, there exist a constant $B > 0$ such that for all t and $\Pi^t_{x,y}$-almost all $(\bar{x}, \bar{y})$,*

$$e^{A^t(\bar{x}) - B\omega\circ d(x,y)} \leqslant e^{A^t(\bar{y})} \leqslant e^{A^t(\bar{x}) + B\omega\circ d(x,y)}.$$

In particular, there exist a constant $C > 0$ such that

$$\frac{1}{C} e^{A^t(\bar{x})} \leqslant e^{A^t(\bar{y})} \leqslant C e^{A^t(\bar{x})}.$$

Proof. By flatness:

$$e^{A^t(\bar{y})} \leqslant e^{|A^t(\bar{y})-A^t(\bar{x})|}e^{A^t(\bar{x})} \leqslant e^{B\omega\circ d(x,y)}e^{A^t(\bar{x})} \leqslant \underbrace{e^{B\omega(\operatorname{diam}\Omega)}}_{C} e^{A^t(\bar{x})}$$

and similarly $e^{A^t(\bar{y})} \geqslant e^{-B\omega\circ d(x,y)}e^{A^t(\bar{x})}$. □

See Section 5.3 for criteria ensuring flatness. Basically, flatness will be shown under a (possibly weak) contraction property by choosing suitably two moduli of continuity: one to control the potential, and the second one to control the Birkhoff sums. The weaker the contraction property, the more gap we will have to take between the two moduli.

3 A Ruelle-Perron-Frobenius theorem

In this Section, we slightly extend the proof given in [KLS15], itself strongly inspired from Denker-Urbanski [DU91], for the existence of a positive eigenfunction for the spectral radius, in as great a generality we could achieve without obscuring the main point. Then we show how to apply this in our present setting.

We consider a Banach algebra $\mathcal{X}$ of continuous, real valued functions on a compact metric space Ω. The norm on $\mathcal{X}$ will be denoted by $\|\cdot\|$ while the uniform norm is $\|\cdot\|_\infty$, and we assume that $\|\cdot\| \gtrsim \|\cdot\|_\infty$ and that $\mathbf{1} \in \mathcal{X}$.

Proposition 3.1. *Assume that the closed balls of $\mathcal{X}$ are sequentially compact in the uniform norm, and let $\mathscr{L} : \mathcal{X} \to \mathcal{X}$ be a linear operator preserving the cone of positive functions, continuous in both norms $\|\cdot\|$ and $\|\cdot\|_\infty$, such that for some constant $C > 0$:*

i. for some $x \in \Omega$ and all n it holds $1/C \leqslant \left(\mathscr{L}^n\mathbf{1}(x)\right)^{\frac{1}{n}} \leqslant C$,

ii. for all $x, y \in \Omega$ and all $n \in \mathbb{N}$ it holds $\mathscr{L}^n\mathbf{1}(x) \leqslant C\mathscr{L}^n\mathbf{1}(y)$,

iii. for all n it holds $\|\mathscr{L}^n\mathbf{1}\| \leqslant C\|\mathscr{L}^n\mathbf{1}\|_\infty$.

Let ρ be the spectral radius of $\mathscr{L}$ in the uniform norm, i.e.

$$\rho = \lim_{n\to\infty}\left(\sup_{f\in\mathcal{X}\setminus\{0\}}\frac{\|\mathscr{L}^n f\|_\infty}{\|f\|_\infty}\right)^{\frac{1}{n}}.$$

Then there exists a positive function $h \in \mathcal{X}$ such that $\mathscr{L}h = \rho h$.

Proof. Note that *ii* implies that *i* holds uniformly for all x, up to changing the constant C. By *i*, the number

$$\tilde{\rho} := \limsup_{n\to\infty}\left(\mathscr{L}^n\mathbf{1}(x)\right)^{\frac{1}{n}}$$

is finite and non zero, and by *ii* it does not depend on the choice of x. Moreover we have $\tilde{\rho} \leqslant \rho$. For all $f \in \mathcal{X}$ such that $\|f\|_\infty = 1$, it holds $-2 < f < 2$ and since $\mathscr{L}$ preserves

the cone of positive functions, we get $-2\mathscr{L}^n\mathbf{1} < \mathscr{L}^n f < 2\mathscr{L}^n\mathbf{1}$ for all n. In particular $\|\mathscr{L}^n f\|_\infty^{\frac{1}{n}} < 2^{\frac{1}{n}}\|\mathscr{L}^n\mathbf{1}\|_\infty^{\frac{1}{n}}$, so that $\rho = \tilde{\rho}$.

We fix $x_0 \in \Omega$; the radius of convergence of the series $\sum_{n\geqslant 1} z^n\mathscr{L}^n\mathbf{1}(x_0)$ is equal to $1/\rho$ (note that we will only consider real values of z). As proved in Denker-Urbanski [DU91] we can chose a positive sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_{n+1}/a_n \to 1$ and the modified series

$$\sum_{n\geqslant 1} a_n z^n \mathscr{L}^n\mathbf{1}(x_0)$$

still has radius of convergence $1/\rho$ and diverges at $z = 1/\rho$. We then set

$$Q(z) = \sum_{n\geqslant 1} a_n z^n \mathscr{L}^n\mathbf{1}(x_0) \quad \text{and} \quad h_z(x) = \frac{1}{Q(z)}\sum_{n\geqslant 1} a_n z^n \mathscr{L}^n\mathbf{1}(x).$$

Assumption *iii* ensures that for all $z < 1/\rho$, we have $h_z \in \mathcal{X}$, and together with assumption *ii* it implies that $\|h_z\|$ is bounded independently of z. Since the closed balls of $\mathcal{X}$ are assumed to be sequentially compact in the uniform norm, we can extract a sequence $z_m \to 1/\rho$ such that $h_m := h_{z_m}$ converges in the uniform norm to some $h \in \mathcal{X}$. The h_z are uniformly bounded away from 0, thus h is positive.

Last we use that $Q(z_m) \to +\infty$ when $m \to \infty$ to deduce that h is the desired eigenfunction. Given any $\varepsilon > 0$, let N_ε be such that for all $n \geqslant N_\varepsilon$ we have $|a_{n-1}/a_n - 1| \leqslant \varepsilon$. Then for all $x \in \Omega$ we have:

$$\begin{aligned}
|\mathscr{L}h(x) - \rho h(x)| &\leqslant \limsup_{m\to\infty} \|\mathscr{L}h_m - \rho h_m\|_\infty \\
&\leqslant \limsup_{m\to\infty} \frac{1}{Q(z_m)} \left| \sum_{n=1}^{N_\varepsilon - 2} a_n z_m^n \mathscr{L}^{n+1}\mathbf{1}(x) - \sum_{n=1}^{N_\varepsilon - 1} a_n z_m^n \rho \mathscr{L}^n\mathbf{1}(x) \right| \\
&\qquad + \limsup_{m\to\infty} \frac{1}{Q(z_m)} \left| \sum_{n\geqslant N_\varepsilon} (a_{n-1} z_m^{n-1} - a_n z_m^n \rho)\mathscr{L}^n\mathbf{1}(x) \right| \\
&\leqslant \limsup_{m\to\infty} \frac{1}{Q(z_m)} \left| \sum_{n\geqslant N_\varepsilon} (a_{n-1} z_m^{n-1} - a_n z_m^n \rho)\mathscr{L}^n\mathbf{1}(x) \right| \\
&\leqslant \limsup_{m\to\infty} \frac{1}{z_m Q(z_m)} \sum_{n\geqslant N_\varepsilon} \left| \frac{a_{n-1}}{a_n} - \rho z_m \right| a_n z_m^n \mathscr{L}^n\mathbf{1}(x) \\
&\leqslant \limsup_{m\to\infty} \frac{\varepsilon + |1 - \rho z_m|}{z_m} \cdot \frac{\sum_{n\geqslant N_\varepsilon} a_n z_m^n \mathscr{L}^n\mathbf{1}(x)}{\sum_{n\geqslant 1} a_n z_m^n \mathscr{L}^n\mathbf{1}(x_0)} \\
&\leqslant \varepsilon\rho C
\end{aligned}$$

hence, $\mathscr{L}h = \rho h$. □

We tried to make what we really use apparent in the statement of Proposition 3.1, but we will only use it in the following particular case.

Proposition 3.2. *Let $\mathsf{M} = (m_x)_{x\in\Omega}$ be a transition kernel on a compact metric space Ω and let ω be a modulus of continuity. Assume that we have a coupling $(\Pi^t_{x,y})_{x,y,t}$ which*

is ω-Hölder at step 1 and that $A \in \mathscr{C}^\omega(\Omega)$ is flat. Then Proposition 3.1 applies to $\mathcal{X} = \mathscr{C}^\omega(\Omega)$ and $\mathscr{L} = \mathscr{L}_{\mathsf{M},A}$, in particular there exist a positive number ρ and a positive function $h \in \mathscr{C}^\omega$ such that $\mathscr{L}_{\mathsf{M},A} h = \rho h$.

The hypothesis on the coupling is very mild, but for general couplings flat potentials might be rare.

Proof. It follows from the Arzelà-Ascoli theorem that closed balls of $\mathscr{C}^\omega(\Omega)$ are sequentially compact in the uniform norm, and by its definition $\mathscr{L}_{\mathsf{M},A}$ preserves the cone of positive functions and acts continuously in the uniform norm; more precisely $\|\mathscr{L}_{\mathsf{M},A} f\|_\infty \leqslant \|e^A\|_\infty \|f\|_\infty$.

