ENTROPY DENSITY AND LARGE DEVIATION PRINCIPLES WITHOUT UPPER SEMI-CONTINUITY OF ENTROPY

ZHIQIANG LI AND XIANGHUI SHI

Abstract. Expanding Thurston maps were introduced by M. Bonk and D. Meyer with motivation from complex dynamics and Cannon's conjecture from geometric group theory via Sullivan's dictionary. In this paper, we show that the entropy map of an expanding Thurston map is upper semi-continuous if and only if the map has no periodic critical points. For all expanding Thurston maps, even in the presence of periodic critical points, we show that ergodic measures are entropy-dense and establish level-2 large deviation principles for the distributions of Birkhoff averages, periodic points, and iterated preimages. It follows that iterated preimages and periodic points are equidistributed with respect to the unique equilibrium state for an expanding Thurston map and a potential that is Hölder continuous with respect to a visual metric on S^2. In particular, our results answer two questions posed in [Li15] and generalize the corresponding results there.

One of the main tools used in this paper is called the subsystems of expanding Thurston maps, which were inspired by a translation of the concept of subgroups from geometric group theory via Sullivan's dictionary, and have been investigated in [LSZ25, LS24].

Contents

2020 *Mathematics Subject Classification.* Primary: 37F10; Secondary: 37A50, 37A35, 37D35, 37F15, 37B99, 57M12.

Key words and phrases. expanding Thurston map, postcritically-finite map, large deviation principles, entropy density, upper semi-continuity, subsystems, thermodynamic formalism.

Z. Li and X. Shi were partially supported by NSFC Nos. 12101017, 12090010, 12090015, and BJNSF No. 1214021.

1. Introduction

A Thurston map is a (non-homeomorphic) branched covering map on a topological 2-sphere S^2 that is postcritically-finite, meaning that each of its critical points has a finite orbit under iteration. The most important examples are given by postcritically-finite rational maps on the Riemann sphere $\widehat{\mathbb{C}}$. While Thurston maps are purely topological objects, a deep theorem due to W.P. Thurston characterizes Thurston maps that are, in a suitable sense, described in the language of topology and combinatorics, equivalent to postcritically-finite rational maps (see [DH93]). This suggests that for the relevant rational maps, an explicit analytic expression is not so important, but rather a geometric-combinatorial description.

In the early 1980s, D.P. Sullivan introduced a "dictionary" that is now known as *Sullivan's dictionary*, which connects two branches of conformal dynamics: the iteration theory of rational maps and the actions of Kleinian groups. Under Sullivan's dictionary, the counterpart to Thurston's theorem in geometric group theory is Cannon's Conjecture [Can94]. Inspired by Sullivan's dictionary and their interest in Cannon's Conjecture, M. Bonk and D. Meyer [BM10, BM17], as well as P. Haïssinsky and K.M. Pilgrim [HP09], studied a subclass of Thurston maps, called *expanding Thurston maps*, by imposing some additional condition of expansion (see Definition 3.5). These maps are characterized by a contraction property for inverse images (see Subsection 3.2 for the precise definition). In particular, a rational Thurston map on $\widehat{\mathbb{C}}$ is expanding if and only if its Julia set is equal to $\widehat{\mathbb{C}}$. For an expanding Thurston map on S^2, we can equip S^2 with a natural class of metrics d, called *visual metrics*, that are snowflake equivalent to each other and are constructed in a similar way as the visual metrics on the boundary $\partial_\infty G$ of a Gromov hyperbolic group G (see [BM17, Chapter 8] for details, and see [HP09] for a related construction).

In this paper we study the dynamics and properties of expanding Thurston maps from the point of view of ergodic theory. The ergodic theory for expanding Thurston maps has been investigated in [BM10, BM17, HP09, Li18, Li15, Li17]. In [Li18], the first-named author developed the thermodynamic formalism and investigated the existence, uniqueness, and other properties of equilibrium states for expanding Thurston maps. In [Li15], for expanding Thurston maps without periodic critical points, by proving that the (measure-theoretic) entropy map is upper semi-continuous and then applying a general framework devised by Y. Kifer [Kif90] and reformulated by H. Comman and J. Rivera-Letelier [CRL11], the first-named author established level-2 large deviation principles for iterated preimages and periodic points with respect to equilibrium states and obtains the corresponding equidistribution results.

However, for expanding Thurston maps with a periodic critical point, upper semi-continuity of the entropy map, level-2 large deviation principles, and equidistribution of periodic points with respect to equilibrium states remained open.

In the present paper, for any expanding Thurston map, even in the presence of periodic critical points, we prove entropy density of ergodic measures, establish level-2 large deviation principles for the distributions of Birkhoff averages, periodic points, and iterated preimages, and conclude that periodic points and iterated preimages are equidistributed with respect to the unique equilibrium state for a potential that is Hölder continuous with respect to a visual metric on S^2. Specifically, these results extend the corresponding findings in [Li15], and the methods we use do not depend on the upper semi-continuity of the entropy map.

Our results answer the two questions posed in [Li15] by the first-named author of the current paper. More precisely, we show that the entropy map of an expanding Thurston map $f\colon S^2 \to S^2$ is not upper semi-continuous when f has at least one periodic critical point. This result gives a negative answer to Question 1 posed in [Li15, p. 523]. Moreover, it suggests that the method used there to prove large deviation principles does not apply to expanding Thurston maps with at least one periodic critical point. In order to answer Question 2 posed in [Li15, p. 523] positively, i.e., obtain the equidistribution results even in the presence of periodic critical points, we show that the equilibrium state is the unique minimizer of the rate function and then apply the level-2 large deviation principles.

One of the tools that used in this paper are called subsystems of expanding Thurston maps (see Subsection 3.3), introduced and investigated in [LSZ25, LS24]. The notion of subsystems is inspired by a translation of the notion of subgroups from geometric group theory via Sullivan's dictionary. We remark that subsystems are not only useful tools for studying the ergodic theory of expanding Thurston maps, but they also have geometric significance in themselves. According to Sullivan's dictionary, an expanding Thurston map is associated with a Gromov hyperbolic group whose boundary at infinity is S^2. In this context, a subsystem corresponds to a Gromov hyperbolic group whose boundary at infinity is a subset of S^2. In particular, for Gromov hyperbolic groups whose boundary at infinity is a Sierpiński carpet, there is an analog of Cannon's conjecture—the Kapovich–Kleiner conjecture. It predicts that these groups arise from some standard situation in hyperbolic geometry. Similar to Cannon's conjecture, one can reformulate the Kapovich–Kleiner conjecture in an equivalent way as a question related to quasisymmetric uniformization. For subsystems, it is easy to find examples where the tile maximal invariant set is homeomorphic to the standard Sierpiński carpet (see Subsection 3.3 for examples of subsystems). In this case, an analog of the Kapovich–Kleiner conjecture for subsystems is under investigation [BLL].

The main challenges in this work arise from two factors: the non-uniform expansion of the dynamics and the lack of upper semi-continuity of the entropy map due to periodic critical points. To tackle the first challenge, we need to investigate the geometric properties of visual metrics and their interaction with the combinatorial structures associated with subsystems. On the one hand, by utilizing visual metrics, we can transform these difficulties into more manageable topological and combinatorial problems. On the other hand, by constructing suitable subsystems and applying the thermodynamic formalism for subsystems developed in the series of papers [LSZ25, LS24], we can obtain measures with controllable measure-theoretic pressures, enabling us to obtain the desired bounds and prove our results. Due to the second challenge, one cannot prove large deviation principles by applying a variant of Y. Kifer's result [Kif90, Theorem 4.3], as formulated by H. Comman and J. Rivera-Letelier [CRL11, Theorem C], which is the main necessary tool used in [CRL11, Li15]. Our proofs follow a different approach and, in particular, do not rely on the upper semi-continuity of the entropy map (see Subsection 1.2 for details).

We remark that in [DPTUZ21], the authors addressed the periodic critical points by first blowing up the sphere S^2 along the preimages of the critical orbits, thereby obtaining a coarse expanding map on a Sierpiński carpet without periodic critical points. They then encoded the dynamics of the system using a geometric coding tree, following the approach in [Prz85], which yields a semi-conjugacy with the shift map. The key point is that this semi-conjugacy preserves the entropy of an equilibrium state (we also use this property to prove that the equilibrium state is the unique minimizer of the rate function, see Proposition 7.3 (iii) and Theorem 7.5). Therefore, the authors can apply classical results for shift maps to a coarse expanding map with periodic critical points without any loss of entropy, thereby establishing the thermodynamic formalism and statistical laws. However, in our context, we do not see any obvious way to extend such a property to general invariant measures, not just the equilibrium state. In addition, the relationship between the quantities associated with periodic orbits in expanding Thurston maps and shift maps remains unclear and maybe insufficient for proving large deviation principles.

1.1. **Main results.** Our results consist of three parts. We first show that the entropy map of an expanding Thurston map $f\colon S^2 \to S^2$ is upper semi-continuous if and only if f has no periodic critical points. Then for every expanding Thurston map, we prove that the set of ergodic measures is entropy-dense in the space of invariant measure. Finally, we establish level-2 large deviation principles for the distributions of Birkhoff averages, periodic points, and iterated preimages, and conclude that periodic points and iterated preimages are equidistributed with respect to the equilibrium state.

We now state our results precisely.

Upper semi-continuity of entropy. The entropy map of a continuous map $T\colon X \to X$ defined on a compact metric space (X, d) is the map $\mu \mapsto h_\mu(T)$ which is defined on the space of T-invariant

Borel probability measures $\mathcal{M}(X,T)$, where $h_\mu(T)$ is the measure-theoretic entropy of T for μ (see Subsection 3.1 and Definition 5.1) and $\mathcal{M}(X,T)$ is equipped with the weak*-topology.

Our first result is about the upper semi-continuity of the entropy map for expanding Thurston maps.

Theorem 1.1. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Then the entropy map of f is upper semi-continuous if and only if f has no periodic critical points.*

We remark that for expanding Thurston maps without periodic critical points, the upper semi-continuity of the entropy map has been established in [Li15, Corollary 1.3] by proving a stronger property called asymptotic h-expansiveness (see [Mis76]). In the present paper, we complete the "only if" part in Theorem 1.1 through concrete constructions. These constructions show that the entropy map is not upper semi-continuous even when restricted to the set of ergodic measures. Moreover, we estimate the defects in semi-continuity (see Theorem 5.5).

The continuity properties of the entropy map have been studied for a long time. A classical result is that for an expansive homeomorphism defined on a compact metric space the entropy map is upper semi-continuous (see for example, [Wal82, Theorem 8.2]). M.Yu. Lyubich [Lyu83, Corollary 1] showed that for rational maps on the Riemann sphere the entropy map is upper semi-continuous. Another fundamental result is that for a C^∞ map defined on a smooth compact manifold the entropy map is upper semi-continuous (see [New89, Theorem 4.1] and [Yom87]). While for C^r diffeomorphisms with finite r, upper semi-continuity of the entropy map may fail (for examples in dimension four see [Mis73] and for examples in dimension two see [Buz14]). In this setting, J. Buzzi, S. Crovisier, and O. Sarig estimated the discontinuities of the entropy map in terms of Lyapunov exponents (see [BCS22]). In the non-compact setting, for transitive countable Markov shift, the entropy map is upper semi-continuous if the shift map has finite topological entropy (see [ITV22, Theorem 8.1]). Otherwise the entropy map may not be upper semi-continuous (see [JMU05, p. 774]).

Entropy density of ergodic measures. Let $T\colon X \to X$ be a continuous map on a compact metric space (X,d) and $\mathcal{M}(X,T)$ be the set of T-invariant Borel probability measures on X. We say that a subset $\mathcal{N} \subseteq \mathcal{M}(X,T)$ is *entropy-dense* in $\mathcal{M}(X,T)$ if, for each $\mu \in \mathcal{M}(X,T)$, there exists a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathcal{N}$ such that $\{\mu_n\}_{n\in\mathbb{N}}$ converges to μ in the weak*-topology and $h_{\mu_n}(T) \to h_\mu(T)$ as $n \to +\infty$.

We show that the set of ergodic measures is entropy-dense for every expanding Thurston map.

Theorem 1.2. *For an expanding Thurston map $f\colon S^2 \to S^2$, the set of ergodic f-invariant measures is entropy-dense in $\mathcal{M}(S^2, f)$.*

Entropy density of ergodic measures guarantees that one can approximate non-ergodic measures with ergodic ones with similar entropy and similar expectations. Such a property plays an important role in large deviation theory, which was used in [FO88], and has been studied in various settings such as $\mathbb{Z}^d$ subshifts of finite type [EKW94], uniformly hyperbolic systems and β-shifts [PS05], ergodic group automorphisms [Yam09], and countable Markov Shifts [Tak20]. In addition, it has applications in the multifractal analysis (see for example, [IJ15]).