Since $\mathscr{C}^\omega(\Omega)$ is a Banach algebra (in particular $e^A \in \mathscr{C}^\omega(\Omega)$), using

$$\int g(x_1)\, \mathrm{d}m_x(x_1) = \int g(x_1)\, \mathrm{d}\Pi^1_{x,y}(x_1, y_1)$$

we get for all $f \in \mathscr{C}^\omega(\Omega)$ and all $x, y \in \Omega$:

$$\begin{aligned}
\left|\mathscr{L}_{\mathsf{M},A} f(x) - \mathscr{L}_{\mathsf{M},A} f(y)\right| &= \left| \int e^{A(x_1)} f(x_1)\, \mathrm{d}\Pi^1_{x,y}(x_1, y_1) - \int e^{A(y_1)} f(y_1)\, \mathrm{d}\Pi^1_{x,y}(x_1, y_1) \right| \\
&\leqslant \int \left| e^{A(x_1)} f(x_1) - e^{A(y_1)} f(y_1) \right| \mathrm{d}\Pi^1_{x,y}(x_1, y_1) \\
&\leqslant \mathrm{Hol}_\omega(e^A f) \int \omega \circ d(x_1, y_1)\, \mathrm{d}\Pi^1_{x,y}(x_1, y_1) \\
&\leqslant \|e^A\|_\omega \|f\|_\omega C \omega \circ d(x, y)
\end{aligned}$$

It follows that $\mathrm{Hol}_\omega(\mathscr{L}_{\mathsf{M},A} f) \leqslant C \|e^A\|_\omega \|f\|_\omega$ and $\mathscr{L}_{\mathsf{M},A}$ acts continuously on $\mathscr{C}^\omega(\Omega)$.

We have left to check conditions *i*, *ii* and *iii* of Proposition 3.1. We have

$$\mathscr{L}^n_{\mathsf{M},A} \mathbf{1}(x) = \int e^{A^n(\bar{x})}\, \mathrm{d}m^n_x(\bar{x})$$

and $A^n(\bar{x})$ lies between $n \min A$ and $n \max A$, so that $e^{\min A} \leqslant \left(\mathscr{L}^n_{\mathsf{M},A} \mathbf{1}(x)\right)^{\frac{1}{n}} \leqslant e^{\max A}$ for all x, proving Condition *i*. Using again that $\Pi^t_{x,y}$ has marginals m^t_x and m^t_y, we have for all x, y:

$$\begin{aligned}
\left|\mathscr{L}^n_{\mathsf{M},A} \mathbf{1}(x) - \mathscr{L}^n_{\mathsf{M},A} \mathbf{1}(y)\right| &= \left| \int e^{A^n(\bar{x})}\, \mathrm{d}\Pi^n_{x,y}(\bar{x}, \bar{y}) - \int e^{A^n(\bar{y})}\, \mathrm{d}\Pi^n_{x,y}(\bar{x}, \bar{y}) \right| \\
&\leqslant \int \left| e^{A^n(\bar{x})} - e^{A^n(\bar{y})} \right| \mathrm{d}\Pi^n_{x,y}(\bar{x}, \bar{y}) \\
&\leqslant \int e^{A^n(\bar{y})} \left| e^{A^n(\bar{x}) - A^n(\bar{y})} - 1 \right| \mathrm{d}\Pi^n_{x,y}(\bar{x}, \bar{y}) \\
&\leqslant C \int e^{A^n(\bar{y})} \left| A^n(\bar{x}) - A^n(\bar{y}) \right| \mathrm{d}\Pi^n_{x,y}(\bar{x}, \bar{y}) \\
&\leqslant C' \omega \circ d(x, y) \int e^{A^n(\bar{y})}\, \mathrm{d}\Pi^n_{x,y}(\bar{x}, \bar{y})
\end{aligned}$$

where the last two lines use flatness of A. Since P is a coupling of M, it follows

$$\left|\mathscr{L}^n_{\mathsf{M},A} \mathbf{1}(x) - \mathscr{L}^n_{\mathsf{M},A} \mathbf{1}(y)\right| \leqslant C' \omega \circ d(x, y) \mathscr{L}_{\mathsf{M},A} \mathbf{1}(y),$$

proving *ii* and *iii*. ∎

Proposition 3.2 is half of the RPF theorem, and the second half (existence of an eigenprobability) will be a consequence of our contraction results (or can be proved directly). As seen in Section 2.1, this half already enables us to normalize the potential A into $\tilde{A} = A + \log h - \log h \circ T - \log \rho$. Note that as soon as composition with T preserves $\mathscr{C}^\omega(\Omega)$ (e.g. if T is Lipschitz), the normalized potential $\tilde{A}$ is still in $\mathscr{C}^\omega(\Omega)$: up to changing h by a constant factor, we can assume $0 < h < 2$ and then $\log h$ develops as a converging power series of h, thus $\log h \in \mathscr{C}^\omega(\Omega)$.

A crucial observation is that if the coupling P is ω-Hölder and A is flat, then $\tilde{A}$ is also flat. Indeed, for all t, x, y and $\Pi^t_{x,y}$-almost all $(\bar{x}, \bar{y})$, we have

$$\begin{aligned}
|\tilde{A}^t(\bar{x}) - \tilde{A}^t(\bar{y})| &\leqslant |A^t(\bar{x}) - A^t(\bar{y})| \\
&\qquad + \Big| \sum_{k=1}^{t} \log h(x_k) - \log h(T(x_k)) - \log h(y_k) + \log h(T(y_k)) \Big| \\
&\leqslant |A^t(\bar{x}) - A^t(\bar{y})| \\
&\qquad + \Big| \sum_{k=1}^{t} \log h(x_k) - \log h(x_{k-1}) - \log h(y_k) + \log h(y_{k-1}) \Big| \\
&\leqslant |A^t(\bar{x}) - A^t(\bar{y})| + |\log h(x_t) - \log h(x) - \log h(y_t) + \log h(y)| \\
&\leqslant |A^t(\bar{x}) - A^t(\bar{y})| + \mathrm{Hol}_\omega(\log h)\Big(\omega \circ d(x_t, y_t) + \omega \circ d(x, y)\Big) \\
&\lesssim \omega \circ d(x, y).
\end{aligned}$$

Let us gather these observations in a statement which reduces the problem to find out whether $\mathscr{L}_{\mathsf{M},A}$ has a spectral gap (or other decay properties) to the case of normalized potentials.

Corollary 3.3. *Let $T : \Omega \to \Omega$ be a map on a compact metric space Ω, let $\mathsf{M} = (m_x)_{x\in\Omega}$ be a backward random walk for T and let ω be a modulus of continuity such that composition with T preserves $\mathscr{C}^\omega(\Omega)$ (this is satisfied if T Lipschitz).*

Assume that we have a coupling $\mathsf{P} = (\Pi^t_{x,y})_{t,x,y}$ which is ω-Hölder and that $A \in \mathscr{C}^\omega(\Omega)$ is flat with respect to P.

Then there is a normalized *and* flat *potential $\tilde{A} \in \mathscr{C}^\omega(\Omega)$ that differs from A by a coboundary and a constant; it follows that $\mathscr{L}_{\mathsf{M},A}$ and $\mathscr{L}_{\mathsf{M},\tilde{A}}$ are conjugated up to a constant. In particular, the property to have a spectral gap is equivalent for $\mathscr{L}_{\mathsf{M},A}$ and $\mathscr{L}_{\mathsf{M},\tilde{A}}$.*

4 The main contraction result

Our core result is the following.

Theorem 4.1. *Let M be a transition kernel on a compact metric space Ω, and let ω be a modulus of continuity. Let $A \in \mathscr{C}^\omega(\Omega)$ be a flat, normalized potential and set $\mathscr{L} = \mathscr{L}_{\mathsf{M},A}$.*

Assume M admits a coupling P such that P_A has ω-decay with decay function F and corresponding half-life $\tau = \tau_{\frac{1}{2}} : (0, +\infty) \to \mathbb{N}$.

Then there exist constants $C > 0$ and $k \in \mathbb{N}$ such that for all $\mu, \nu \in \mathcal{P}(\Omega)$ with $\mathrm{W}_\omega(\mu,\nu) =: r$ it holds

$$\mathrm{W}_\omega\left(\mathscr{L}^{*k\tau(r/k)}\mu, \mathscr{L}^{*k\tau(r/k)}\nu\right) \leqslant \frac{1}{2}\,\mathrm{W}_\omega(\mu,\nu)$$

and

$$\mathrm{W}_\omega\left(\mathscr{L}^{*t}\mu, \mathscr{L}^{*t}\nu\right) \leqslant C\,\mathrm{W}_\omega(\mu,\nu) \quad \forall t \in \mathbb{N}.$$

In particular:

- *if F is exponential then $\tau(r)$ is bounded, and so is $k\tau(r/k)$, so that $\mathscr{L}^*_{\mathsf{M},A}$ decays exponentially in the metric W_ω, and $\mathscr{L}_{\mathsf{M},A}$ has a spectral gap on $\mathscr{C}^\omega(\Omega)$.*
- *if F is polynomial then $\tau(r) \leqslant D/r^\alpha$ so that $k\tau(r/k) \leqslant D'/r^\alpha$ and $\mathscr{L}^*_{\mathsf{M},A}$ decays polynomially, with the same degree.*

It follows from this result that as soon as P_A decays to 0, whatever the speed, $\mathscr{L}^*_{\mathsf{M},A}$ fixes a unique probability measure μ_A, which thus must be the RPF measure of A (so we get the second half of the RPF theorem for flat potentials).

In the set-up of our main results, there is only one reasonable candidate for the coupling P, and to apply Theorem 4.1 our main task will be to identify sufficient conditions on potentials ensuring flatness.

Proof. The proof mostly follows the strategy of [KLS15], the extra generality coming mostly from the use of decay times and moduli of continuity.

Step 1: construct a transport plan between $\mathscr{L}^{*t}\delta_x$ and $\mathscr{L}^{*t}\delta_y$.