Level-2 large deviation principles. In order to present our results precisely, we briefly review some key concepts. We refer the reader to Subsection 7.1 for a detailed discussion.

Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of Borel probability measures on a Hausdorff topological space $\mathcal{X}$. We say that $\{\xi_n\}_{n\in\mathbb{N}}$ satisfies a *large deviation principle* in $\mathcal{X}$ if there exists a lower semi-continuous function $I\colon \mathcal{X} \to [0,+\infty]$ such that

$$\liminf_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{G}) \geqslant -\inf_{\mathcal{G}} I \qquad \text{for all open } \mathcal{G} \subseteq \mathcal{X},$$

and

$$\limsup_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{K}) \leqslant -\inf_{\mathcal{K}} I \qquad \text{for all closed } \mathcal{K} \subseteq \mathcal{X},$$

where $\log 0 = -\infty$, $\inf \emptyset = +\infty$, and $\sup \emptyset = -\infty$. Such a function I is called a *rate function*, and we call $x \in \mathcal{X}$ a *minimizer* if $I(x) = 0$ holds.

For expanding Thurston maps, we establish level-2 large deviation principles for the distributions of Birkhoff averages, periodic points, and iterated preimages.

Theorem 1.3. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map and d be a visual metric on S^2 for f. Let ϕ be a real-valued Hölder continuous function on S^2 with respect to the metric d. Let μ_ϕ be the unique equilibrium state for f and ϕ. We denote by $P(f,\phi)$ the topological pressure of f with respect to ϕ, and by $\mathcal{P}(S^2)$ the space of Borel probability measures on S^2 equipped with the weak*-topology.*

For each $n \in \mathbb{N}$, let $V_n\colon S^2 \to \mathcal{P}(S^2)$ be the continuous function defined by

$$V_n(x) \coloneqq \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(x)}, \tag{1.1}$$

and define $S_n\phi(x) \coloneqq \sum_{i=0}^{n-1} \phi(f^i(x))$ for each $x \in S^2$. Fix an arbitrary sequence $\{w_n\}_{n\in\mathbb{N}}$ of real-valued functions on S^2 satisfying $w_n(x) \in \left[1, \deg_{f^n}(x)\right]$ for each $n \in \mathbb{N}$ and each $x \in S^2$. For each $n \in \mathbb{N}$, we consider the following Borel probability measures on $\mathcal{P}(S^2)$.

Birkhoff averages. *$\Sigma_n \coloneqq (V_n)_*(\mu_\phi)$ (i.e., Σ_n is the push-forward of μ_ϕ by $V_n\colon S^2 \to \mathcal{P}(S^2)$).*

Periodic points. *With $\mathrm{Per}_n(f) \coloneqq \{p \in S^2 : f^n(p) = p\}$, put*

$$\Omega_n \coloneqq \sum_{p\in \mathrm{Per}_n(f)} \frac{w_n(p)\exp(S_n\phi(p))}{\sum_{p'\in\mathrm{Per}_n(f)} w_n(p')\exp(S_n\phi(p'))}\delta_{V_n(p)}. \tag{1.2}$$

Iterated preimages. *Given a sequence $\{x_j\}_{j\in\mathbb{N}}$ of points in S^2, put*

$$\Omega_n(x_n) \coloneqq \sum_{y\in f^{-n}(x_n)} \frac{w_n(y)\exp(S_n\phi(y))}{\sum_{y'\in f^{-n}(x_n)} w_n(y')\exp(S_n\phi(y'))}\delta_{V_n(y)}. \tag{1.3}$$

Then each of the sequences $\{\Sigma_n\}_{n\in\mathbb{N}}$, $\{\Omega_n\}_{n\in\mathbb{N}}$, and $\{\Omega_n(x_n)\}_{n\in\mathbb{N}}$ satisfies a large deviation principle in $\mathcal{P}(S^2)$ with the rate function $I_\phi\colon \mathcal{P}(S^2) \to [0,+\infty]$ given by

$$I_\phi(\mu) \coloneqq -\inf_{\mathcal{G}\ni\mu} \sup_{\mathcal{G}} F_\phi, \tag{1.4}$$

where the infimum is taken over all open sets $\mathcal{G} \subseteq \mathcal{P}(S^2)$ containing μ, and $F_\phi\colon \mathcal{P}(S^2) \to [-\infty, 0]$ is defined by

$$F_\phi(\mu) \coloneqq \begin{cases} h_\mu(f) + \int \phi \,\mathrm{d}\mu - P(f,\phi) & \text{if } \mu \in \mathcal{M}(S^2, f); \\ -\infty & \text{if } \mu \in \mathcal{P}(S^2) \setminus \mathcal{M}(S^2, f). \end{cases} \tag{1.5}$$

Moreover, μ_ϕ is the unique minimizer of the rate function I_ϕ, and each of the sequences $\{\Sigma_n\}_{n\in\mathbb{N}}$, $\{\Omega_n\}_{n\in\mathbb{N}}$, and $\{\Omega_n(x_n)\}_{n\in\mathbb{N}}$ converges to δ_{μ_ϕ} in the weak topology. Furthermore, for each convex open subset $\mathcal{G}$ of $\mathcal{P}(S^2)$ containing some invariant measure, we have $\inf_{\mathcal{G}} I_\phi = \inf_{\overline{\mathcal{G}}} I_\phi$,*

$$\lim_{n\to+\infty} \frac{1}{n}\log \Sigma_n(\mathcal{G}) = \lim_{n\to+\infty} \frac{1}{n}\log \Omega_n(\mathcal{G}) = \lim_{n\to+\infty} \frac{1}{n}\log \Omega_n(x_n)(\mathcal{G}) = -\inf_{\mathcal{G}} I_\phi, \tag{1.6}$$

and (1.6) remains true with $\mathcal{G}$ replaced by its closure $\overline{\mathcal{G}}$.

Remark 1.4. In Theorem 1.3, it can be verified that $\sup_{\mathcal{G}} F_\phi = \sup_{\mathcal{G}}(-I_\phi)$ for all open subsets $\mathcal{G} \subseteq \mathcal{P}(S^2)$, and the rate function I_ϕ is convex and lower semi-continuous. We call $-I_\phi$ the upper semi-continuous regularization of F_ϕ. Note that if f has at least one periodic critical point, then the rate function I_ϕ is not equal to $-F_\phi$, because the entropy map of f is not upper semi-continuous (see Theorem 1.1). More precisely, it follows from Theorem 1.1 that $I_\phi = -F_\phi$ if and only if f has no periodic critical points.

As an immediate consequence of Theorem 1.3, we get the following corollary.

Corollary 1.5. *Under the assumptions of Theorem 1.3, for each* $\mu \in \mathcal{M}(S^2, f)$ *and each convex local basis* G_μ *of* $\mathcal{P}(S^2)$ *at* μ*, we have*

$$\begin{aligned}
-I_\phi(\mu) &= \inf_{\mathcal{G}\in G_\mu}\left\{\lim_{n\to+\infty}\frac{1}{n}\log \mu_\phi(\{x\in S^2 : V_n(x)\in\mathcal{G}\})\right\}\\
&= \inf_{\mathcal{G}\in G_\mu}\left\{\lim_{n\to+\infty}\frac{1}{n}\log \sum_{p\in\mathrm{Per}_n(f), V_n(p)\in\mathcal{G}} w_n(p)\exp(S_n\phi(p))\right\} - P(f,\phi)\\
&= \inf_{\mathcal{G}\in G_\mu}\left\{\lim_{n\to+\infty}\frac{1}{n}\log \sum_{y\in f^{-n}(x_n), V_n(y)\in\mathcal{G}} w_n(y)\exp(S_n\phi(y))\right\} - P(f,\phi).
\end{aligned} \tag{1.7}$$

By applying the general theory of large deviations, particularly the contraction principle (see Theorem 7.1), we derive the following level-1 large deviation principles from Theorem 1.3.

Corollary 1.6. *Let* $\psi\colon S^2\to\mathbb{R}$ *be a continuous function, and let* $\widehat{\psi}\colon \mathcal{P}(S^2)\to\mathbb{R}$ *be defined by* $\widehat{\psi}(\mu) := \int\psi\,\mathrm{d}\mu$*. Under the assumptions of Theorem 1.3, each of the sequences* $\bigl\{\widehat{\psi}_*(\Sigma_n)\bigr\}_{n\in\mathbb{N}}$*,* $\bigl\{\widehat{\psi}_*(\Omega_n)\bigr\}_{n\in\mathbb{N}}$*, and* $\bigl\{\widehat{\psi}_*(\Omega_n(x_n)\bigr\}_{n\in\mathbb{N}}$ *satisfies a large deviation principle in* $\mathbb{R}$ *with the rate function* $J\colon\mathbb{R}\to[0,+\infty]$ *defined by*

$$J(\alpha) := \inf\left\{I_\phi(\mu) : \mu\in\mathcal{P}(S^2),\ \int\psi\,\mathrm{d}\mu=\alpha\right\}\qquad \text{for } \alpha\in\mathbb{R}. \tag{1.8}$$

Furthermore, if $c_\psi < d_\psi$*, where* $c_\psi := \min\bigl\{\int\psi\,\mathrm{d}\nu : \nu\in\mathcal{M}(S^2,f)\bigr\}$ *and* $d_\psi := \max\bigl\{\int\psi\,\mathrm{d}\nu : \nu\in\mathcal{M}(S^2,f)\bigr\}$*, then for each interval* $K\subseteq\mathbb{R}$ *intersecting* (c_ψ,d_ψ)*,*

$$\begin{aligned}
-\inf_{\alpha\in K} J(\alpha) &= \lim_{n\to+\infty}\frac{1}{n}\log\mu_\phi\left(\left\{x\in S^2 : \frac{1}{n}S_n\psi(x)\in K\right\}\right)\\
&= \lim_{n\to+\infty}\frac{1}{n}\log\left(\frac{\sum_{p\in\mathrm{Per}_n(f),\,\frac{1}{n}S_n\psi(p)\in K} w_n(p)\exp(S_n\phi(p))}{\sum_{p'\in\mathrm{Per}_n(f)} w_n(p')\exp(S_n\phi(p'))}\right)\\
&= \lim_{n\to+\infty}\frac{1}{n}\log\left(\frac{\sum_{y\in f^{-n}(x_n),\,\frac{1}{n}S_n\psi(y)\in K} w_n(y)\exp(S_n\phi(y))}{\sum_{y'\in f^{-n}(x_n)} w_n(y')\exp(S_n\phi(y'))}\right).
\end{aligned} \tag{1.9}$$

In addition, for all $\alpha\in\mathbb{R}$*,*

$$J(\alpha)\leqslant \widetilde{J}(\alpha) := \inf\left\{P(f,\phi)-h_\mu(f)-\int\phi\,\mathrm{d}\mu : \mu\in\mathcal{M}(S^2,f),\ \int\psi\,\mathrm{d}\mu=\alpha\right\}, \tag{1.10}$$

and $J(\alpha)=\widetilde{J}(\alpha)$ *for each* $\alpha\in(c_\psi,d_\psi)$ *if* $\psi\colon S^2\to\mathbb{R}$ *is Hölder continuous with respect to the visual metric* d*.*

Remark. When the function ψ is Hölder continuous, the level-1 rate function $\widetilde{J}$ defined in (1.10) is close related to multifractal analysis, which studies level sets of the form $K_\alpha^\psi := \{x\in X : \lim_{n\to+\infty}\frac{1}{n}\sum_{i=0}^{n-1}\psi(f^i(x)) = \alpha\}$. Each K_α^ψ is f-invariant, and the *entropy spectrum* of Birkhoff averages, i.e., the function $H_\psi\colon[c_\psi,d_\psi]\to[0,+\infty)$ defined by $H_\psi(\alpha)=h_{\mathrm{top}}\bigl(K_\alpha^\psi\bigr)$ (see Bowen [34] and Pesin and Pitskel [137] for the precise definition), is described by the family of equilibrium states $\{\mu_{t\psi}\}_{t\in\mathbb{R}}$, in the sense that $\eta\colon t\mapsto\int\psi\,\mathrm{d}\mu_{t\psi}$ is a homeomorphism from $\mathbb{R}$ onto (c_ψ,d_ψ), and for each $\alpha\in(c_\psi,d_\psi)$,

$$H_\psi(\alpha) = h_{\mu_{\eta^{-1}(\alpha)\psi}}(f) = \sup\left\{h_\mu(f) : \mu\in\mathcal{M}(S^2,f),\ \int\psi\,\mathrm{d}\mu=\alpha\right\}.$$

The following equidistribution results follow from the corresponding level-2 large deviation principles and the uniqueness of the minimizer of the rate function.