Here we need the normalization assumption, to ensure these two measures are both of the same mass. Fix $t \in \mathbb{N}$, $x, y \in \Omega$ and observe that $\mathscr{L}^{*t}\delta_x = (e_t)_*\left(e^{A^t}\,\mathrm{d}m_x^t\right)$ where $e_t : \Omega^t \to \Omega$ is the projection to the last coordinates. We seek an efficient transport plan between $\mathscr{L}^{*t}\delta_x$ and $\mathscr{L}^{*t}\delta_y$, and we will construct it as $(e_t, e_t)_*\Pi$ where Π is a transport plan between $e^{A^t}\,\mathrm{d}m_x^t$ and $e^{A^t}\,\mathrm{d}m_y^t$. What we are given by the coupling P is a transport plan $\Pi^t_{x,y}$ between m_x^t and m_y^t, and we will modify it to take into account the e^{A^t} factors. Define a function

$$\begin{aligned} a : \Omega^t \times \Omega^t &\to \mathbb{R} \\ (\bar{x}, \bar{y}) &\mapsto \min\left(e^{A^t(\bar{x})}, e^{A^t(\bar{y})}\right) \end{aligned}$$

so that $a\,\mathrm{d}\Pi^t_{x,y}$ is a positive measure whose marginals are less than $e^{A^t}\,\mathrm{d}m_x^t$ and $e^{A^t}\,\mathrm{d}m_y^t$, respectively. There must thus exist some positive measure Λ on $\Omega^t \times \Omega^t$ such that

$$\Pi := a\,\mathrm{d}\Pi^t_{x,y} + \Lambda$$

is a probability measure with marginals exactly $e^{A^t}\,\mathrm{d}m_x^t$ and $e^{A^t}\,\mathrm{d}m_y^t$. We want to bound above the ω-cost of Π; the basic idea is that the first term will be small by the decay hypothesis, the second one will be small because Λ has small mass.

Step 2: Bound from above the mass of Λ.

By Lemma 2.15, for $\Pi^t_{x,y}$ almost all $(\bar{x}, \bar{y})$ and for some constant B,

$$a(\bar{x}, \bar{y}) \geqslant e^{A^t(\bar{x})} e^{-B\omega \circ d(x,y)};$$

using that A is normalized it comes that the total mass of $a \, \mathrm{d}\Pi^t_{x,y}$ is at least $e^{-B\omega\circ d(x,y)}$. Since Ω has finite diameter, up to enlarging B this total mass can bounded from below both by a constant $e^{-B} \in (0,1)$ and by $1 - B\omega \circ d(x,y)$. The total mass of Λ is therefore bounded above as follows:

$$\int \mathbf{1} \, \mathrm{d}\Lambda \leqslant \min\left(B\omega \circ d(x,y), 1 - e^{-B}\right).$$

Step 3: Bound the cost of Π for a modified metric.

We introduce a new modulus of continuity

$$\omega' = \min\left(K\omega, \omega(\operatorname{diam}\Omega)\right)$$

where K is a positive constant to be specified later on (independently of x, y). We have $\omega' \circ d(x,y) \geqslant \omega \circ d(x,y)$ for all $x, y \in \Omega$ and $\omega' \leqslant K\omega$, so that $\omega \circ d$ and $\omega' \circ d$ are Lipschitz-equivalent metrics on Ω, and as a consequence W_ω and $\mathrm{W}_{\omega'}$ are Lipschitz-equivalent (with the same constants).

We decompose the cost as

$$\int \omega' \circ d(x_t, y_t) \, \mathrm{d}\Pi(\bar{x}, \bar{y}) = \int \omega' \circ d(x_t, y_t) a(\bar{x}, \bar{y}) \, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) + \int \omega' \circ d(x_t, y_t) \, \mathrm{d}\Lambda(\bar{x}, \bar{y}).$$

For the first term we get from ω-decay of P_A (with D only depending on A):

$$\begin{aligned}\int \omega' \circ d(x_t, y_t) a(\bar{x}, \bar{y}) \, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) &\leqslant K \int \omega \circ d(x_t, y_t) e^{A^t(\bar{x})} \, \mathrm{d}\Pi^t_{x,y}(\bar{x}, \bar{y}) \\ &\leqslant DK \cdot F(t, \omega \circ d(x,y)) \\ &\leqslant DK \cdot F(t, \omega' \circ d(x,y)).\end{aligned}$$

For the second term, we distinguish two cases. If $\omega \circ d(x,y) \geqslant \omega(\operatorname{diam}\Omega)/K$, then $\omega' \circ d(x,y) = \omega(\operatorname{diam}\Omega) = \omega'(\operatorname{diam}\Omega)$ and we bound the mass of Λ by $1 - e^{-B}$, so that

$$\begin{aligned}\int \omega' \circ d(x_t, y_t) \, \mathrm{d}\Lambda(\bar{x}, \bar{y}) &\leqslant (1 - e^{-B}) \omega'(\operatorname{diam}\Omega) \\ &\leqslant (1 - e^{-B}) \omega' \circ d(x,y).\end{aligned}$$

If $\omega \circ d(x,y) \leqslant \omega(\operatorname{diam}\Omega)/K$, then $\omega' \circ d(x,y) = K\omega \circ d(x,y)$ and we bound the mass of Λ by $B\omega \circ d(x,y)$:

$$\begin{aligned}\int \omega' \circ d(x_t, y_t) \, \mathrm{d}\Lambda(\bar{x}, \bar{y}) &\leqslant B\omega \circ d(x,y) \omega'(\operatorname{diam}\Omega) \\ &\leqslant \frac{B\omega(\operatorname{diam}\Omega)}{K} \omega' \circ d(x,y).\end{aligned}$$

Choosing K large enough to ensure $\frac{B\omega(\operatorname{diam}\Omega)}{K} \leqslant 1 - e^{-B}$, we get in both cases

$$\int \omega' \circ d(x_t, y_t) \, \mathrm{d}\Pi(\bar{x}, \bar{y}) \leqslant DK \cdot F(t, \omega' \circ d(x,y)) + (1 - e^{-B}) \omega'(d(x,y)). \tag{3}$$

Step 4: $\mathrm{W}_\omega\left(\mathscr{L}^{*t}\mu, \mathscr{L}^{*t}\nu\right) \leqslant C\,\mathrm{W}_\omega(\mu,\nu) \quad \forall t \in \mathbb{N},\ \forall \mu,\nu \in \mathcal{P}(\Omega)$.

Since $F(t,r) \lesssim r$, the previous step implies in particular that Π, as a restricted coupling at time t, is ω'-Hölder; but $\omega \leqslant \omega' \leqslant K\omega$ on $[0, \operatorname{diam}\Omega]$ so that Π is also ω-Hölder. Then the claim follows from Lemma 2.13.

Step 5: There exist $\theta_1 \in (0,1)$ and $k_1 \in \mathbb{N}$ such that for all r, all $x, y \in \Omega$ such that $\omega' \circ d(x,y) \geqslant r$ and all $t \geqslant k_1\tau(r/2^{k_1})$,

$$\mathrm{W}_{\omega'}(\mathscr{L}^{*t}\delta_x, \mathscr{L}^{*t}\delta_y) \leqslant \theta_1 \omega' \circ d(x,y).$$

We choose any $\theta_1 \in (1-e^{-B}, 1)$ and k_1 large enough to ensure $DK/2^{k_1} + (1-e^{-B}) \leqslant \theta_1$, and then apply (3) (note that $k_1\tau(r/2^{k_1}) \geqslant \tau(r) + \tau(r/2) + \cdots + \tau(r/2^{k_1})$).

Step 6: There exist $\theta \in (0,1)$ and $k_2 \in \mathbb{N}$ such that for all r, all $\mu,\nu \in \mathcal{P}(\Omega)$ such that $\mathrm{W}_{\omega'}(\mu,\nu) = r$ and all $t \geqslant k_2\tau(r/k_2)$,

$$\mathrm{W}_{\omega'}(\mathscr{L}^{*t}\mu, \mathscr{L}^{*t}\nu) \leqslant \theta\,\mathrm{W}_{\omega'}(\mu,\nu).$$

Choose any $\theta \in (\theta_1, 1)$ and let $\eta > 0$ be small enough to ensure $\theta_1 + C\eta \leqslant \theta$, where C is the constant of Step 4. Let μ,ν be any two probability measure and let $\Pi \in \Gamma(\mu,\nu)$ be optimal for $\mathrm{W}_{\omega'}(\mu,\nu) =: r$. Define $s := \eta r$ and $E := \{(x,y) \mid \omega' \circ d(x,y) \geqslant s\}$. For all $t \geqslant k_1\tau(s/2^{k_1})$, using Lemma 2.9 it comes

$$\begin{aligned}
\mathrm{W}_{\omega'}(\mathscr{L}^{*t}\mu, \mathscr{L}^{*t}\nu) &\leqslant \int \mathrm{W}_{\omega'}(\mathscr{L}^{*t}\delta_x, \mathscr{L}^{*t}\delta_y)\,\mathrm{d}\Pi(x,y) \\
&\leqslant \int_E \mathrm{W}_{\omega'}(\mathscr{L}^{*t}\delta_x, \mathscr{L}^{*t}\delta_y)\,\mathrm{d}\Pi(x,y) + \int_{\Omega\times\Omega\setminus E} \mathrm{W}_{\omega'}(\mathscr{L}^{*t}\delta_x, \mathscr{L}^{*t}\delta_y)\,\mathrm{d}\Pi(x,y) \\
&\leqslant \theta_1 \int_E \omega' \circ d(x,y)\,\mathrm{d}\Pi(x,y) + C\int_{\Omega\times\Omega\setminus E} \omega' \circ d(x,y)\,\mathrm{d}\Pi(x,y) \\
&\leqslant \theta_1\,\mathrm{W}_{\omega'}(\mu,\nu) + C\eta r \\
&\leqslant \theta\,\mathrm{W}_{\omega'}(\mu,\nu).
\end{aligned}$$

It suffices to choose $k_2 \geqslant 2^{k_1}/\eta$.

Step 7: conclude.

We deduce that the θ decay time $\tau_\theta^{\omega'}(r)$ of $\mathscr{L}^*$ with respect to $\mathrm{W}_{\omega'}$ is at most $k_2\tau(r/k_2)$. Then for all $n \in \mathbb{N}$,

$$\tau_{\theta^n}^{\omega'}(r) \leqslant k_2\tau(r/k_2) + k_2\tau(\theta r/k_2) + \cdots + k\tau(\theta^{n-1}r/k_2)$$

and taking n large enough to ensure $\theta^n \leqslant \frac{1}{2K}$ we get

$$\tau_{\frac{1}{2K}}^{\omega'}(r) \leqslant k_2 n\tau(\theta^{n-1}r/k_2) \leqslant k\tau(r/k) \quad \text{for some } k.$$

Now since $\mathrm{W}_\omega \leqslant \mathrm{W}_{\omega'} \leqslant K\,\mathrm{W}_\omega$, the decay time for $\mathscr{L}^*$ with respect to W_ω satisfies $\tau_{\frac{1}{2}}^{\omega} \leqslant \tau_{\frac{1}{2K}}^{\omega'}$, and we are done. □

5 Weakly contracting 1-to-k transition kernels

This Section develops the tools we will need next to apply our general results of Sections 3 and 4 to the cases considered in the Introduction. We introduce a class of transition kernels that appear naturally as backward random walks for the maps of Theorems A-G; we construct natural couplings for this class of transition kernels, and give flatness criteria with respect to these natural couplings and moduli of continuity $\omega_{\alpha+\beta\log}$; to be able to apply these criteria we study the rate of decay of $c^n(r)$ as $n \to \infty$ for relevant families of weak contractions $c : [0, +\infty) \to [0, +\infty)$; and finally we state a half-general result encompassing the results of Sections 3 and 4 in a way that is easy to reuse.

5.1 A criterion for stable ω-decay

We start by giving a sufficient condition for certain couplings to have stable ω-decay rate, ensuring applicability of Theorem 4.1 without restriction on A. We assume that P is itself Markovian, i.e. is given by $(\Pi^t_{x,y})_{x,y} = (\pi_{x,y})_{x,y} \circ \cdots \circ (\pi_{x,y})_{x,y}$ for some $\pi_{x,y} \in \Gamma(m_x, m_y)$.

Definition 5.1. We say that a Markov transition kernel $(\pi_{x,y})_{x,y}$ on Ω^2 is *non-dilating and contracting with positive probability* when there exist $\lambda, p \in (0,1)$ such that for all $x, y \in \Omega$:

- for $\pi_{x,y}$-almost all (x_1, y_1), $d(x_1, y_1) \leqslant d(x, y)$,
- there exist a set $E_{x,y} \subset \Omega^2$ such that for all $(x_1, y_1) \in E$, $d(x_1, y_1) \leqslant \lambda d(x, y)$ and $\pi_{x,y}(E_{x,y}) \geqslant p$,

Lemma 5.2. *Let $(\pi_{x,y})_{x,y}$ be a Markovian coupling of M and A be a bounded potential normalized with respect to M. If $(\pi_{x,y})_{x,y}$ is non-dilating and contracting with positive probability, then $(e^{A(x_1)}\,\mathrm{d}\pi_{x,y}(x_1, y_1))_{x,y}$ is non-dilating and contracting with positive probability.*

Proof. Both points are essentially obvious: $e^{A(x_1)}\,\mathrm{d}\pi_{x,y}(x_1, y_1)$ is absolutely continuous with respect to $\pi_{x,y}$ and $\int_{E_{x,y}} e^{A(x_1)}\,\mathrm{d}\pi_{x,y} \geqslant p \min(e^A) > 0$. For the claim to make sense though, one has to observe that $(e^A \pi_{x,y})_{x,y}$ is a Markov transition kernel, which follows from the normalization hypothesis:

$$\int \mathbf{1} e^{A(x_1)}\,\mathrm{d}\pi_{x,y}(x_1, y_1) = \int \mathbf{1} e^{A(x_1)}\,\mathrm{d}m_x(x_1) = \mathbf{1}.$$

□

Lemma 5.3. *If $(\pi_{x,y})_{x,y}$ is a Markov transition kernel on Ω^2 which is non-dilating and contracting with positive probability, then for some $\lambda \in (0,1)$, all $(x, y) \in \Omega^2$ and all $t \in \mathbb{N}$,*

$$\int d(x_t, y_t)\,\mathrm{d}\pi^t_{x,y}(\bar{x}, \bar{y}) \leqslant \lambda^t d(x, y)$$

Here we denoted by $\pi^t_{x,y}$ the iterates of the Markov transition kernel, so that when $(\pi_{x,y})_{x,y}$ comes from the coupling P, $\pi^t_{x,y} = \Pi^t_{x,y}$.

Proof. It suffices to consider the case $t = 1$, then conclude by induction. But this follows from the definition: denoting by λ_1 the contraction factor of $\pi_{x,y}$ on $E_{x,y}$,

$$\begin{aligned}\int d(x_1, y_1)\, \mathrm{d}\pi_{x,y}(x_1, y_1) &= \int_{E_{x,y}} d(x_1, y_1)\, \mathrm{d}\pi_{x,y}(x_1, y_1) + \int_{\Omega^2 \setminus E_{x,y}} d(x_1, y_1)\, \mathrm{d}\pi_{x,y}(x_1, y_1) \\ &\leqslant p\lambda_1 d(x, y) + (1 - p)d(x, y)\end{aligned}$$

then we take $\lambda = p\lambda_1 + (1 - p)$. □

Corollary 5.4. *Assume* P *is Markovian with a transition kernel that is non-dilating and contracting with positive probability. Then for all bounded normalized potentials* A, P_A *has exponential decay with respect to* $\omega_{\alpha+\beta\log}$ *for all* $\alpha \in (0, 1)$ *and all* $\beta \in \mathbb{R}$*; and* P_A *has polynomial decay of degree* β *with respect to* $\omega_{\beta\log}$ *for all* $\beta > 0$.

Proof. Let A be a bounded, normalized potential. Then for all modulus of continuity ω, some $\lambda \in (0, 1)$, and all x, y, t:

$$\begin{aligned}\int \omega \circ d(x_t, y_t)\, \mathrm{d}\Big(e^{A^t}\Pi^t_{x,y}\Big)(\bar{x}, \bar{y}) &\leqslant \omega\Big(\int d(x_t, y_t)\, \mathrm{d}\Big(e^{A^t}\Pi^t_{x,y}\Big)(\bar{x}, \bar{y})\Big) \\ &\leqslant \omega(\lambda^t d(x, y)).\end{aligned}$$

There is only left to observe that for all $\alpha' < \alpha$, $\omega_{\alpha+\beta\log}(\lambda^t r) \lesssim \lambda^{\alpha' t}\omega_{\alpha+\beta\log}(r)$ and that

$$\omega_{\beta\log}(\lambda^t r) = \frac{\omega_{\beta\log}(r)}{\Big(1 + t \cdot \omega_{\beta\log}(r)^{\frac{1}{\beta}} \log\frac{1}{\lambda}\Big)^{\beta}}$$

which provides exactly the polynomial decay of degree β. □

5.2 Weakly contracting 1-to-k transition kernels

Let $\mathsf{M} = (m_x)_{x\in\Omega}$ be a transition kernel of the form $m_x = \sum_{j=1}^k \frac{1}{k}\delta_{x^j}$ where for each $x \in \Omega$, $B(x) = (x^j)_{1\leqslant j\leqslant k}$ is a family of k points. We will sometimes call M a 1-to-k transition kernel, but note that some x^j might coincide. Note that compared to the setup of [KLS15], this is just another way to encode a k-multiset.

Definition 5.5. Let us call *contraction function* any continuous map $c : [0, +\infty) \to [0, +\infty)$ such that $c(0) = 0$ and $c(r) < r$ for all $r > 0$. We say that M is *weakly contracting* if there are a contraction function c and a real number $\lambda > 1$ such that, for all $x, y \in \Omega$, there exist permutations η, σ of $\{1, \dots, k\}$ such that

i. for all $j \in \{1, \dots, k\}$ it holds $d(x^{\eta(j)}, y^{\sigma(j)}) \leqslant c(d(x, y))$,

ii. $d(x^{\eta(k)}, y^{\sigma(k)}) \leqslant d(x, y)/\lambda$.

Most usually, we will have $c(r) \sim r$ when $r \to 0$ so that the second contraction property will be the stronger one; it will often be satisfied for more j's than just $j = k$, but the above is sufficient for our purposes.

Definition 5.6. When M is weakly contracting, we define a *natural coupling* P as follows (it need not be unique given M, but is uniquely defined by the family of permutations η and σ above). For each $(x, y) \in \Omega \times \Omega$, we fix permutations η, σ realizing item *i* above (the dependency of η, σ on both x and y will be kept implicit); then for each pair (x, y) and each word $w = (j_1, \dots, j_t) \in \{1, \dots, k\}^t$ we let $\bar{x}_t^w, \bar{y}_t^w \in \Omega^t$ be the sequences $(x_1, \dots, x_t), (y_1, \dots, y_t)$ such that $x_1 = x^{\eta(j_1)} \in B(x)$ and $y_1 = y^{\sigma(j_1)} \in B(y)$, and for all n: $x_{n+1} = (x_n)^{\eta_n(j_n)} \in B(x_n)$ and $y_{n+1} = (y_n)^{\sigma_n(j_n)} \in B(y_n)$ where η_n and σ_n are the permutation associated to the pair (x_n, y_n). Then the natural coupling is

$$\Pi_{x,y}^t = \sum_{w \in \{1,\dots,k\}^t} \frac{1}{k^t} \delta_{(\bar{x}_t^w, \bar{y}_t^w)}.$$

In other words we pair together the orbits according to the pairing given in the definition of "weakly contracting".

The following lemma will be fed into Theorem 4.1, and we see that it is really the regularity of observables that drives the speed of decay.