Theorem 1.7. *Let* $f\colon S^2\to S^2$ *be an expanding Thurston map and* d *be a visual metric on* S^2 *for* f*. Let* ϕ *be a real-valued Hölder continuous function on* S^2 *with respect to the metric* d*. Let* μ_ϕ *be* *the unique equilibrium state for f and ϕ. Fix an arbitrary sequence $\{w_n\}_{n\in\mathbb{N}}$ of real-valued functions on S^2 satisfying $w_n(x) \in \left[1, \deg_{f^n}(x)\right]$ for each $n \in \mathbb{N}$ and each $x \in S^2$. For each $n \in \mathbb{N}$, denote $S_n\phi(x) \coloneqq \sum_{i=0}^{n-1} \phi(f^i(x))$ for each $x \in S^2$, and consider the following Borel probability measures on S^2.*

Periodic points. *With* $\operatorname{Per}_n(f) \coloneqq \{p \in S^2 : f^n(p) = p\}$, *put*

$$\mu_n \coloneqq \sum_{p\in\operatorname{Per}_n(f)} \frac{w_n(p)\exp(S_n\phi(p))}{\sum_{p'\in\operatorname{Per}_n(f)} w_n(p')\exp(S_n\phi(p'))} \frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(p)}.$$

Iterated preimages. *Given a sequence $\{x_j\}_{j\in\mathbb{N}}$ of points in S^2, put*

$$\nu_n \coloneqq \sum_{y\in f^{-n}(x_n)} \frac{w_n(y)\exp(S_n\phi(y))}{\sum_{z\in f^{-n}(x_n)} w_n(z)\exp(S_n\phi(z))} \frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(y)}.$$

Then each of the sequences $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}}$ converges to μ_ϕ in the weak-topology.*

See Subsection 7.7 for the proof of Theorem 1.7.

Remark 1.8. Since $S_n\phi(f^i(p)) = S_n\phi(p)$ for each $i \in \mathbb{N}$ if $p \in \operatorname{Per}_n(f)$, we get

$$\mu_n = \sum_{p\in\operatorname{Per}_n(f)} \frac{\frac{S_n w_n(p)}{n}\exp(S_n\phi(p))}{\sum_{p'\in\operatorname{Per}_n(f)} w_n(p')\exp(S_n\phi(p'))}\delta_p$$

for each $n \in \mathbb{N}$. In particular, when $w_n(\cdot) \equiv 1$, we have

$$\mu_n = \sum_{p\in\operatorname{Per}_n(f)} \frac{\exp(S_n\phi(p))}{\sum_{p'\in\operatorname{Per}_n(f)} \exp(S_n\phi(p'))}\delta_p;$$

when $w_n(\cdot) \equiv \deg_{f^n}(\cdot)$, since $\deg_{f^n}(f^i(p)) = \deg_{f^n}(p)$ for each $i \in \mathbb{N}$ if $p \in \operatorname{Per}_n(f)$, we have

$$\mu_n = \sum_{p\in\operatorname{Per}_n(f)} \frac{\deg_{f^n}(p)\exp(S_n\phi(p))}{\sum_{p'\in\operatorname{Per}_n(f)} \deg_{f^n}(p')\exp(S_n\phi(p'))}\delta_p.$$

We remark that the novelty of Theorem 1.3, Corollary 1.5, and Theorem 1.7, which generalize the corresponding results in [Li15], lies in their application to *all* expanding Thurston maps, including those with periodic critical points. Additionally, Corollary 1.6 partially extends [DPTUZ21, Theorem 1.2 (5)].

1.2. **Strategy and organization.** We now discuss the strategies of proofs of our main results and describe the organization of the paper.

To prove Theorem 1.1, by [Li15, Corollary 1.3], it suffices to study expanding Thurston maps with periodic critical points and show that their entropy maps are not upper semi-continuous. The main point here is to find a sequence of invariant measures that converges in the weak*-topology and has an entropy drop at the limit. Our strategy is to construct a suitable sequence of subsystems and then apply the main results in [LSZ25, LS24] (see for example, Theorem 3.25) to verify that the sequence of measures of maximal entropy associated with the sequence of subsystems satisfies our desired properties. The construction of such a sequence of subsystems is based on the key observation that the local degree at a periodic critical point increases exponentially under iteration. Hence we can find subsystems around periodic critical points such that entropies of the associated measures of maximal entropy have a uniform positive lower bound.

The main idea behind the proof of Theorem 1.2 is to find an ergodic measure that is close to a given invariant measure in terms of both topology and entropy. To achieve this, we construct a suitable subsystem and use the corresponding measure of maximal entropy to approximate the given invariant measure. The construction of the subsystem poses the main difficulty. Our strategy involves using pairs (as described in Subsection 3.2) instead of tiles to build a strongly primitive subsystem. While the set of pairs is not a generator (unlike the set of tiles, as shown in Lemma 6.3), we first approximate ergodic measures with a finite collection of tiles in a specific sense (Lemma 6.2). We then construct pairs from these tiles. Furthermore, to obtain a primitive subsystem, we add one suitable pair contained in the interior of the corresponding 0-tile for each color. The existence of such pairs is guaranteed by Lemma 6.4. Finally, with these preparations we are able to prove the entropy density of ergodic measures.

The proof of Theorem 1.3 is more involved and we divide the proof into four parts: the uniqueness of the minimizer, characterizations of topological pressures, the large deviation lower bound, and the large deviation upper bound. Here the characterizations of topological pressures are used in the proof of large deviation lower and upper bounds, and the uniqueness of the minimizer is used to prove the convergence of distributions and equidistribution results, which is necessary since in our setting the entropy map may not be upper semi-continuous. It should be noted that one cannot derive the uniqueness of the minimizer directly from the uniqueness of the equilibrium states (recall Remark 1.4).

To prove the uniqueness of the minimizer, we use a semi-conjugacy of an expanding Thurston map to a shift map, and show that the uniqueness of the minimizer follows from ergodic properties of the shift map and the uniqueness of the equilibrium state. Here a key property of such a semi-conjugacy is that even in the presence of periodic critical points, the entropy at the equilibrium state does not drop under the projection from the symbolic space (see Proposition 7.3 (iii)). The existence and properties of such a semi-conjugacy are proved in [DPTUZ21].

The characterization of topological pressures in terms of periodic points for expanding Thurston maps has been established in [Li15, Propositions 6.8] (see Theorem 7.7), while for iterated preimages such characterization was only obtained for expanding Thurston maps without periodic critical points (compare Theorem 7.8 with Theorem 7.7). By carefully analyzing the combinatorics of critical points, we establish the characterization of topological pressures in terms of iterated preimages for all expanding Thurston maps (see Theorem 7.9).

In the proof of large deviation lower and upper bounds we take an indirect approach. We first consider an expanding Thurston map that has an invariant Jordan curve containing the postcritical set. In such situations the associated cell decompositions have nice compatibility properties, enabling the application of the results in the ergodic theory of subsystems established in [LSZ25, LS24]. However, such an invariant Jordan curve may not exist (see for example, [BM17, Example 15.11]). Our strategy is to first establish large deviation bounds for sufficiently high iterates of an expanding Thurston map, where such an invariant Jordan curve does exist (see Lemma 3.7), and then prove for the original expanding Thurston map.

An important property for the proof of the large deviation lower bound for all open sets is *entropy approachability of ergodic measures* (as defined in Definition 7.11), which guarantees that any invariant measure can be approximated by an ergodic measure with entropy sufficiently large. This property and is weaker than the entropy density of ergodic measures and allows for the simplification of the proof of the lower bound to the case where the measure being considered is ergodic (see also [EKW94, FO88, You90, Tak19]). Then we obtain estimates for ergodic measures by using the approximations (Lemma 6.2) in the proof of the entropy density established in Section 6.

The main idea for the large deviation upper bound is to construct measures whose measure-theoretic pressures provide desired upper bounds for the deviations. Our strategy is to construct certain strongly primitive subsystems (see Definition 3.20) and apply the results in [LSZ25, LS24] to produce such measures (see Proposition 7.19). Similar strategies are used in one-dimensional real dynamics [CRLT19] and countable Markov shift [Tak19]. Constructions and estimates for the upper bound are much harder than those for the lower bound and are necessarily involved due to the presence of critical points and the lack of upper semi-continuity of entropy.

Finally, to prove Theorem 1.7, we use the property that the equilibrium state is the unique minimizer of the rate function and then apply the results from large deviation principles.

We now describe the structure of this paper.

In Section 2, we fix some notation that will be used throughout the paper. In Section 3, we first review some notions from ergodic theory and dynamical systems and go over some key concepts and results on Thurston maps. Then we summarize some concepts and results on subsystems of expanding Thurston maps. In Section 4, we state the assumptions regarding some of the objects in this paper, which we will repeatedly refer to later as the *Assumptions* in Section 4.

In Section 5, we investigate the upper semi-continuity of the entropy map of an expanding Thurston map and prove Theorem 1.1. In Section 6, we show that ergodic measures are entropy-dense for expanding Thurston maps and prove Theorem 1.2.

Section 7 is devoted to the study of large deviation principles and equidistribution results. In Subsection 7.1, we give a brief review of large deviation principles. In Subsection 7.2, we show that the equilibrium state is the unique minimizer of the rate function. In Subsection 7.3, we establish characterizations of the topological pressure in terms of periodic points and iterated preimages. In Subsections 7.4 and 7.5, we prove the large deviation lower bound for all open sets and upper bound for all closed sets. In Subsection 7.6, by combining the previous results together, we finish the proof of Theorem 1.3 and its two corollaries. In Subsection 7.7, we show that periodic points and iterated preimages are equidistributed with respect to the unique equilibrium state, establishing Theorem 1.7.

Acknowledgments. The first-named author wants to thank Peter Haïssinky and Juan Rivera-Letelier for discussions on the upper semi-continuity property of the entropy map. Z. Li and X. Shi were partially supported by NSFC Nos. 12101017, 12090010, 12090015, 12471083, and BJNSF No. 1214021.

2. Notation

Let $\mathbb{C}$ be the complex plane and $\widehat{\mathbb{C}}$ be the Riemann sphere. Let S^2 denote an oriented topological 2-sphere. We use $\mathbb{N}$ to denote the set of integers greater than or equal to 1 and write $\mathbb{N}_0 \coloneqq \{0\} \cup \mathbb{N}$. For $x \in \mathbb{R}$, we define $\lfloor x \rfloor$ as the greatest integer $\leqslant x$, and $\lceil x \rceil$ the smallest integer $\geqslant x$. The cardinality of a set A is denoted by $\operatorname{card}(A)$. The symmetric difference of two subsets A and B of a set X is defined as

$$A \bigtriangleup B \coloneqq (A \setminus B) \cup (B \setminus A).$$

Let $g\colon X \to Y$ be a map between two sets X and Y. We denote the restriction of g to a subset Z of X by $g|_Z$.

Consider a map $f\colon X \to X$ on a set X. The inverse map of f is denoted by f^{-1}. We write f^n for the n-th iterate of f, and $f^{-n} \coloneqq (f^n)^{-1}$, for $n \in \mathbb{N}$. We set $f^0 \coloneqq \mathrm{id}_X$, the identity map on X. For a real-valued function $\varphi\colon X \to \mathbb{R}$, we write

$$S_n\varphi(x) = S_n^f\varphi(x) \coloneqq \sum_{j=0}^{n-1} \varphi\bigl(f^j(x)\bigr)$$

for $x \in X$ and $n \in \mathbb{N}_0$. We omit the superscript f when the map f is clear from the context. Note that when $n = 0$, by definition we always have $S_0\varphi = 0$.

Let (X, d) be a metric space. For subsets A, $B \in X$, we set $d(A, B) \coloneqq \inf\{d(x, y) : x \in A,\, y \in B\}$. For each subset $Y \subseteq X$, we denote the diameter of Y by $\operatorname{diam}_d(Y) \coloneqq \sup\{d(x, y) : x,\, y \in Y\}$, the interior of Y by $\operatorname{int}(Y)$, and the characteristic function of Y by $\mathbb{1}_Y$, which maps each $x \in Y$ to $1 \in \mathbb{R}$ and vanishes otherwise. For each $r > 0$ and each $x \in X$, we denote the open (resp. closed) ball of radius r centered at x by $B_d(x, r)$ (resp. $\overline{B_d}(x, r)$). We often omit the metric d in the subscript when it is clear from the context.