Lemma 5.7. *Let M be a weakly contracting 1-to-k transition kernel and P be the natural coupling. For all bounded normalized potential A, P_A has exponential decay with respect to $\omega_{\alpha+\beta\log}$, for all $\alpha \in (0, 1)$ and all $\beta \in \mathbb{R}$, and P_A has polynomial decay of degree β with respect to $\omega_{\beta\log}$ for all $\beta > 0$.*

Proof. The natural coupling is Markovian, with transition kernel $\pi_{x,y} = \sum_j \frac{1}{k} \delta_{(x^{\eta(j)}, y^{\sigma(j)})}$, which is non-dilating and contracting with positive probability thanks to the hypothesis that M is weakly contracting. Corollary 5.4 provides the conclusion. □

Note that the above holds regardless of the contraction function (and would even hold with $c(r) = r$): a single contracting branch is sufficient to ensure the natural coupling of the transition kernel has good decay. By contrast, the specific features of c will prove important for flatness: for weak c our method will only apply to very regular potentials.

5.3 Flatness criteria

Lemma 5.8. *Let ω, $\tilde{\omega}$ be moduli of continuity such that $\tilde{\omega} \leqslant C\omega$ on $[0, 1]$, so that $\mathscr{C}^{\tilde{\omega}}(\Omega) \subset \mathscr{C}^{\omega}(\Omega)$.*

If M is weakly contracting with contraction function c and if there exist a constant C such that for all $r \in (0, \operatorname{diam} \Omega]$,

$$\sum_{n \geqslant 1} \tilde{\omega}(c^n(r)) \leqslant C\omega(r)$$

then every potential $A \in \mathscr{C}^{\tilde{\omega}}(\Omega)$ is flat with respect to P and ω.

Proof. Fix a potential $A \in \mathscr{C}^{\tilde{\omega}}(\Omega)$. $\Pi^t_{x,y}$ is supported on pairs of sequences $(\bar{x} = (x_1, \dots, x_k), \bar{y} = (y_1, \dots, y_k))$ such that for all $n \leqslant k$ it holds $d(x_n, y_n) \leqslant c^n(d(x,y))$. For all x, y, t and $\Pi^t_{x,y}$-almost all $\bar{x} = (x_1, \dots, x_t)$,$\bar{y} = (y_1, \dots, y_t)$ we thus have

$$\begin{aligned}\left|A^t(\bar{x}) - A^t(\bar{y})\right| &\leqslant \sum_{n=1}^{t} |A(x_n) - A(y_n)| \\ &\leqslant \mathrm{Hol}_{\tilde{\omega}}(A) \sum_{n=1}^{t} \tilde{\omega} \circ d(x_n, y_n) \\ &\leqslant \mathrm{Hol}_{\tilde{\omega}}(A) \sum_{n=1}^{t} \tilde{\omega}(c^n(d(x,y))) \\ &\lesssim \omega \circ d(x,y),\end{aligned}$$

which is the definition of flatness. □

In some circumstances, we will need a slightly finer analysis.

Lemma 5.9. *Let* M *be a weakly contracting* 1*-to-*k *transition kernel with contraction function* c *and ratio* $\lambda > 1$*. As above,* $x^1, \dots, x^k$ *denote the points supporting* m_x *and* η, σ *(the dependence on* (x,y) *being kept implicit) define the natural coupling* P*.*

Assume that there is a set $N \subset \Omega$ *such that for all* $x, y \in \Omega$ *and all* j*, if either* $x^{\eta(j)} \notin N$ *or* $y^{\sigma(j)} \notin N$*, then* $d(x^{\eta(j)}, y^{\sigma(j)}) \leqslant d(x,y)/\lambda$*. In other words, in the definition of "weakly contracting", the indices for which the stronger contraction does not hold are detected by the condition* $(x^{\eta(j)}, y^{\sigma(j)}) \in N^2$*.*

Let $\alpha \in (0,1]$ *and* $A \in \mathscr{C}^{\alpha}(\Omega)$*. If for some* C*, all* t*, all* $x, y \in \Omega$ *and* $\Pi^t_{x,y}$*-almost all* $(\bar{x}, \bar{y})$ *staying in* N *(i.e.* $(x_n, y_n) \in N^2$ *for all* $n \in \{1, \dots, t\}$*) we have*

$$\left| \sum_{n=1}^{t} A(x_n) - A(y_n) \right| \leqslant C d(x,y)^{\alpha}$$

then A *is* ω_α*-flat.*

Proof. Given t, x, y and $\Pi^t_{x,y}$ almost any $(\bar{x}, \bar{y})$, we decompose the time interval into "runs in N", i.e. write $\{1, \dots, t\} = I_1 \cup \dots \cup I_k$ where $I_j = \{n_j, \dots, n_{j+1} - 1\}$ are intervals of integers, with (n_j) is an increasing sequence and:

- $(x_{n+1}, y_{n+1}) \in N^2$ whenever n and $n+1$ both lie in some I_j,
- $d(x_{n_{j+1}}, y_{n_{j+1}}) \leqslant d(x_{n_{j+1}-1}, y_{n_{j+1}-1})/\lambda$ for all j.

We use the hypothesis and that distance decreases as n increases to get

$$\begin{aligned}\Big|\sum_{n=1}^{t} A(x_n) - A(y_n)\Big| &\leqslant \sum_{j=1}^{k} \Big|\sum_{n=n_j}^{n_{j+1}-1} A(x_n) - A(y_n)\Big| \\ &\leqslant \sum_{j=1}^{k} C d(x_{n_j-1}, y_{n_j-1})^\alpha \\ &\leqslant C \sum_{j=1}^{k} d(x,y)^\alpha / \lambda^{(j-1)\alpha} \\ &\leqslant C' d(x,y)^\alpha.\end{aligned}$$

□

5.4 Decay of iterated contraction functions

In order to use Lemma 5.8, we shall determine the speed of convergence to 0 of iterates of c in the cases of interest. Recall that c is assumed to be a contraction function: it is continuous, $c(0) = 0$ and $c(r) < r$ for all $r > 0$.

Lemma 5.10. *If for some $q \in (0,1)$ and $D > 0$ we have*

$$c(r) = (1 - Dr^q)r + o_{r\to 0}(r^{1+q})$$

then for some $a > 0$, all $r \in (0,1]$ and all $n \in \mathbb{N}$ we have

$$c^n(r) \leqslant \frac{1}{\left(an + \frac{1}{r^q}\right)^{\frac{1}{q}}}.$$

Proof. We have

$$\begin{aligned}\frac{1}{c(r)^q} &= \frac{1}{r^q}\big(1 - Dr^q + o(r^q)\big)^{-q} \\ &= \frac{1}{r^q}\big(1 + qDr^q + o(r^q)\big) \\ &= \frac{1}{r^q} + qD + o(1).\end{aligned}$$

For some $a > 0$ and $r_1 \in (0,1]$ and all $r \in (0, r_1)$, we thus have

$$\frac{1}{c(r)^q} \geqslant \frac{1}{r^q} + a. \tag{4}$$

Since $c(r) < r$ on $[r_1, 1]$ and c is continuous, Inequality (4) holds on $(0,1]$ up to taking a smaller a. Then by induction, for all n it holds

$$\frac{1}{c^n(r)^q} \geqslant \frac{1}{r^q} + an,$$

which is the claim. □

Lemma 5.11. *If for some $q \in (0, +\infty)$ we have*

$$c(r) = \Big(1 - \frac{1}{\big(\log \frac{1}{r}\big)^q}\Big) r + o_{r\to 0}\Big(\frac{r}{\big(\log \frac{1}{r}\big)^q}\Big),$$

then for some $a > 0$, all $r \in (0, 1]$ and all $n \in \mathbb{N}$ we have

$$c^n(r) \leqslant e^{-\left(an + (\log \frac{1}{r})^{q+1}\right)^{\frac{1}{q+1}}}.$$

Proof. We have for $r > 0$ and arbitrary $\alpha > 0$:

$$\begin{aligned}
\log \frac{1}{c(r)} &= \log \frac{1}{r} - \log\Big(1 - \frac{1}{\big(\log \frac{1}{r}\big)^q} + o\Big(\frac{1}{\big(\log \frac{1}{r}\big)^q}\Big)\Big) \\
&= \log \frac{1}{r} + \frac{1}{\big(\log \frac{1}{r}\big)^q} + o\Big(\frac{1}{\big(\log \frac{1}{r}\big)^q}\Big) \\
\Big(\log \frac{1}{c(r)}\Big)^\alpha &= \Big(\log \frac{1}{r}\Big)^\alpha \Big(1 + \frac{1}{\big(\log \frac{1}{r}\big)^{q+1}} + o\Big(\frac{1}{\big(\log \frac{1}{r}\big)^{q+1}}\Big)\Big)^\alpha \\
&= \Big(\log \frac{1}{r}\Big)^\alpha + \alpha\Big(\log \frac{1}{r}\Big)^{\alpha-(q+1)} + o\Big(\Big(\log \frac{1}{r}\Big)^{\alpha-(q+1)}\Big).
\end{aligned}$$

Taking $\alpha = q + 1$ it comes $\Big(\log \frac{1}{c(r)}\Big)^{q+1} = \Big(\log \frac{1}{r}\Big)^{q+1} + (q+1) + o(1)$ so that for some $a, r_0 > 0$ and all $r \in (0, r_0)$ we get

$$\Big(\log \frac{1}{c(r)}\Big)^{q+1} \geqslant \Big(\log \frac{1}{r}\Big)^{q+1} + a.$$

Up to adjusting a, the same also holds true on $[r_0, 1]$ since c is continuous and $c(r) < r$ for positive r. It follows that for all $r \in (0, 1]$ and all $n \in \mathbb{N}$:

$$\Big(\log \frac{1}{c^n(r)}\Big)^{q+1} \geqslant \Big(\log \frac{1}{r}\Big)^{q+1} + an$$

and the result follows. □

5.5 A half-general result

Now we specialize the general results above to the case of k-to-1 maps, getting closer to Theorems A-G.

Theorem 5.12. *Let $T : \Omega \to \Omega$ be a Lipschitz k-to-1 map of a compact metric space Ω, such that the uniform backward random walk $\mathsf{M} = (m_x)_{x\in\Omega}$ defined by $m_x = \frac{1}{k}\sum_{y\in T^{-1}(x)} \delta_y$ is weakly contracting. Let $A \in \mathscr{C}^{\alpha+\beta\log}(\Omega)$ be a flat potential with respect to some modulus $\omega_{\alpha+\beta\log}$ and to the natural coupling P.*

Then the transfer operator $\mathscr{L}_A = \mathscr{L}_{T,A}$ has a positive eigenfunction h_A in $\mathscr{C}^{\alpha+\beta\log}(\Omega)$ with a positive eigenvalue λ_A; its dual $\mathscr{L}_A^*$ has an eigenprobability ν_A for the same eigenvalue and $d\mu_A := h_A \,\mathrm{d}\nu_A$ defines a T-invariant probability (the RPF measure of A).

If $\alpha > 0$, then moreover $\mathscr{L}_A$ has spectral gap in $\mathscr{C}^{\alpha+\beta\log}$: for some $C > 0$, $\delta \in (0,1)$ and for all observable $f \in \mathscr{C}^{\alpha+\beta\log}(\Omega)$ such that $\int f \,\mathrm{d}\mu_A = 0$ we have

$$\|\mathscr{L}_A^t f\|_{\alpha+\beta\log} \leqslant C(1-\delta)^t \lambda_A^t \|f\|_{\alpha+\beta\log}.$$

In this case, we thus have exponential decay of correlations: for some C and all f, g it holds

$$\left| \int f \cdot g \circ T^t \,\mathrm{d}\mu_A - \int f \,\mathrm{d}\mu_A \int g \,\mathrm{d}\mu_A \right| \leqslant C(1-\delta)^t \operatorname{Hol}_{\alpha+\beta\log}(f) \|g\|_{L^1(\mu_A)}.$$

Otherwise, i.e. if $\alpha = 0$ and $\beta > 0$, $\mathscr{L}_{T,A}$ decays in the uniform norm as a degree β polynomial: for some $C > 0$ and all observable $f \in \mathscr{C}^{\beta\log}(\Omega)$ such that $\int f \,\mathrm{d}\mu_A = 0$ we have

$$\|\mathscr{L}_A^t f\|_\infty \leqslant C \frac{\|f\|_{\beta\log}}{t^\beta} \lambda_A^t.$$

In that case, we thus have polynomial decay of correlation of degree β: for some C and all f, g it holds

$$\left| \int f \cdot g \circ T^t \,\mathrm{d}\mu_A - \int f \,\mathrm{d}\mu_A \int g \,\mathrm{d}\mu_A \right| \leqslant C \frac{\operatorname{Hol}_{\beta\log}(f) \|g\|_{L^1(\mu_A)}}{t^\beta}.$$

Proof. Since M is weakly contracting, the natural coupling is ω-Hölder for all ω. Since T is Lipschitz, it preserves $\mathscr{C}^{\alpha+\beta\log}(\Omega)$ by composition, and we can apply Corollary 3.3: there is a positive eigenvalue λ_A and a positive eigenfunction $h_A \in \mathscr{C}^{\alpha+\beta\log}(\Omega)$, and the normalized potential $\tilde{A} = A + \log h_A - \log h_A \circ T - \log \lambda_A$ is flat and in $\mathscr{C}^{\alpha+\beta\log}(\Omega)$.

By Lemma 5.7, the natural coupling P is $\omega_{\alpha+\beta\log}$-decaying with a rate which is either exponential (if $\alpha > 0$) or polynomial of degree β (if $\alpha = 0$ and $\beta > 0$). Theorem 4.1 then shows that $\mathscr{L}_{\tilde{A}}^*$ has at least decay rate F which is either exponential (if $\alpha > 0$) or polynomial of degree β (if $\alpha = 0$ and $\beta > 0$), with respect to the Wasserstein metric $\mathrm{W}_{\alpha+\beta\log}$. In particular it has a unique fixed probability $\nu_{\tilde{A}}$. It follows that $\mathscr{L}_A^*$ has a unique eigenprobability $\mathrm{d}\nu_A = \frac{1}{h_A} \mathrm{d}\nu_{\tilde{A}}$, and $\mu_A = \nu_{\tilde{A}}$ is T-invariant (see Section 2.1).

Now we apply Proposition 2.8 to $\mathscr{L}_{\tilde{A}}$. First, if $\alpha > 0$ we can take $F(t,r) = C(1-\delta)^t r$ for some $C > 0$ and $\delta \in (0,1)$ and we get

$$\|\mathscr{L}_{\tilde{A}}^t f - \mu_A(f)\|_{\alpha+\beta\log} \leqslant C \operatorname{Hol}_{\alpha+\beta\log}(f)(1-\delta)^t$$

where C depends notably on $\operatorname{diam}\Omega$. Since $\mathscr{L}_A^t f = \lambda_A^t h_A \mathscr{L}_{\tilde{A}}^t (h_A^{-1} f)$, when $\mu_A(f) = 0$ it comes

$$\|\mathscr{L}_A^t f\|_{\alpha+\beta\log} \leqslant C\lambda_A^t \operatorname{Hol}_{\alpha+\beta\log}(h_A^{-1} f)(1-\delta)^t \leqslant C'(1-\delta)^t \lambda_A^t \|f\|_{\alpha+\beta\log}.$$

Note that we have to switch from the Hölder constant of $h_A^{-1} f$ to the Hölder *norm* of f since the norm is multiplicative, but not the constant.

Second, if $\alpha = 0$ and $\beta > 0$, we can take $F(t,r) = \frac{Br}{\left(tr^{\frac{1}{\beta}}+b\right)^{\beta}}$ and we get

$$\|\mathscr{L}_{\tilde{A}}^t f - \mu_A(f)\|_\infty \leqslant \mathrm{Hol}_{\beta\log}(f)\frac{C}{t^\beta},$$

from which we deduce as above $\|\mathscr{L}_A^t f\|_\infty \leqslant C\frac{\|f\|_{\beta\log}}{t^\beta}\lambda_A^t$ when $\mu_A(f) = 0$.

In both case, the decay of correlations follows in the classical way: one observes that $\mathscr{L}_{\tilde{A}}^t(f \cdot g \circ T^t) = \mathscr{L}_{\tilde{A}}^t(f) \cdot g$, assumes $\mu_A(f) = 0$ by adding a constant, and then writes

$$\begin{aligned}\left|\int f \cdot g \circ T^t \,\mathrm{d}\mu_A\right| &= \left|\int f \cdot g \circ T^t \,\mathrm{d}\left(\mathscr{L}_{\tilde{A}}^{*t}\mu_A\right)\right| \\ &= \left|\mathscr{L}_{\tilde{A}}^t(f \cdot g \circ T^t)\,\mathrm{d}\mu_A\right| \\ &= \left|\mathscr{L}_{\tilde{A}}^t(f) \cdot g\,\mathrm{d}\mu_A\right| \\ &\leqslant \|\mathscr{L}_{\tilde{A}}^t f\|_\infty \int |g|\,\mathrm{d}\mu_A,\end{aligned}$$

ending the proof. □

In the case $\alpha = 0$, we could obtain decay in an intermediate generalized Hölder norm (between the uniform norm and $\|\cdot\|_{\beta\log}$) by using the full extent of Proposition 2.8.

Remark 5.13. Theorem 5.12 applies only when the number of inverse images is constant ("full branches"), and a natural question is whether the method applies more generally.

Since we rely on spaces of continuous functions, one cannot hope to deal with the traditional transfer operator (corresponding to the uniform backward random walk) when branches are not full. However, Section 4 applies as soon as one finds a backward random walk that does preserve continuity and contracts appropriately. In order to preserve continuity, one needs to jump only with small probability to a point located near the endpoint of its branch; it should be possible to carry this out in some non-Markovian examples.

6 Application to circle maps

6.1 The Pomeau-Manneville family

In this section we finish the proof of Theorems A and G by considering the Pomeau-Manneville family: given an exponent $q > 0$, we consider on the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ parametrized by $[0,1)$ the map

$$\begin{aligned}T_q : \mathbb{T} &\to \mathbb{T} \\ x &\mapsto \begin{cases}\left(1+(2x)^q\right)x & \text{if } x \in [0,\frac{1}{2}] \\ 2x-1 & \text{if } x \in [\frac{1}{2},1)\end{cases}\end{aligned}$$

and we let M be the corresponding 1-to-k backward random walk.

6.1.1 Inverse branches and weak contraction

A point $y \in [0,1)$ has two inverse images, denoted by $b_1(y)$ and $b_2(y) = (y+1)/2$ (the maps b_1 and b_2 are called *branches* of T_q). We do not search for an explicit expression of $b_1(y)$, but we observe that

- b_1 is concave,
- $b_1(x) \geqslant x/2$, and
- $b_1(y) = \big(1-(2y)^q\big)y + O(y^{2q+1})$.

Turning back to the notation of Section 5, to prove that M is weakly contracting we consider $x, y \in \mathbb{T}$ and their inverse images $x^1 = b_1(x), x^2 = (x+1)/2$ and $y^1 = b_1(y), y^2 = (y+1)/2$.

When a shortest path from x to y avoids 0, i.e. when $d(x,y) = |x-y|$, we pair x^1 with y^1 and x^2 with y^2 (i.e. $\sigma = \eta = \mathrm{Id}$). On the one hand we have $d(x^2,y^2) = d(x,y)/2$; on the other hand, by concavity $d(x^1,y^1) \leqslant b_1(d(x,y))$.