For a compact metrizable topological space X, we denote by $C(X)$ (resp. $B(X)$) the space of continuous (resp. bounded Borel) functions from X to $\mathbb{R}$, by $\mathcal{M}(X)$ the set of finite signed Borel measures, and $\mathcal{P}(X)$ the set of Borel probability measures on X. For $\mu \in \mathcal{M}(X)$, we denote by $\operatorname{supp}\mu$ the support of μ (the smallest closed set $A \subseteq X$ such that $|\mu|(X \setminus A) = 0$). For a point $x \in X$, we denote by δ_x the Dirac measure supported on $\{x\}$. If we do not specify otherwise, we equip $C(X)$ with the uniform norm $\|\cdot\|_{C(X)} \coloneqq \|\cdot\|_\infty$, and equip $\mathcal{M}(X)$ and $\mathcal{P}(X)$ with the weak* topology. According to the Riesz representation theorem (see for example, [Fol99, Theorems 7.17 and 7.8]), we identify the dual of $C(X)$ with the space $\mathcal{M}(X)$.

The space of real-valued Hölder continuous functions with an exponent $\beta \in (0,1]$ on a compact metric space (X,d) is denoted as $C^{0,\beta}(X,d)$. For each $\phi \in C^{0,\beta}(X,d)$,

$$|\phi|_\beta \coloneqq \sup\biggl\{ \frac{|\phi(x)-\phi(y)|}{d(x,y)^\beta} : x,\, y \in X,\, x \neq y \biggr\},$$

and the Hölder norm is defined as $\|\phi\|_{C^{0,\beta}} \coloneqq |\phi|_\beta + \|\phi\|_{C(X)}$.

3. Preliminaries

3.1. **Thermodynamic formalism.** We first review some basic concepts from ergodic theory and dynamical systems. For more detailed studies of these concepts, we refer the reader to [Wal82, Chapter 9] and [KH95, Chapter 20].

Let (X,d) be a compact metric space and $g\colon X \to X$ a continuous map. Given $n \in \mathbb{N}$,

$$d_g^n(x,y) \coloneqq \max\bigl\{ d\bigl(g^k(x), g^k(y)\bigr) : k \in \{0,\,\dots,\,n-1\} \bigr\}, \qquad \text{for } x,\, y \in X,$$

defines a metric on X. A set $F \subseteq X$ is (n,ϵ)*-separated* (with respect to g), for some $n \in \mathbb{N}$ and $\epsilon > 0$, if for each pair of distinct points $x,\, y \in F$, we have $d_g^n(x,y) \geqslant \epsilon$.

For each real-valued continuous function $\psi \in C(X)$, the following two limits exist and are equal (see for example, [KH95, Subsection 20.2]), which we denote by $P(g,\psi)$:

$$P(g,\psi) \coloneqq \lim_{\epsilon \to 0^+} \limsup_{n\to+\infty} \frac{1}{n} \log N_d(g,\psi,\varepsilon,n) = \lim_{\epsilon \to 0^+} \liminf_{n\to+\infty} \frac{1}{n} \log N_d(g,\psi,\varepsilon,n), \tag{3.1}$$

where $N_d(g,\psi,\varepsilon,n) \coloneqq \sup\bigl\{ \sum_{x\in E} \exp\bigl(S_n\psi(x)\bigr) : E \subseteq X \text{ is } (n,\varepsilon)\text{-separated with respect to } g \bigr\}$. We call $P(g,\psi)$ the *topological pressure* of g with respect to the *potential* ψ. In particular, when $\psi = 0$, the quantity $h_{\mathrm{top}}(g) \coloneqq P(g,0)$ is called the *topological entropy* of g. Note that $P(g,\psi)$ is independent of d as long as the topology on X defined by d remains the same (see for example, [KH95, Subsection 20.2]). Moreover, the topological pressure is well-behaved under iteration. Indeed, if $n \in \mathbb{N}$, then $P(g^n, S_n\psi) = nP(g,\psi)$ (see for example, [Wal82, Theorem 9.8 (i)]).

We denote by $\mathcal{B}$ the σ-algebra of all Borel sets on X. A measure on X is understood to be a Borel measure, i.e., one defined on $\mathcal{B}$. We call a measure μ on X g-invariant if

$$\mu\bigl(g^{-1}(A)\bigr) = \mu(A)$$

for all $A \in \mathcal{B}$. We denote by $\mathcal{M}(X,g)$ the set of all g-invariant Borel probability measures on X.

Let $\mu \in \mathcal{M}(X,g)$. Then we say that g is *ergodic* for μ (or μ is *ergodic* for g) if for each set $A \in \mathcal{B}$ with $g^{-1}(A) = A$ we have $\mu(A) = 0$ or $\mu(A) = 1$. We denote by $\mathcal{M}_{\mathrm{erg}}(X,g)$ the set of all g-invariant ergodic measures on X.

Let $\mu \in \mathcal{M}(X,g)$. A *measurable partition* ξ for (X,μ) is a countable collection $\xi = \{A_i : i \in I\}$ of sets in $\mathcal{B}$ such that $\mu(A_i \cap A_j) = 0$ for each pair of $i,\, j \in I$ with $i \neq j$, and

$$\mu\Bigl(X \setminus \bigcup_{i\in I} A_i \Bigr) = 0.$$

Here I is a countable (i.e., finite or countably infinite) index set. The measurable partition ξ is finite if the index set J is a finite set.

Two measurable partitions ξ and η for (X,μ) are called *equivalent* if there exists a bijection between the sets of positive measure in ξ and the sets of positive measure in η such that corresponding sets have a symmetric difference of μ-measure zero. Roughly speaking, this means that the partitions are the same up to sets of measure zero.

Let $\xi = \{A_j : j \in J\}$ and $\eta = \{B_k : k \in K\}$ be measurable partitions for (X,μ). We say ξ is a *refinement* of η if for each $A_j \in \xi$, there exists $B_k \in \eta$ such that $A_j \subseteq B_k$. The *common refinement* (or *join*) $\xi \vee \eta$ of ξ and η defined as

$$\xi \vee \eta \coloneqq \{A_j \cap B_k : j \in J,\, k \in K\}$$

is also a measurable partition. Put $g^{-1}(\xi) \coloneqq \{g^{-1}(A_j) : j \in J\}$, and for each $n \in \mathbb{N}$ define

$$\xi_g^n \coloneqq \bigvee_{j=0}^{n-1} g^{-j}(\xi) = \xi \vee g^{-1}(\xi) \vee \cdots \vee g^{-(n-1)}(\xi).$$

Let ξ be a finite measurable partition for (g, μ) and $\mathcal{A}$ be the smallest σ-algebra containing all sets in the partitions ξ_g^n, $n \in \mathbb{N}$. We call ξ a *generator* for (g, μ) if for each Borel set $B \in \mathcal{B}$ there exists a set $A \in \mathcal{A}$ such that $\mu(A \triangle B) = 0$. Note that if for every set $B \in \mathcal{B}$ and every $\varepsilon > 0$, there exists $n \in \mathbb{N}$ and a union A of sets in ξ_g^n with $\mu(A \triangle B) < \varepsilon$, then ξ is a generator for (g, μ).

Let $\xi = \{A_j : j \in J\}$ be a measurable partition of X and $\mu \in \mathcal{M}(X, g)$ be a g-invariant Borel probability measure on X. The *entropy* of ξ is $H_\mu(\xi) \coloneqq -\sum_{j\in J} \mu(A_j) \log\bigl(\mu(A_j)\bigr) \in [0, +\infty]$, where $0 \log 0$ is defined to be zero. One can show that (see for example, [Wal82, Chapter 4]) if $H_\mu(\xi) < +\infty$, then the following limit exists:

$$h_\mu(g, \xi) \coloneqq \lim_{n\to+\infty} \frac{1}{n} H_\mu(\xi_g^n) \in [0, +\infty).$$

The quantity $h_\mu(g, \xi)$ is called the *measure-theoretic entropy of g relative to ξ*. The *measure-theoretic entropy* of g for μ is defined as

$$h_\mu(g) \coloneqq \sup\{h_\mu(g, \xi) : \xi \text{ is a measurable partition of } X \text{ with } H_\mu(\xi) < +\infty\}.$$

If $\mu \in \mathcal{M}(X, g)$ and $n \in \mathbb{N}$, then (see for example, [KH95, Proposition 4.3.16 (4)])

$$h_\mu(g^n) = n h_\mu(g). \tag{3.2}$$

If $t \in [0, 1]$ and $\nu \in \mathcal{M}(X, g)$ is another measure, then (see for example, [Wal82, Theorem 8.1])

$$h_{t\mu+(1-t)\nu}(g) = t h_\mu(g) + (1 - t) h_\nu(g). \tag{3.3}$$

For each real-valued continuous function $\psi \in C(X)$, the *measure-theoretic pressure* $P_\mu(g, \psi)$ of g for the measure $\mu \in \mathcal{M}(X, g)$ and the potential ψ is

$$P_\mu(g, \psi) \coloneqq h_\mu(g) + \int \psi \,\mathrm{d}\mu.$$

The topological pressure is related to the measure-theoretic pressure by the so-called *Variational Principle*. It states that (see for example, [KH95, Theorem 20.2.4])

$$P(g, \psi) = \sup\{P_\mu(g, \psi) : \mu \in \mathcal{M}(X, g)\} \tag{3.4}$$

for each $\psi \in C(X)$. In particular, when ψ is the constant function 0,

$$h_{\mathrm{top}}(g) = \sup\{h_\mu(g) : \mu \in \mathcal{M}(X, g)\}. \tag{3.5}$$

A measure μ that attains the supremum in (3.4) is called an *equilibrium state* for the map g and the potential ψ. A measure μ that attains the supremum in (3.5) is called a *measure of maximal entropy* of g.

Let $\widetilde{X}$ be another compact metric space. If μ is a measure on X and the map $\pi \colon X \to \widetilde{X}$ is continuous, then the *push-forward* $\pi_*\mu$ of μ by π is the measure given by $\pi_*\mu(A) \coloneqq \mu\bigl(\pi^{-1}(A)\bigr)$ for each Borel set $A \subseteq \widetilde{X}$. Note that if $\widetilde{X} = X$, then μ is π-invariant if and only if $\pi_*\mu = \mu$.

Suppose $\widetilde{g} \colon \widetilde{X} \to \widetilde{X}$ is a continuous map, $\mu \in \mathcal{M}(X, g)$, and $\widetilde{\mu} \in \mathcal{M}(\widetilde{X}, \widetilde{g})$. Then the dynamical system $(\widetilde{X}, \widetilde{g}, \widetilde{\mu})$ is called a *factor* of (X, g, μ) if there exists a continuous and surjective map $\pi \colon X \to \widetilde{X}$ such that $\pi_*\mu = \widetilde{\mu}$ and $\widetilde{g} \circ \pi = \pi \circ g$. In this case, $h_{\widetilde{\mu}}(\widetilde{g}) \leqslant h_\mu(g)$ (see for example, [KH95, Proposition 4.3.16 (1)]).

3.2. **Thurston maps.** In this subsection, we go over some key concepts and results on Thurston maps, and expanding Thurston maps in particular. For a more thorough treatment of the subject, we refer to [BM17, Li17, LSZ25].

Let S^2 denote an oriented topological 2-sphere and $f\colon S^2 \to S^2$ be a branched covering map. We denote by $\deg_f(x)$ the local degree of f at $x \in S^2$. The *degree* of f is $\deg f = \sum_{x\in f^{-1}(y)} \deg_f(x)$ for $y \in S^2$ and is independent of y.

A point $x \in S^2$ is a *critical point* of f if $\deg_f(x) \geqslant 2$. The set of critical points of f is denoted by $\operatorname{crit} f$. A point $y \in S^2$ is a *postcritical point* of f if $y = f^n(x)$ for some $x \in \operatorname{crit} f$ and $n \in \mathbb{N}$. The set of postcritical points of f is denoted by $\operatorname{post} f$.

Definition 3.1 (Thurston maps)**.** A Thurston map is a branched covering map $f\colon S^2 \to S^2$ on S^2 with $\deg f \geqslant 2$ and $\operatorname{card}(\operatorname{post} f) < +\infty$.

We now recall the notation for cell decompositions of S^2 used in [BM17] and [Li17]. A *cell of dimension* n in S^2, $n \in \{1, 2\}$, is a subset $c \subseteq S^2$ that is homeomorphic to the closed unit ball $\overline{\mathbb{B}^n}$ in $\mathbb{R}^n$, where $\mathbb{B}^n$ is the open unit ball in $\mathbb{R}^n$. We define the *boundary of* c, denoted by ∂c, to be the set of points corresponding to $\partial\mathbb{B}^n$ under such a homeomorphism between c and $\overline{\mathbb{B}^n}$. The *interior of* c is defined to be $\operatorname{inte}(c) = c \setminus \partial c$. For each point $x \in S^2$, the set $\{x\}$ is considered as a *cell of dimension* 0 in S^2. For a cell c of dimension 0, we adopt the convention that $\partial c = \emptyset$ and $\operatorname{inte}(c) = c$.

Let $f\colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subseteq S^2$ be a Jordan curve containing $\operatorname{post} f$. Then the pair f and $\mathcal{C}$ induces natural cell decompositions $\mathbf{D}^n(f,\mathcal{C})$ of S^2, for each $n \in \mathbb{N}_0$, in the following way:

By the Jordan curve theorem, the set $S^2 \setminus \mathcal{C}$ has two connected components. We call the closure of one of them the *white* 0*-tile* for $(f,\mathcal{C})$, denoted by $X^0_{\mathfrak{w}}$, and the closure of the other one the *black* 0*-tile* for $(f,\mathcal{C})$, denoted be $X^0_{\mathfrak{b}}$. The set of 0*-tiles* is $\mathbf{X}^0(f,\mathcal{C}) \coloneqq \big\{X^0_{\mathfrak{b}},\, X^0_{\mathfrak{w}}\big\}$. The set of 0*-vertices* is $\mathbf{V}^0(f,\mathcal{C}) \coloneqq \operatorname{post} f$. We set $\overline{\mathbf{V}}^0(f,\mathcal{C}) \coloneqq \big\{\{x\} : x \in \mathbf{V}^0(f,\mathcal{C})\big\}$. The set of 0*-edges* $\mathbf{E}^0(f,\mathcal{C})$ is the set of the closures of the connected components of $\mathcal{C} \setminus \operatorname{post} f$. Then we get a cell decomposition

$$\mathbf{D}^0(f,\mathcal{C}) \coloneqq \mathbf{X}^0(f,\mathcal{C}) \cup \mathbf{E}^0(f,\mathcal{C}) \cup \overline{\mathbf{V}}^0(f,\mathcal{C})$$

of S^2 consisting of *cells of level* 0, or 0*-cells*.

We can recursively define the unique cell decomposition $\mathbf{D}^n(f,\mathcal{C})$, $n \in \mathbb{N}$, consisting of n*-cells* such that f is cellular for $(\mathbf{D}^{n+1}(f,\mathcal{C}), \mathbf{D}^n(f,\mathcal{C}))$. We refer to [BM17, Lemma 5.12] for more details. We denote by $\mathbf{X}^n(f,\mathcal{C})$ the set of n-cells of dimension 2, called n*-tiles*; by $\mathbf{E}^n(f,\mathcal{C})$ the set of n-cells of dimension 1, called n*-edges*; by $\overline{\mathbf{V}}^n(f,\mathcal{C})$ the set of n-cells of dimension 0; and by $\mathbf{V}^n(f,\mathcal{C})$ the set $\{x : \{x\} \in \overline{\mathbf{V}}^n(f,\mathcal{C})\}$, called the set of n*-vertices*. The k*-skeleton*, for $k \in \{0, 1\}$, of $\mathbf{D}^n(f,\mathcal{C})$ is the union of all n-cells of dimension k in this cell decomposition.

The following result proved in [BM17, Proposition 5.16] summarizes some properties of the cell decompositions $\mathbf{D}^n(f,\mathcal{C})$ defined above.

Proposition 3.2 (M. Bonk & D. Meyer [BM17])**.** *Let $k, n \in \mathbb{N}_0$, $f\colon S^2 \to S^2$ be a Thurston map, $\mathcal{C} \subseteq S^2$ be a Jordan curve with* $\operatorname{post} f \subseteq \mathcal{C}$*, and $m \coloneqq \operatorname{card}(\operatorname{post} f)$.*

(i) *The map f^k is cellular for $(\mathbf{D}^{n+k}(f,\mathcal{C}), \mathbf{D}^n(f,\mathcal{C}))$. In particular, if c is any $(n+k)$-cell, then $f^k(c)$ is an n-cell, and $f^k|_c$ is a homeomorphism of c onto $f^k(c)$.*

(ii) *The* 1*-skeleton of $\mathbf{D}^n(f,\mathcal{C})$ is equal to $f^{-n}(\mathcal{C})$. The* 0*-skeleton of $\mathbf{D}^n(f,\mathcal{C})$ is the set $\mathbf{V}^n(f,\mathcal{C}) = f^{-n}(\operatorname{post} f)$, and we have $\mathbf{V}^n(f,\mathcal{C}) \subseteq \mathbf{V}^{n+k}(f,\mathcal{C})$.*

(iii) *Every n-tile is an m-gon, i.e., the number of n-edges and the number of n-vertices contained in its boundary are equal to m.*

(iv) *Let $F \coloneqq f^k$ be an iterate of f with $k \in \mathbb{N}$. Then $\mathbf{D}^n(F,\mathcal{C}) = \mathbf{D}^{nk}(f,\mathcal{C})$.*

Remark 3.3. Note that for each n-edge $e^n \in \mathbf{E}^n(f,\mathcal{C})$, $n \in \mathbb{N}_0$, there exist exactly two n-tiles in $\mathbf{X}^n(f,\mathcal{C})$ containing e^n.

For $n \in \mathbb{N}_0$, we define the *set of black n-tiles* as

$$\mathbf{X}^n_{\mathfrak{b}}(f,\mathcal{C}) := \{X \in \mathbf{X}^n(f,\mathcal{C}) : f^n(X) = X^0_{\mathfrak{b}}\},$$

and the *set of white n-tiles* as

$$\mathbf{X}^n_{\mathfrak{w}}(f,\mathcal{C}) := \{X \in \mathbf{X}^n(f,\mathcal{C}) : f^n(X) = X^0_{\mathfrak{w}}\}.$$

From now on, if the map f and the Jordan curve $\mathcal{C}$ are clear from the context, we will sometimes omit $(f,\mathcal{C})$ in the notation above.

If we fix the cell decomposition $\mathbf{D}^n(f,\mathcal{C})$, $n \in \mathbb{N}_0$, we can define for each $v \in \mathbf{V}^n$ the *n-flower of v* as

$$W^n(v) := \bigcup \{\operatorname{inte}(c) : c \in \mathbf{D}^n(f,\mathcal{C}), v \in c\}. \tag{3.6}$$

Note that flowers are open (in the standard topology on S^2). Let $\overline{W}^n(v)$ be the closure of $W^n(v)$.

Remark 3.4. For each $n \in \mathbb{N}_0$ and each $v \in \mathbf{V}^n$, we have

$$\overline{W}^n(v) = X_1 \cup X_2 \cup \cdots \cup X_m,$$

where $m := 2\deg_{f^n}(v)$, and $X_1, X_2, \dots X_m$ are all the n-tiles that contain v as a vertex (see [BM17, Lemma 5.28]). Moreover, each flower is mapped under f to another flower in such a way that is similar to the map $z \mapsto z^k$ on the complex plane. More precisely, for each $n \in \mathbb{N}_0$ and each $v \in \mathbf{V}^{n+1}$, there exist orientation preserving homeomorphisms $\varphi\colon W^{n+1}(v) \to \mathbb{D}$ and $\eta\colon W^n(f(v)) \to \mathbb{D}$ such that $\mathbb{D}$ is the unit disk on $\mathbb{C}$, $\varphi(v) = 0$, $\eta(f(v)) = 0$, and

$$(\eta \circ f \circ \varphi^{-1})(z) = z^k$$

for all $z \in \mathbb{D}$, where $k := \deg_f(v)$. Let $\overline{W}^{n+1}(v) = X_1 \cup X_2 \cup \cdots \cup X_m$ and $\overline{W}^n(f(v)) = X'_1 \cup X'_2 \cup \cdots \cup X'_{m'}$, where $X_1, X_2, \dots X_m$ are all the $(n+1)$-tiles that contain v as a vertex, listed counterclockwise, and $X'_1, X'_2, \dots X'_{m'}$ are all the n-tiles that contain $f(v)$ as a vertex, listed counterclockwise, and $f(X_1) = X'_1$. Then $m = m'k$, and $f(X_i) = X'_j$ if $i \equiv j \pmod{k}$, where $k = \deg_f(v)$ (see Case 3 of the proof of [BM17, Lemma 5.24] for more details). In particular, $W^n(v)$ is simply connected. Moreover, it follows from Proposition 3.2 that the map f preserves the structure of flowers, or more precisely,

$$f(W^n(x)) = W^{n-1}(f(x)) \tag{3.7}$$

for each $n \in \mathbb{N}$ and each $x \in \mathbf{V}^n(f,\mathcal{C})$.

We denote, for each $x \in S^2$ and each $n \in \mathbb{Z}$, the *n-bouquet of x*

$$U^n(x) := \bigcup \{Y^n \in \mathbf{X}^n : \text{there exists } X^n \in \mathbf{X}^n \text{ with } x \in X^n,\ X^n \cap Y^n \neq \emptyset\} \tag{3.8}$$

if $n \geqslant 0$, and set $U^n(x) := S^2$ otherwise.

We can now define expanding Thurston maps.

Definition 3.5 (Expansion)**.** A Thurston map $f\colon S^2 \to S^2$ is called *expanding* if there exists a metric d on S^2 that induces the standard topology on S^2 and a Jordan curve $\mathcal{C} \subseteq S^2$ containing $\operatorname{post} f$ such that

$$\lim_{n\to+\infty} \max\{\operatorname{diam}_d(X) : X \in \mathbf{X}^n(f,\mathcal{C})\} = 0.$$

It is clear from Proposition 3.2 (iv) and Definition 3.5 that if f is an expanding Thurston map, so is f^n for each $n \in \mathbb{N}$. We observe that being expanding is a topological property of a Thurston map and independent of the choice of the metric d that generates the standard topology on S^2. By Lemma 6.2 in [BM17], it is also independent of the choice of the Jordan curve $\mathcal{C}$ containing $\operatorname{post} f$.

For an expanding Thurston map f, we can fix a particular metric d on S^2 called a *visual metric for f*. For the existence and properties of such metrics, see [BM17, Chapter 8]. For a visual metric d for f, there exists a unique constant $\Lambda > 1$ called the *expansion factor* of d (see [BM17, Chapter 8] for more details). One major advantage of a visual metric d is that in (S^2, d) we have good quantitative control over the sizes of the cells in the cell decompositions. We record [BM17, Proposition 8.4] here, which characterizes visual metrics based on this quantitative control.

Proposition 3.6. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map and d be a metric on S^2. Then d is a visual metric for f with expansion factor $\Lambda > 1$ if and only if there exists a constant $C \geqslant 1$ such that the following two conditions hold for all $n \in \mathbb{N}_0$:*

(i) *$d(\sigma, \tau) \geqslant C^{-1}\Lambda^{-n}$ whenever σ and τ are disjoint n-cells.*

(ii) *$C^{-1}\Lambda^{-n} \leqslant \operatorname{diam}_d(\tau) \leqslant C\Lambda^{-n}$ for all n-edges and all n-tiles τ.*

Here cells are defined in terms of some Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} f \subseteq \mathcal{C}$, and the constant $C \geqslant 1$ is independent of the cells and their level n.

A Jordan curve $\mathcal{C} \subseteq S^2$ is *f-invariant* if $f(\mathcal{C}) \subseteq \mathcal{C}$. If $\mathcal{C}$ is f-invariant with $\operatorname{post} f \subseteq \mathcal{C}$, then the cell decompositions $\mathbf{D}^n(f, \mathcal{C})$ have nice compatibility properties. In particular, $\mathbf{D}^{n+k}(f, \mathcal{C})$ is a refinement of $\mathbf{D}^n(f, \mathcal{C})$, whenever $n, k \in \mathbb{N}_0$. According to Example 15.11 in [BM17], such f-invariant Jordan curves containing $\operatorname{post} f$ need not exist. However, M. Bonk and D. Meyer [BM17, Theorem 15.1] proved that there exists an f^n-invariant Jordan curve $\mathcal{C}$ containing $\operatorname{post} f$ for each sufficiently large n depending on f. We record it below for the convenience of the reader.

Lemma 3.7 (M. Bonk & D. Meyer [BM17])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\widetilde{\mathcal{C}} \subseteq S^2$ be a Jordan curve with $\operatorname{post} f \subseteq \widetilde{\mathcal{C}}$. Then there exists an integer $N(f, \widetilde{\mathcal{C}}) \in \mathbb{N}$ such that for each $n \geqslant N(f, \widetilde{\mathcal{C}})$ there exists an f^n-invariant Jordan curve $\mathcal{C}$ isotopic to $\widetilde{\mathcal{C}}$ rel. $\operatorname{post} f$.*

For the convenience of the reader, we record Proposition 12.5 (ii) of [BM17] here.

Proposition 3.8 (M. Bonk & D. Meyer [BM17])**.** *Let $k, n \in \mathbb{N}_0$, $f\colon S^2 \to S^2$ be a Thurston map, and $\mathcal{C} \subseteq S^2$ be an f-invariant Jordan curve with $\operatorname{post} f \subseteq \mathcal{C}$. Then every $(n+k)$-tile X^{n+k} is contained in a unique k-tile X^k.*

The following distortion lemma follows immediately from [Li18, Lemma 5.1].