When the unique shortest path from x to y goes through 0, we pair x^1 with y^2 and x^2 with y^1. Up to swapping x and y we assume that y is closer to 1 and x closer to 0, so that $d(x,y) = x+1-y$ and $y \geqslant 1/2$. Then on the one hand, since $y \geqslant 1/2$ the derivative of T_q' is greater than 2 on $[y, \frac{1}{2}]$ and we have $d(x^2,y^1) \leqslant d(x,y)/2$. On the other hand, setting $d = d(x,y)$ we have

$$d(x^1,y^2) = b^1(x) + \frac{1-y}{2} = \frac{d+x}{2} - 2^q x^{q+1} + O(x^{2q+1}).$$

The $O(x^{2q+1})$ is also an $O(d^{2q+1})$, and d being fixed the quantity $\frac{d+x}{2} - 2^q x^{q+1}$ is increasing in x when $d \leqslant x_0 := 2^{-\frac{q+1}{q}}(q+1)^{-\frac{1}{q}}$. We deduce

$$d(x^1,y^2) \leqslant (1-(2d)^q)d + O(d^{2q+1}).$$

This proves that M is weakly contracting, with coefficient $\lambda = 1/2$ and function $c(r) = (1-(2r)^q)r + O(r^{2q+1})$.

6.1.2 Case $q < 1$.

We start with the case $q < 1$ of Theorem A. Let thus $\alpha, \gamma \in (0,1)$ be such that $\gamma - \alpha \geqslant q$ and let $A \in \mathscr{C}^\gamma(\mathbb{T})$. Since $\gamma > \alpha$, we have $\omega_\gamma \leqslant C\omega_\alpha$ and in particular $A \in \mathscr{C}^\alpha(\mathbb{T})$. By Lemma 5.10 for some $a > 0$ and all $r > 0$ we have $c^n(r) \leqslant \big(an + r^{-q}\big)^{-\frac{1}{q}}$, so that (using $\gamma > q$):

$$\begin{aligned}
\sum_{n\geqslant 1} \omega_\gamma(c^n(r)) &\leqslant \sum_{n\geqslant 1} \big(an + r^{-q}\big)^{-\frac{\gamma}{q}} \\
&\leqslant \int_0^\infty \big(ax + r^{-q}\big)^{-\frac{\gamma}{q}}\,\mathrm{d}x \\
&\lesssim \int_{r^{-q}}^\infty y^{-\frac{\gamma}{q}}\,\mathrm{d}y \\
&\lesssim (r^{-q})^{1-\frac{\gamma}{q}} = r^{\gamma-q} \leqslant r^\alpha.
\end{aligned}$$

By Lemma 5.8, we deduce that A is flat with respect to ω_α. Then Theorem 5.12 applies, showing that $\mathscr{L}_A$ satisfies a Ruelle-Perron-Frobenius Theorem with a spectral gap, and exponential decay of correlations.

For Theorem G, we consider $\beta > 1$ and $A \in \mathscr{C}^{q+\beta\log}(\mathbb{T})$ and have to prove that A is $\omega_{(\beta-1)\log}$-flat; then Theorem 5.12 shows the Ruelle-Perron-Frobenius Theorem for $\mathscr{L}_{T_q,A}$ and degree $\beta-1$ polynomial decay of correlations. Flatness follows again from Lemma 5.8:

$$\begin{aligned}
\sum_{n\geqslant 1} \omega_{q+\beta\log}(c^n(r)) &\leqslant \sum_{n\geqslant 1} \frac{1}{(an+r^{-q})\Big(\log\big(r_0(an+r^{-q})^{\frac{1}{q}}\big)\Big)^\beta} \\
&\leqslant \int_0^\infty \frac{1}{(ax+r^{-q})\Big(\log\big(r_0(ax+r^{-q})^{\frac{1}{q}}\big)\Big)^\beta}\,\mathrm{d}x \\
&\lesssim \int_{r^{-q}}^\infty \frac{1}{y(\log(r_0 y))^\beta}\,\mathrm{d}y \\
&\lesssim \frac{1}{\big(\log(r_0 r^{-q})\big)^{\beta-1}} \\
&\lesssim \frac{1}{|\log(r)|^{\beta-1}}.
\end{aligned}$$

6.1.3 Case $q \geqslant 1$.

The case $q \geqslant 1$ of Theorem A is dealt with in a similar way. Let α, γ be such that $\gamma > q$ and $\gamma + 1 - \alpha \geqslant q$ and let $A \in \mathscr{C}^\alpha(\mathbb{T})$ be differentiable near 0 with $A'(r) = O_{r\to 0}(r^\gamma)$. We will be done if we can apply Theorem 5.12, for which we have left to check that A is ω_α-flat. With Lemma 5.9 in mind, we observe that there is a neighborhood N of 0 where $A'(r) \leqslant Cr^\alpha$ and outside which T_q is uniformly expanding, so that for some $\lambda > 1$ and all x, y, if x_1, y_1 are inverse images of x, y which are paired by the natural coupling, either $d(x_1, y_1) \leqslant d(x, y)/\lambda$ or $x_1, y_1 \in N$ and $d(x_1, y_1) \leqslant c(d(x, y))$. Lemma 5.9 shows that it suffices to check flatness for pairs of backward orbits $\bar{x}, \bar{y}$ such that $(x_n, y_n) \in N^2$ for all $n \in \{1, \dots, t\}$. Setting $r = d(x_1, y_1)$ we then have $d(x_n, y_n) \leqslant (an + r^{-q})^{-\frac{1}{q}}$ and $|x_n|, |y_n| \leqslant Cn^{-\frac{1}{q}}$ for some $a, C > 0$ and all n (Lemma 5.10). It follows from the Mean Value Theorem that

$$\begin{aligned}
\Big|\sum_{n=1}^t A(x_n) - A(y_n)\Big| &\lesssim \sum_{n=1}^t n^{-\frac{\gamma}{q}}(an+r^{-q})^{-\frac{1}{q}} \\
&\lesssim \sum_{an\leqslant r^{-q}} n^{-\frac{\gamma}{q}}(r^{-q})^{-\frac{1}{q}} + \sum_{an>r^{-q}} n^{-\frac{\gamma}{q}}(an)^{-\frac{1}{q}} \\
&\lesssim \sum_{an\leqslant r^{-q}} n^{-\frac{\gamma}{q}} r + \sum_{an>r^{-q}} n^{-\frac{\gamma+1}{q}} \\
&\lesssim r + r^{\gamma+1-q} \qquad (\text{since } \gamma > q) \\
&\lesssim d(x,y)^\alpha,
\end{aligned}$$

finishing the proof of Theorem A.

6.2 Non-uniformly expanding maps with a weak neutral point

Let us now prove Theorem B. We consider $q > 0$ and the map

$$T_{q\log} : \mathbb{T} \to \mathbb{T}$$
$$x \mapsto \begin{cases} 0 & \text{if } x = 0 \\ \big(1 + (1 - \log 2x)^{-q}\big)x & \text{if } x \in (0, \frac{1}{2}] \\ 2x - 1 & \text{if } x \in [\frac{1}{2}, 1) \end{cases}$$

whose uniform backward walk M is weakly contracting, with contraction function of the form

$$c(r) = \Big(1 - \frac{1}{\big(\log \frac{1}{r}\big)^q}\Big)r + o_{r\to 0}\Big(\frac{r}{\big(\log \frac{1}{r}\big)^q}\Big).$$

Lemma 5.11 tells us that for some $a > 0$,

$$c^n(r) \leqslant e^{-\left(an + (\log \frac{1}{r})^{q+1}\right)^{\frac{1}{q+1}}}.$$

Given any $\alpha \in (0, 1)$, for all $r > 0$ we thus have:

$$\begin{aligned} \sum_{n\geqslant 1} \omega_{\alpha+q\log}(c^n(r)) &\leqslant \sum_{n\geqslant 1} \frac{e^{-\alpha\left(an+(\log \frac{1}{r})^{q+1}\right)^{\frac{1}{q+1}}}}{\left(an + (\log \frac{1}{r})^{q+1}\right)^{\frac{q}{q+1}}} \\ &\leqslant \int_0^\infty \frac{e^{-\alpha\left(ax+(\log \frac{1}{r})^{q+1}\right)^{\frac{1}{q+1}}}}{\left(ax + (\log \frac{1}{r})^{q+1}\right)^{\frac{q}{q+1}}}\,\mathrm{d}x \\ &\lesssim \int_{(\log \frac{1}{r})^{q+1}}^\infty \frac{e^{-\alpha y^{\frac{1}{q+1}}}}{y^{\frac{q}{q+1}}}\,\mathrm{d}y \\ &\lesssim e^{-\alpha\left((\log \frac{1}{r})^{q+1}\right)^{\frac{1}{q+1}}} = r^\alpha, \end{aligned}$$

so that Lemma 5.8 ensures that every potential $A \in \mathscr{C}^{\alpha+q\log}(\mathbb{T})$ is flat with respect to ω_α and the natural coupling P. Then Theorem 5.12 can be applied and Theorem B follows.

6.3 General non-uniformly expanding maps

We turn to the proof of Theorem C. We consider $\alpha \in (0, 1]$ and a k-to-1 map $T : \mathbb{T} \to \mathbb{T}$ with a neutral fixed point at 0 and such that for all neighborhood N of 0, there is some $\lambda = \lambda(N) > 1$ such that $|T'| \geqslant \lambda$ outside N. Taking N small enough we can assume that for some λ and all $x, y \in \mathbb{T}$, given the permutations σ, η defining a pairing of

their inverse images $x^1, \dots, x^k, y^1, \dots, y^k$ to nearest neighbors, we have $d(x^{\sigma(j)}, y^{\eta(j)}) \leqslant d(x,y)/y$ except possibly for one pair $(x^{\sigma(1)}, y^{\eta(1)}) \in N^2$ (this can be considered a suitable interpretation of the hypothesis that T is non-uniformly expanding in order to include non-derivable maps).

It follows at once from Lemma 5.9 that every potential $A \in \mathscr{C}^\alpha(\mathbb{T})$ which is constant on a neighborhood N of 0 is ω_α-flat, and Theorem 5.12 ensures that $\mathscr{L}_A$ satisfies a Ruelle-Perron-Frobenius theorem with a spectral gap.