Lemma 3.9. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map, and $\mathcal{C} \subseteq S^2$ be a Jordan curve that satisfies $\operatorname{post} f \subseteq \mathcal{C}$ and $f^{n_{\mathcal{C}}}(\mathcal{C}) \subseteq \mathcal{C}$ for some $n_{\mathcal{C}} \in \mathbb{N}$. Let d be a visual metric on S^2 for f with expansion factor $\Lambda > 1$. Let $\phi \in C^{0,\beta}(S^2, d)$ be a real-valued Hölder continuous function with an exponent $\beta \in (0, 1]$. Then there exists a constant $C_1 \geqslant 0$ depending only on f, $\mathcal{C}$, d, ϕ, and β such that for all $n \in \mathbb{N}_0$, $X^n \in \mathbf{X}^n(f, \mathcal{C})$, and $x, y \in X^n$,*

$$|S_n\phi(x) - S_n\phi(y)| \leqslant C_1 d(f^n(x), f^n(y))^\beta \leqslant C_1 (\operatorname{diam}_d(S^2))^\beta. \tag{3.9}$$

Quantitatively, we choose

$$C_1 \coloneqq C_0 |\phi|_\beta \big/ \big(1 - \Lambda^{-\beta}\big), \tag{3.10}$$

where $C_0 > 1$ is a constant depending only on f, $\mathcal{C}$, and d.

We summarize the existence, uniqueness, and some basic properties of equilibrium states for expanding Thurston maps in the following theorem.

Theorem 3.10 (Z. Li [Li18])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map and d a visual metric on S^2 for f. Let $\phi \in C^{0,\beta}(S^2, d)$ be real-valued Hölder continuous functions with an exponent $\beta \in (0, 1]$. Then the following statements are satisfied:*

(i) *There exists a unique equilibrium state μ_ϕ for the map f and the potential ϕ.*

(ii) *If $\mathcal{C} \subseteq S^2$ is a Jordan curve containing $\operatorname{post} f$ with the property that $f^{n_{\mathcal{C}}}(\mathcal{C}) \subseteq \mathcal{C}$ for some $n_{\mathcal{C}} \in \mathbb{N}$, then $\mu_\phi\big(\bigcup_{i=0}^{+\infty} f^{-i}(\mathcal{C})\big) = 0$.*

Theorem 3.10 (i) is part of [Li18, Theorem 1.1]. Theorem 3.10 (ii) was established in [Li18, Proposition 7.1].

Actually, by [Li18, Theorem 5.16, Proposition 5.17], the equilibrium state μ_ϕ is a Gibbs measure. We explicitly formulate this result below for the convenience of the reader.

Proposition 3.11 (Z. Li [Li18])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map and $\mathcal{C} \subseteq S^2$ be a Jordan curve containing* $\operatorname{post} f$ *with the property that $f^{n_{\mathcal{C}}}(\mathcal{C}) \subseteq \mathcal{C}$ for some $n_{\mathcal{C}} \in \mathbb{N}$. Let d be a visual metric on S^2 for f and $\phi \in C^{0,\beta}(S^2, d)$ be a real-valued Hölder continuous function with an exponent $\beta \in (0,1]$. Then the equilibrium state μ_ϕ for f and ϕ is a Gibbs measure with respect to f, $\mathcal{C}$, and ϕ, with the constant $P_{\mu_\phi} = P(f,\phi)$, i.e., there exists a constant $C_{\mu_\phi} \geqslant 1$ such that for each $n \in \mathbb{N}_0$, each n-tile $X^n \in \mathbf{X}^n(f,\mathcal{C})$, and each $x \in X^n$, we have*

$$\frac{1}{C_{\mu_\phi}} \leqslant \frac{\mu_\phi(X^n)}{\exp(S_n\phi(x) - nP(f,\phi))} \leqslant C_{\mu_\phi}. \tag{3.11}$$

We next introduce pair structures associated with tile structures induced by an expanding Thurston map. We refer the reader to [LSZ25, Subsection 7.2] for details.

Definition 3.12 (Pair structures)**.** Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$. Fix an arbitrary 0-edge $e^0 \in \mathbf{E}^0(f,\mathcal{C})$. For each $n \in \mathbb{N}$, we can pair a white n-tile $X^n_{\mathfrak{w}} \in \mathbf{X}^n_{\mathfrak{w}}$ and a black n-tile $X^n_{\mathfrak{b}} \in \mathbf{X}^n_{\mathfrak{b}}$ whose intersection $X^n_{\mathfrak{w}} \cap X^n_{\mathfrak{b}}$ contains an n-edge contained in $f^{-n}(e^0)$. We define the *set of n-pairs* (with respect to f, $\mathcal{C}$, and e^0), denoted by $\mathbf{P}^n(f,\mathcal{C},e^0)$, to be the collection of the union $X^n_{\mathfrak{w}} \cup X^n_{\mathfrak{b}}$ of such pairs (called *n-pairs*), i.e.,

$$\mathbf{P}^n(f,\mathcal{C},e^0) \coloneqq \bigl\{X^n_{\mathfrak{w}} \cup X^n_{\mathfrak{b}} : X^n_{\mathfrak{w}} \in \mathbf{X}^n_{\mathfrak{w}},\, X^n_{\mathfrak{b}} \in \mathbf{X}^n_{\mathfrak{b}},\, X^n_{\mathfrak{w}} \cap X^n_{\mathfrak{b}} \cap f^{-n}\bigl(e^0\bigr) \in \mathbf{E}^n(f,\mathcal{C})\bigr\}. \tag{3.12}$$

From now on, if the map f, the Jordan curve $\mathcal{C}$, and the 0-edge e^0 are clear from the context, we will sometimes omit $(f,\mathcal{C},e^0)$ in the notation above.

We record [LSZ25, Lemmas 7.6 and 7.9] in the following. Recall that $U^n(x)$ is defined in (3.8).

Lemma 3.13 (Z. Li, X. Shi, Y. Zhang [LSZ25])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying* $\operatorname{post} f \subseteq \mathcal{C}$*. Fix an arbitrary* 0*-edge $e^0 \in \mathbf{E}^0(f,\mathcal{C})$. Then for each $n \in \mathbb{N}$ and any two distinct n-pairs P^n, $\widetilde{P}^n \in \mathbf{P}^n(f,\mathcal{C},e^0)$, their interiors are disjoint.*

Lemma 3.14 (Z. Li, X. Shi, Y. Zhang [LSZ25])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying* $\operatorname{post} f \subseteq \mathcal{C}$*. Let d be a visual metric on S^2 for f. Fix an arbitrary* 0*-edge $e^0 \in \mathbf{E}^0(f,\mathcal{C})$. We assume in addition that $f(\mathcal{C}) \subseteq \mathcal{C}$. Then there exists an integer $M \in \mathbb{N}$ depending only on f, $\mathcal{C}$, d, and e^0 such that for each color $\mathfrak{c} \in \{\mathfrak{b},\mathfrak{w}\}$, there exists an M-pair $P^M_{\mathfrak{c}} \in \mathbf{P}^M(f,\mathcal{C},e^0)$ such that for each integer $n \geqslant M$ and each $x \in P^M_{\mathfrak{c}}$, we have $U^n(x) \subseteq \operatorname{inte}\bigl(X^0_{\mathfrak{c}}\bigr)$.*

3.3. **Subsystems of expanding Thurston maps.** In this subsection, we review some concepts and results on subsystems of expanding Thurston maps. We refer the reader to [LSZ25, LS24] for details.

We first introduce the definition of subsystems along with relevant concepts and notations. Additionally, we provide several examples to illustrate these ideas.

Definition 3.15. Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$. We say that a map $F\colon \operatorname{dom}(F) \to S^2$ is a *subsystem of f with respect to $\mathcal{C}$* if $\operatorname{dom}(F) = \bigcup \mathfrak{X}$ for some non-empty subset $\mathfrak{X} \subseteq \mathbf{X}^1(f,\mathcal{C})$ and $F = f|_{\operatorname{dom}(F)}$. We denote by $\operatorname{Sub}(f,\mathcal{C})$ the set of all subsystems of f with respect to $\mathcal{C}$. Define

$$\operatorname{Sub}_*(f,\mathcal{C}) \coloneqq \{F \in \operatorname{Sub}(f,\mathcal{C}) : \operatorname{dom}(F) \subseteq F(\operatorname{dom}(F))\}.$$

Consider a subsystem $F \in \operatorname{Sub}(f,\mathcal{C})$. For each $n \in \mathbb{N}_0$, we define the *set of n-tiles of F* to be

$$\mathfrak{X}^n(F,\mathcal{C}) \coloneqq \{X^n \in \mathbf{X}^n(f,\mathcal{C}) : X^n \subseteq F^{-n}(F(\operatorname{dom}(F)))\}, \tag{3.13}$$

where we set $F^0 \coloneqq \operatorname{id}_{S^2}$ when $n = 0$. We call each $X^n \in \mathfrak{X}^n(F,\mathcal{C})$ an *n-tile* of F. We define the *tile maximal invariant set* associated with F with respect to $\mathcal{C}$ to be

$$\Omega(F,\mathcal{C}) \coloneqq \bigcap_{n\in\mathbb{N}} \Bigl(\bigcup \mathfrak{X}^n(F,\mathcal{C})\Bigr), \tag{3.14}$$

which is a compact subset of S^2. Indeed, $\Omega(F,\mathcal{C})$ is forward invariant with respect to F, namely, $F(\Omega(F,\mathcal{C})) \subseteq \Omega(F,\mathcal{C})$ (see Proposition 3.17 (i)). We denote by F_Ω the map $F|_{\Omega(F,\mathcal{C})}\colon \Omega(F,\mathcal{C}) \to \Omega(F,\mathcal{C})$.

Let $X^0_{\mathfrak{b}}$, $X^0_{\mathfrak{w}} \in \mathbf{X}^0(f,\mathcal{C})$ be the black 0-tile and the white 0-tile, respectively. For each $n \in \mathbb{N}_0$, we define the *set of black n-tiles of F* as

$$\mathfrak{X}^n_{\mathfrak{b}}(F,\mathcal{C}) := \big\{X \in \mathfrak{X}^n(F,\mathcal{C}) : F^n(X) = X^0_{\mathfrak{b}}\big\},$$

and the *set of white n-tiles of F* as

$$\mathfrak{X}^n_{\mathfrak{w}}(F,\mathcal{C}) := \big\{X \in \mathfrak{X}^n(F,\mathcal{C}) : F^n(X) = X^0_{\mathfrak{w}}\big\}.$$

Moreover, for each $n \in \mathbb{N}_0$ and each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b},\mathfrak{w}\}$ we define

$$\mathfrak{X}^n_{\mathfrak{c}\mathfrak{c}'}(F,\mathcal{C}) := \big\{X \in \mathfrak{X}^n_{\mathfrak{c}}(F,\mathcal{C}) : X \subseteq X^0_{\mathfrak{c}'}\big\}.$$

In other words, for example, a tile $X \in \mathfrak{X}^n_{\mathfrak{b}\mathfrak{w}}(F,\mathcal{C})$ is a *black n-tile of F contained in $X^0_{\mathfrak{w}}$*, i.e., an n-tile of F that is contained in the white 0-tile $X^0_{\mathfrak{w}}$ as a set, and is mapped by F^n onto the black 0-tile $X^0_{\mathfrak{b}}$.

By abuse of notation, we often omit $(F,\mathcal{C})$ in the notations above when it is clear from the context.

We discuss two examples below and refer the reader to [LSZ25, Subsection 5.1] for more examples.

Example 3.16. Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$. Consider $F \in \operatorname{Sub}(f,\mathcal{C})$.

(i) The map $F\colon \operatorname{dom}(F) \to S^2$ is represented by Figure 3.1. Here S^2 is identified with a pillow that is obtained by gluing two squares together along their boundaries. Moreover, each square is subdivided into 3×3 subsquares, and $\operatorname{dom}(F)$ is obtained from S^2 by removing the interior of the middle subsquare $X^1_{\mathfrak{w}} \in \mathbf{X}^1_{\mathfrak{w}}(f,\mathcal{C})$ and $X^1_{\mathfrak{b}} \in \mathbf{X}^1_{\mathfrak{b}}(f,\mathcal{C})$ of the respective squares. In this case, Ω is a Sierpiński carpet. It consists of two copies of the standard square Sierpiński carpet glued together along the boundaries of the squares.