The fact that potentials which are constant in some neighborhood of 0 are dense in $\mathscr{C}^\alpha(\mathbb{T})$ with respect to $\|\cdot\|_\gamma$ for all $\gamma \in (0, \alpha)$ is a classical fact, but let us give a proof nonetheless.

Proof of denseness of V. For any $B \in \mathscr{C}^\alpha(\mathbb{T})$ and any $\varepsilon > 0$, let A be the potential such that $A(x) = B(0)$ for all x at distance at most $\varepsilon/2$ from 0, $A(x) = B(x)$ for all x at least ε away from 0, and which is affine on both $[\varepsilon/2, \varepsilon]$ and $[1-\varepsilon, 1-\varepsilon/2]$. Then A is α-Hölder with $\mathrm{Hol}_\alpha(A) \lesssim \mathrm{Hol}_\alpha(B)$, and for all $x \in \mathbb{T}$ we moreover have

$$|A(x) - B(x)| \lesssim \mathrm{Hol}_\alpha(B)\varepsilon^{\alpha-\gamma} d(x,0)^\gamma.$$

This already shows that A and B are close in the uniform norm. Let $x, y \in \mathbb{T}$ and first assume $d(x,y) \geqslant \varepsilon$. Then either A and B coincide at both x and y, or they coincide at one of them (y, say) and $d(x,0) \leqslant \varepsilon \leqslant d(x,y)$; then

$$|(A-B)(x) - (A-B)(y)| \lesssim \mathrm{Hol}_\alpha(B)\varepsilon^{\alpha-\gamma} d(x,0)^\gamma \lesssim \varepsilon^{\alpha-\gamma} d(x,y)^\gamma.$$

In the case when $d(x,y) \leqslant \varepsilon$ we can estimate directly

$$|(A-B)(x) - (A-B)(y)| \lesssim \mathrm{Hol}_\alpha(B) d(x,y)^\alpha \lesssim \varepsilon^{\alpha-\gamma} d(x,y)^\gamma.$$

We deduce that $\|A - B\|_\gamma \lesssim \varepsilon^{\alpha-\gamma}$, concluding the proof. □

6.4 Uniformly expanding maps

We conclude by the proof of Theorem E and Corollary F. Assume that $T : \Omega \to \Omega$ is a k-to-1 uniformly expanding map, by which we means that the uniform backward random walk $\mathsf{M} = (m_x)_{x \in \Omega}$ defined by

$$m_x = \sum_{y \in T^{-1}(x)} \frac{1}{k} \delta_y$$

is λ contracting for some $\lambda > 1$ (which in our terminology is the same as being weakly contracting with contraction function $c(r) = r/\lambda$). This definition in particular includes the classical definition, i.e. the case when Ω is a manifold and $\|D_x T(v)\| \geqslant \lambda \|v\|$ for all $x \in \Omega$ and all $v \in T_x\Omega$ (as briefly explained in [KLS15], Example 2.9).

Consider a parameter $\beta > 1$ and a potential $A \in \mathscr{C}^{\beta \log}(\Omega)$. We will again apply Lemma 5.8, observing that

$$\begin{aligned}
\sum_{n\geqslant 1} \omega_{\beta\log}(c^n(r)) &= \sum_{n\geqslant 1} \frac{1}{\left(\log(\frac{r_0}{r}\lambda^n)\right)^\beta} \\
&\leqslant \int_0^\infty \frac{1}{\left(x\log\lambda + \log\frac{r_0}{r}\right)^\beta}\,\mathrm{d}x \\
&\lesssim \int_{\log\frac{r_0}{r}} y^{-\beta}\,\mathrm{d}y \\
&\lesssim \left(\log\frac{1}{r}\right)^{1-\beta} \\
&\lesssim \omega_{(\beta-1)\log}(r);
\end{aligned}$$

it follows that A is $\omega_{(\beta-1)\log}$-flat. Theorem 5.12 then implies Theorem E.

Now, assume further that Ω is a connected Riemannian manifold and that T is $\mathscr{C}^1$. If the Jacobian JT is of regularity $\mathscr{C}^{\beta\log}$ for some $\beta > 1$, then the "natural" potential $A = -\log JT$ can be applied Theorem E. But this potential is named natural for the reason that if ν denotes the Riemannian volume (normalized to have total mass 1), then $\mathscr{L}_A^*(\nu) = \nu$; from Theorem E we know that there is a positive eigenfunction $f \in \mathscr{C}^{(\beta-1)\log}(\mathbb{T})$, and $f\,\mathrm{d}\nu$ is T-invariant (it is the eigenprobability for $\mathscr{L}_{\tilde{A}}$ where $\tilde{A}$ is the normalized potential). The polynomial decay of correlations also follows from Theorem E, ending the proof of Corollary F.

References

[BT08] Henk Bruin and Mike Todd. Equilibrium states for interval maps: potentials with $\sup\phi - \inf\phi < h_{\mathrm{top}}(f)$. *Comm. Math. Phys.*, 283(3):579–611, 2008. 1.3

[CS09] Van Cyr and Omri Sarig. Spectral gap and transience for Ruelle operators on countable Markov shifts. *Comm. Math. Phys.*, 292(3):637–666, 2009. 1.3

[CV13] A. Castro and P. Varandas. Equilibrium states for non-uniformly expanding maps: decay of correlations and strong stability. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 30(2):225–249, 2013. 1.3

[DU91] Manfred Denker and Mariusz Urbański. On the existence of conformal measures. *Trans. Amer. Math. Soc.*, 328(2):563–587, 1991. 1, 3, 3

[FJ01a] Aihua Fan and Yunping Jiang. On Ruelle-Perron-Frobenius operators. I. Ruelle theorem. *Comm. Math. Phys.*, 223(1):125–141, 2001. 1.2, 1.2

[FJ01b] Aihua Fan and Yunping Jiang. On Ruelle-Perron-Frobenius operators. II. Convergence speeds. *Comm. Math. Phys.*, 223(1):143–159, 2001. 1.2

[GKLMF15] Paolo Giulietti, Benoît R. Kloeckner, Artur O. Lopes, and Diego Marcon Farias. The calculus of thermodynamical formalism. arXiv:1508.01297, to appear in *J. Eur. Math. Soc.*, 2015. 1.1, 1.4, 2.1, 2.3

[Gou04] Sébastien Gouëzel. Sharp polynomial estimates for the decay of correlations. *Israel Journal of Mathematics*, 139(1):29–65, 2004. 1.3

[Gou05] Sébastien Gouëzel. Berry-Esseen theorem and local limit theorem for non uniformly expanding maps. *Ann. Inst. H. Poincaré Probab. Statist.*, 41(6):997–1024, 2005. 1.3

[Gui15] Pierre-Antoine Guihéneuf. Dynamical properties of spatial discretizations of a generic homeomorphism. *Ergodic Theory Dynam. Systems*, 35(5):1474–1523, 2015. 1.3

[HK82] Franz Hofbauer and Gerhard Keller. Equilibrium states for piecewise monotonic transformations. *Ergodic Theory Dynam. Systems*, 2(1):23–43, 1982. 1.3

[Hol05] Mark Holland. Slowly mixing systems and intermittency maps. *Ergodic Theory and Dynamical Systems*, 25(1):133–159, 2005. 1.1, 1.3

[Hu04] Huyi Hu. Decay of correlations for piecewise smooth maps with indifferent fixed points. *Ergodic Theory Dynam. Systems*, 24(2):495–524, 2004. 1.3

[Klo17] Benoît R. Kloeckner. Effective high-temperature estimates for intermittent maps. arXiv:1704.00586, 2017. 1.3, 2.1

[KLS15] Benoît R. Kloeckner, Artur O. Lopes, and Manuel Stadlbauer. Contraction in the Wasserstein metric for some Markov chains, and applications to the dynamics of expanding maps. *Nonlinearity*, 28(11):4117–4137, 2015. 1, 1.1, 2.5, 3, 4, 5.2, 6.4

[LRL14a] Huaibin Li and Juan Rivera-Letelier. Equilibrium states of interval maps for hyperbolic potentials. *Nonlinearity*, 27(8):1779–1804, 2014. 1.1

[LRL14b] Huaibin Li and Juan Rivera-Letelier. Equilibrium states of weakly hyperbolic one-dimensional maps for Hölder potentials. *Comm. Math. Phys.*, 328(1):397–419, 2014. 1.1

[LSV98] Carlangelo Liverani, Benoit Saussol, and Sandro Vaienti. Conformal measure and decay of correlation for covering weighted systems. *Ergodic Theory Dynam. Systems*, 18(6):1399–1420, 1998. 1.3

[LSV99] Carlangelo Liverani, Benoît Saussol, and Sandro Vaienti. A probabilistic approach to intermittency. *Ergodic Theory Dynam. Systems*, 19(3):671–685, 1999. 1.3

[Lyn06] Vincent Lynch. Decay of correlations for non-Hölder observables. *Discrete Contin. Dyn. Syst.*, 16(1):19–46, 2006. 1.3

[MN05] Ian Melbourne and Matthew Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.*, 260(1):131–146, 2005. 1.3

[PM80] Yves Pomeau and Paul Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Comm. Math. Phys.*, 74(2):189–197, 1980. 1.3

[PS92] Thomas Prellberg and Joseph Slawny. Maps of intervals with indifferent fixed points: thermodynamic formalism and phase transitions. *J. Statist. Phys.*, 66(1-2):503–514, 1992. 1.3

[Sar02] Omri Sarig. Subexponential decay of correlations. *Invent. Math.*, 150(3):629–653, 2002. 1.3

[TK05] Marta Tyran-Kamińska. An invariance principle for maps with polynomial decay of correlations. *Comm. Math. Phys.*, 260(1):1–15, 2005. 1.2

[Wal82] P. Walters. *An Introduction to Ergodic Theory*, volume 79 of *Graduate Texts in Mathematics*. Springer Verlag, 1982. 2.3

[You98] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998. 1.3

[You99] Lai-Sang Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999. 1.3