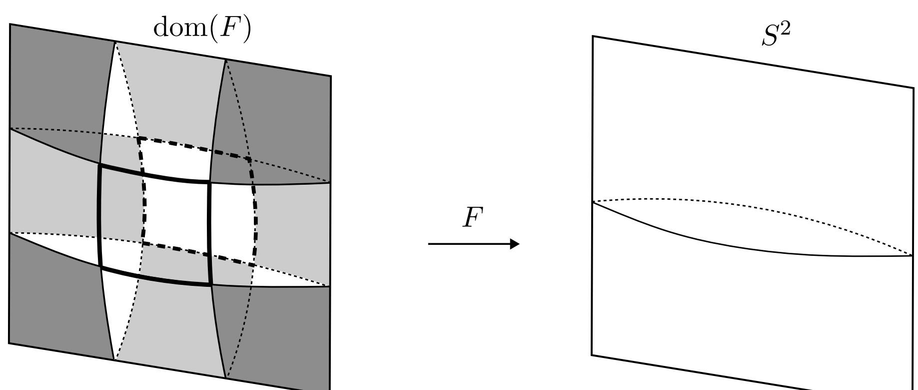

FIGURE 3.1. A Sierpiński carpet subsystem.

(ii) The map $F\colon \operatorname{dom}(F) \to S^2$ is represented by Figure 3.2. Here S^2 is identified with a pillow that is obtained by gluing two equilateral triangles together along their boundaries. Moreover, each triangle is subdivided into 4 small equilateral triangles, and $\operatorname{dom}(F)$ is obtained from S^2 by removing the interior of the middle small triangle $X^1_{\mathfrak{b}} \in \mathbf{X}^1_{\mathfrak{b}}(f,\mathcal{C})$ and $X^1_{\mathfrak{w}} \in \mathbf{X}^1_{\mathfrak{w}}(f,\mathcal{C})$ of the respective triangle. In this case, Ω is a Sierpiński gasket. It consists of two copies of the standard Sierpiński gasket glued together along the boundaries of the triangles.

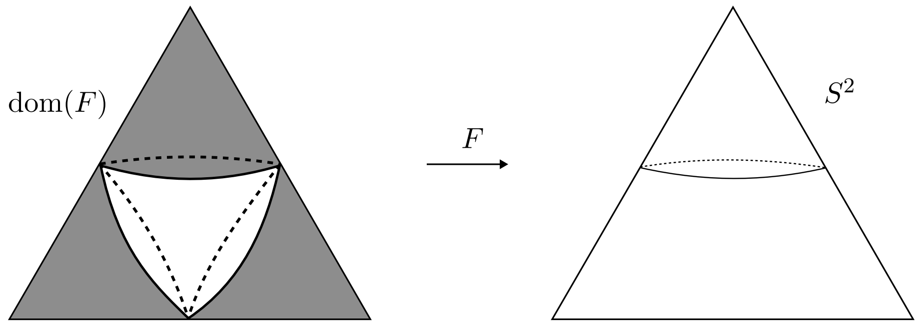

FIGURE 3.2. A Sierpiński gasket subsystem.

We summarize some preliminary results for subsystems in the following proposition.

Proposition 3.17 (Z. Li, X. Shi, Y. Zhang [LSZ25])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying* $\operatorname{post} f \subseteq \mathcal{C}$*. Consider $F \in \operatorname{Sub}(f, \mathcal{C})$. Then the following statements hold:*

(i) *The tile maximal invariant set Ω is forward invariant with respect to F, i.e., $F(\Omega) \subseteq \Omega$.*
(ii) *If $f(\mathcal{C}) \subseteq \mathcal{C}$, then $\mathfrak{X}^m_{\mathfrak{c}}(F, \mathcal{C}) = \mathfrak{X}^m_{\mathfrak{c}\mathfrak{b}}(F, \mathcal{C}) \cup \mathfrak{X}^m_{\mathfrak{c}\mathfrak{w}}(F, \mathcal{C})$ for each $m \in \mathbb{N}_0$ and each $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$.*
(iii) *If $f(\mathcal{C}) \subseteq \mathcal{C}$, then $F^{-1}(\Omega \setminus \mathcal{C}) \subseteq \Omega \setminus \mathcal{C}$.*

Proposition 3.17 (i) was proved in [LSZ25, Proposition 5.4 (iii)]. Proposition 3.17 (ii) and (iii) are from [LSZ25, Proposition 5.5 (ii) and (iii)].

We introduce a 2×2 matrix called the tile matrix to describe tiles of a subsystem according to their colors and locations.

Definition 3.18 (Tile matrices)**.** Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$. Consider $F \in \operatorname{Sub}(f, \mathcal{C})$. We define the *tile matrix* of F with respect to $\mathcal{C}$ as

$$A = A(F, \mathcal{C}) \coloneqq \begin{bmatrix} N_{\mathfrak{w}\mathfrak{w}} & N_{\mathfrak{b}\mathfrak{w}} \\ N_{\mathfrak{w}\mathfrak{b}} & N_{\mathfrak{b}\mathfrak{b}} \end{bmatrix} \tag{3.15}$$

where

$$N_{\mathfrak{c}\mathfrak{c}'} = N_{\mathfrak{c}\mathfrak{c}'}(A) \coloneqq \operatorname{card}\bigl\{X \in \mathfrak{X}^1_{\mathfrak{c}}(F, \mathcal{C}) : X \subseteq X^0_{\mathfrak{c}'}\bigr\} = \operatorname{card}\bigl(\mathfrak{X}^1_{\mathfrak{c}\mathfrak{c}'}(F, \mathcal{C})\bigr)$$

for each pair of colors $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$. For example, $N_{\mathfrak{b}\mathfrak{w}}$ is the number of black tiles in $\mathfrak{X}^1(F, \mathcal{C})$ that are contained in the white 0-tile $X^0_{\mathfrak{w}}$. Recall that $X^0_{\mathfrak{b}}$, $X^0_{\mathfrak{w}} \in \mathbf{X}^0(f, \mathcal{C})$ is the black 0-tile and the white 0-tile, respectively.

Remark 3.19. Note that the tile matrix $A(F, \mathcal{C})$ of F with respect to $\mathcal{C}$ is completely determined by the set $\mathfrak{X}^1(F, \mathcal{C})$. Thus for each integer $n \in \mathbb{N}_0$ and each set of n-tiles $\mathbf{T} \subseteq \mathbf{X}^n(f, \mathcal{C})$, similarly, we can define the tile matrix $A(\mathbf{T})$ of $\mathbf{T}$ as

$$A(\mathbf{T}) \coloneqq \begin{bmatrix} N_{\mathfrak{w}\mathfrak{w}}(\mathbf{T}) & N_{\mathfrak{b}\mathfrak{w}}(\mathbf{T}) \\ N_{\mathfrak{w}\mathfrak{b}}(\mathbf{T}) & N_{\mathfrak{b}\mathfrak{b}}(\mathbf{T}), \end{bmatrix}$$

where $N_{\mathfrak{c}\mathfrak{c}'}(\mathbf{T}) \coloneqq \operatorname{card}\bigl(\bigl\{X \in \mathbf{T} : X \in \mathbf{X}^n_{\mathfrak{c}}(f, \mathcal{C}),\, X \subseteq X^0_{\mathfrak{c}'}\bigr\}\bigr) = \operatorname{card}\bigl(\mathfrak{X}^n_{\mathfrak{c}\mathfrak{c}'}(F, \mathcal{C})\bigr)$ for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$.

Definition 3.20 (Primitivity)**.** Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$. Consider $F \in \operatorname{Sub}(f, \mathcal{C})$. We say that F is a *primitive* (resp. *strongly primitive*) subsystem (of f with respect to $\mathcal{C}$) if there exists an integer $n_F \in \mathbb{N}$ such that for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$ and each integer $n \geqslant n_F$, there exists $X^n \in \mathfrak{X}^n_{\mathfrak{c}}(F, \mathcal{C})$ satisfying $X^n \subseteq X^0_{\mathfrak{c}'}$ (resp. $X^n \subseteq \operatorname{inte}\bigl(X^0_{\mathfrak{c}'}\bigr)$).

Remark 3.21. By [Li18, Lemma 5.10], every expanding Thurston map f is a strongly primitive subsystem of itself with respect to every Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$.

We record [LSZ25, Lemmas 5.22] below, which shows that primitive subsystems have nice combinatorial and topological properties.

Lemma 3.22 (Z. Li, X. Shi, Y. Zhang [LSZ25])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying* $\operatorname{post} f \subseteq \mathcal{C}$*. Let $F \in \operatorname{Sub}(f, \mathcal{C})$ be primitive (resp. strongly primitive). Let $n_F \in \mathbb{N}$ be the constant from Definition 3.20, which depends only on F and $\mathcal{C}$. Then for each $n \in \mathbb{N}$ with $n \geqslant n_F$, each $m \in \mathbb{N}_0$, each $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, and each m-tile $X^m \in \mathfrak{X}^m(F, \mathcal{C})$, there exists an $(n+m)$-tile $X^{n+m}_{\mathfrak{c}} \in \mathfrak{X}^{n+m}_{\mathfrak{c}}(F, \mathcal{C})$ such that $X^{n+m}_{\mathfrak{c}} \subseteq X^m$ (resp. $X^{n+m}_{\mathfrak{c}} \subseteq \operatorname{inte}(X^m)$).*

We now review some concepts and results on the ergodic theory of subsystems of expanding Thurston maps. We refer the reader to [LSZ25, Section 6] for more details and the proofs.

Definition 3.23 (Topological pressures for subsystems)**.** Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$. Consider $F \in \operatorname{Sub}(f, \mathcal{C})$. For a real-valued function $\varphi\colon S^2 \to \mathbb{R}$, we denote

$$Z_n(F, \varphi) \coloneqq \sum_{X^n \in \mathfrak{X}^n(F, \mathcal{C})} \exp\bigl(\sup\bigl\{S_n^F \varphi(x) : x \in X^n\bigr\}\bigr)$$

for each $n \in \mathbb{N}$. We define the *topological pressure* of F with respect to the *potential* φ by

$$P(F, \varphi) \coloneqq \liminf_{n \to +\infty} \frac{1}{n} \log(Z_n(F, \varphi)). \tag{3.16}$$

In particular, when φ is the constant function 0, the quantity $h_{\mathrm{top}}(F) \coloneqq P(F, 0)$ is called the *topological entropy* of F.

Remark. We note that the definition (3.16) differs from the classical definition in (3.1) presented in Subsection 3.1, which applies to $F|_\Omega$ and $\varphi|_\Omega$. Indeed, they coincide for a strongly primitive subsystem and a Hölder continuous potential (see [LSZ25, Theorems 6.29 and 6.30]).

We record [LSZ25, Proposition 6.5] below, which shows that the topological entropy of F can in fact be computed explicitly via tile matrices defined in Definition 3.18.

Proposition 3.24 (Z. Li, X. Shi, Y. Zhang [LSZ25])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying* $\operatorname{post} f \subseteq \mathcal{C}$ *and $f(\mathcal{C}) \subseteq \mathcal{C}$. Consider a subsystem $F \in \operatorname{Sub}(f, \mathcal{C})$. Let A be the tile matrix of F with respect to $\mathcal{C}$. Then we have*

$$h_{\mathrm{top}}(F) = \log(\rho(A)), \tag{3.17}$$

where $\rho(A)$ is the spectral radius of A.

Remark. The spectral radius $\rho(A)$ can easily be computed from any matrix norm. If for an $(m \times m)$-matrix $B = (b_{ij})$ we set $\|B\| \coloneqq \sum_{i,j=1}^m |b_{ij}|$ for example, then $\rho(A) = \lim_{n \to +\infty} (\|A^n\|)^{1/n}$.

We summarize the existence and some basic properties of equilibrium states for strongly primitive subsystems in the following theorem, which is part of [LS24, Theorem 1.1]. Recall that $F(\Omega) \subseteq \Omega$ by Proposition 3.17 (i).

Theorem 3.25 (Z. Li & X. Shi [LS24])**.** *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying* $\operatorname{post} f \subseteq \mathcal{C}$ *and $f(\mathcal{C}) \subseteq \mathcal{C}$. Let d be a visual metric on S^2 for f, and ϕ be a real-valued Hölder continuous function on S^2 with respect to the metric d. Consider a strongly primitive subsystem $F \in \operatorname{Sub}(f, \mathcal{C})$. Then there exists a unique equilibrium state $\mu_{F,\phi}$ for $F|_\Omega$ and $\phi|_\Omega$. Moreover, $\mu_{F,\phi}$ is ergodic for $F|_\Omega$.*

4. The Assumptions

We state below the hypotheses under which we will develop our theory in most parts of this paper. We will selectively use some of the following assumptions in the remaining part of this paper.

The Assumptions.

(1) $f\colon S^2 \to S^2$ is an expanding Thurston map.

(2) $\mathcal{C} \subseteq S^2$ is a Jordan curve containing $\operatorname{post} f$ with the property that there exists an integer $n_{\mathcal{C}} \in \mathbb{N}$ such that $f^{n_{\mathcal{C}}}(\mathcal{C}) \subseteq \mathcal{C}$ and $f^m(\mathcal{C}) \nsubseteq \mathcal{C}$ for each $m \in \{1, \dots, n_{\mathcal{C}} - 1\}$.
(3) $F \in \operatorname{Sub}(f, \mathcal{C})$ is a subsystem of f with respect to $\mathcal{C}$.
(4) d is a visual metric on S^2 for f with expansion factor $\Lambda > 1$.
(5) $\beta \in (0, 1]$.
(6) $\phi \in C^{0,\beta}(S^2, d)$ is a real-valued Hölder continuous function with exponent β.
(7) μ_ϕ is the unique equilibrium state for the map f and the potential ϕ.
(8) $e^0 \in \mathbf{E}^0(f, \mathcal{C})$ is a 0-edge.

Observe that by Lemma 3.7, for each f in (1), there exists at least one Jordan curve $\mathcal{C}$ that satisfies (2). Since for a fixed f, the number $n_{\mathcal{C}}$ is uniquely determined by $\mathcal{C}$ in (2), in this paper, we say that a quantity depends on $\mathcal{C}$ even if it also depends on $n_{\mathcal{C}}$.

Recall that the expansion factor Λ of a visual metric d on S^2 for f is uniquely determined by d and f. We will say that a quantity depends on f and d if it depends on Λ.

In the discussion below, depending on the conditions we will need, we will sometimes say "Let f, $\mathcal{C}$, d, ϕ satisfy the Assumptions.", and sometimes say "Let f and $\mathcal{C}$ satisfy the Assumptions.", etc.

5. Upper semi-continuity

In this section we show that the entropy map of an expanding Thurston map is upper semi-continuous if and only if the map has no periodic critical points.

Definition 5.1. Let X be a compact metrizable topological space and $T\colon X \to X$ be a continuous map. The *entropy map* of T is the map $\mu \mapsto h_\mu(T)$ which is defined on $\mathcal{M}(X,T)$ and has values in $[0, +\infty]$. Here $\mathcal{M}(X,T)$ is the set of all T-invariant Borel probability measures on X and is equipped with the weak* topology. We say that the entropy map of T is *upper semi-continuous* if $\limsup_{n\to+\infty} h_{\mu_n}(T) \leqslant h_\mu(T)$ holds for every sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of Borel probability measures on X that converges to $\mu \in \mathcal{M}(X,T)$ in the weak* topology.

The proof of the following lemma is straightforward, and we include it for the sake of completeness.

Lemma 5.2. *Let X be a compact metrizable topological space and $T\colon X \to X$ be a continuous map. Consider arbitrary $n \in \mathbb{N}$ and $\mu \in \mathcal{M}(X, T^n)$. Define $\nu \coloneqq \frac{1}{n}\sum_{j=0}^{n-1} T_*^j \mu$. Then $\nu \in \mathcal{M}(X,T)$ and*

$$h_{T_*^j\mu}(T^n) = h_\mu(T^n) = h_\nu(T^n) = n h_\nu(T) \qquad \text{for each } j \in \{0, 1, \dots, n-1\}. \tag{5.1}$$

Moreover, if μ is ergodic for T^n, then ν is ergodic for T.

Proof. Fix arbitrary $n \in \mathbb{N}$ and $\mu \in \mathcal{M}(X, T^n)$. Then $T_*\nu = \nu \in \mathcal{M}(X,T)$ since $T_*^n\mu = \mu$. By (3.2) and (3.3), we have

$$n h_\nu(T) = h_\nu(T^n) = \frac{1}{n}\sum_{j=0}^{n-1} h_{T_*^j\mu}(T^n). \tag{5.2}$$

We now show that $h_{T_*^j\mu}(T^n) = h_\mu(T^n)$ for each $j \in \{0, \dots, n-1\}$. Indeed, the measure $T_*\mu$ is T^n-invariant and the triple $(X, T^n, T_*\mu)$ is a factor of (X, T^n, μ) by the map T. It follows that $h_{T_*\mu}(T^n) \leqslant h_\mu(T^n)$ (see Subsection 3.1). Iterating this and noting that $T_*^n\mu = \mu$ by T^n-invariance of μ, we obtain

$$h_\mu(T^n) = h_{T_*^n\mu}(T^n) \leqslant h_{T_*^{n-1}\mu}(T^n) \leqslant \cdots \leqslant h_{T_*\mu}(T^n) \leqslant h_\mu(T^n).$$

Hence $h_{T_*^j\mu}(T^n) = h_\mu(T^n)$ for each $j \in \{0, \dots, n-1\}$. Combining this with (5.2), we establish (5.1).

Finally, we assume that μ is ergodic for T^n. Let A be a Borel subset of X satisfying $T^{-1}(A) = A$. Since $T^{-n}(A) = A$ and μ is ergodic for T^n, we have $\mu(A) \in \{0, 1\}$. Then

$$\nu(A) = \frac{1}{n}\sum_{i=0}^{n-1} T_*^j\mu = \frac{1}{n}\sum_{i=0}^{n-1} \mu\bigl(T^{-i}(A)\bigr) = \mu(A) \in \{0, 1\}.$$

This implies that ν is ergodic for T. $\square$

The following proposition is a consequence of Lemma 5.2.

Proposition 5.3. *Let X be a compact metrizable topological space and $T\colon X \to X$ be a continuous map. Consider arbitrary $n \in \mathbb{N}$. Then the entropy map of T^n is upper semi-continuous if and only if the entropy map of T is upper semi-continuous.*

Proof. Fix arbitrary $n \in \mathbb{N}$.

Suppose that the entropy map of T^n is upper semi-continuous. Since $\mathcal{M}(X,T) \subseteq \mathcal{M}(X,T^n)$, it follows immediately from (3.2) that the entropy map of T is also upper semi-continuous.

For the converse direction suppose that the entropy map of T^n is not upper semi-continuous. Then there exists a T^n-invariant Borel probability measure $\mu_0 \in \mathcal{M}(X,T^n)$ and a sequence $\{\mu_k\}_{k\in\mathbb{N}}$ of T^n-invariant Borel probability measures in $\mathcal{M}(X,T^n)$ such that $\{\mu_k\}_{k\in\mathbb{N}}$ converges to μ_0 in the weak* topology and satisfies

$$\limsup_{k\to+\infty} h_{\mu_k}(T^n) > h_{\mu_0}(T^n). \tag{5.3}$$

We define $\nu_0 \coloneqq \frac{1}{n}\sum_{j=0}^{n-1} T_*^j\mu_0$ and $\nu_k \coloneqq \frac{1}{n}\sum_{j=0}^{n-1} T_*^j\mu_k$ for each $k \in \mathbb{N}$. Then $\{\nu_k\}_{k\in\mathbb{N}}$ converges to ν_0 in the weak* topology. Indeed, for each $\varphi \in C(X)$, since $S_n^T\varphi \in C(X)$ and $\{\mu_k\}_{k\in\mathbb{N}}$ converges to μ_0 in the weak* topology, we obtain

$$\int \varphi \,\mathrm{d}\nu_k = \frac{1}{n}\int S_n^T\varphi \,\mathrm{d}\mu_k \longrightarrow \frac{1}{n}\int S_n^T\varphi \,\mathrm{d}\mu_0 = \int \varphi \,\mathrm{d}\nu_0$$

as $k \to +\infty$.

Now we show that the entropy map of T is not upper semi-continuous at ν_0. It follows immediately from (5.3) and Lemma 5.2 that

$$\limsup_{k\to+\infty} h_{\nu_k}(T) = \frac{1}{n}\limsup_{k\to+\infty} h_{\mu_k}(T^n) > \frac{1}{n}h_{\mu_0}(T^n) = h_{\nu_0}(T).$$

This completes the proof. $\square$

By constructing suitable subsystems, we can prove the "only if" part in Theorem 1.1. We first establish the following proposition and then prove the general cases (Theorem 5.5).

Proposition 5.4. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map with a Jordan curve $\mathcal{C} \subseteq S^2$ satisfying $\operatorname{post} f \subseteq \mathcal{C}$ and $f(\mathcal{C}) \subseteq \mathcal{C}$. Suppose that f has a fixed critical point p. Then there exists a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of ergodic f-invariant Borel probability measures in $\mathcal{M}(S^2,f)$ such that $\{\mu_n\}_{n\in\mathbb{N}}$ converges to δ_p in the weak* topology and satisfies*

$$\lim_{n\to+\infty} h_{\mu_n}(f) = \log\bigl(\deg_f(p)\bigr) > 0 = h_{\delta_p}(f). \tag{5.4}$$

In particular, the entropy map of f is not upper semi-continuous at δ_p.

Proof. Let $p \in S^2$ be a critical point of f that is fixed by f. Set $k \coloneqq \deg_f(p)$. Then $k > 1$. Note that $\delta_p \in \mathcal{M}(S^2,f)$ and $h_{\delta_p}(f) = 0$, where δ_p is the Dirac measure supported on $\{p\}$. It suffices to construct a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of ergodic f-invariant Borel probability measures in $\mathcal{M}(S^2,f)$ such that $\{\mu_n\}_{n\in\mathbb{N}}$ converges to δ_p in the weak* topology and satisfies (5.4).

We first give the construction of $\{\mu_n\}_{n\in\mathbb{N}}$.

Fix arbitrary $n \in \mathbb{N}$. The set of n-tiles of f at p is defined as $\mathfrak{X}^n(f,\mathcal{C},p) \coloneqq \{X \in \mathbf{X}^n(f,\mathcal{C}) : p \in X\}$. Then by Remark 3.4, we have $\overline{W}^n(p) = \bigcup \mathfrak{X}^n(f,\mathcal{C},p)$ and $\operatorname{card}(\mathfrak{X}^n(f,\mathcal{C},p)) = 2(\deg_f(p))^n = 2k^n$, where $W^n(p)$ is defined in (3.6) and $\overline{W}^n(p)$ is the closure of $W^n(p)$.

Since $f \in \operatorname{Sub}(f,\mathcal{C})$ is strongly primitive, by Lemma 3.22, there exists an integer $n_f \in \mathbb{N}$ depending only on f and $\mathcal{C}$ such that for each n-tile $X^n \in \mathbf{X}^n(f,\mathcal{C})$, there exists a black $(n+n_f)$-tile $X_{\mathfrak{b}}^{n+n_f} \in \mathbf{X}_{\mathfrak{b}}^{n+n_f}(f,\mathcal{C})$ and a white $(n+n_f)$-tile $X_{\mathfrak{w}}^{n+n_f} \in \mathbf{X}_{\mathfrak{w}}^{n+n_f}(f,\mathcal{C})$ such that $X_{\mathfrak{b}}^{n+n_f} \cup X_{\mathfrak{w}}^{n+n_f} \subseteq \operatorname{inte}(X^n)$. We denote by $\mathbf{E}_n$ the set consisting of two such $(n+n_f)$-tiles, one black and one white, for each n-tile $X^n \in \mathfrak{X}^n(f,\mathcal{C},p)$. In particular, we have $\operatorname{card}(\mathbf{E}_n) = 2\operatorname{card}(\mathfrak{X}^n(f,\mathcal{C},p)) = 4k^n$.

We set $F_n \coloneqq f^{n+n_f}|_{\bigcup \mathbf{E}_n}$ and $\widehat{F}_n \coloneqq F_n|_{\Omega_n}$, where $\Omega_n \coloneqq \Omega(F_n, \mathcal{C})$ is the tile maximal invariant set associated with F_n with respect to $\mathcal{C}$. Note that $p \in \operatorname{post} f \subseteq \mathcal{C}$. By Remark 3.4 and Proposition 3.8, $F_n \in \operatorname{Sub}(f^{n+n_f}, \mathcal{C})$ is strongly primitive. Then it follows from Theorem 3.25 and [LSZ25, Theorem 1.1] that there exists $\widehat{\mu}_n \in \mathcal{M}(\Omega_n, \widehat{F}_n) \subseteq \mathcal{M}(S^2, f^{n+n_f})$ such that $\operatorname{supp} \widehat{\mu}_n \subseteq \Omega_n \subseteq \bigcup \mathbf{E}_n \subseteq \bigcup \mathfrak{X}^n(f, \mathcal{C}, p) = \overline{W}^n(p)$ and

$$h_{\widehat{\mu}_n}(f^{n+n_f}) = h_{\widehat{\mu}_n}(\widehat{F}_n) = P(F_n, 0) = h_{\operatorname{top}}(F_n),$$

where $P(F_n, 0)$ and $h_{\operatorname{top}}(F_n)$ are defined in Definition 3.23. Put

$$\mu_n \coloneqq \frac{1}{n+n_f} \sum_{j=0}^{n+n_f-1} f_*^j \widehat{\mu}_n.$$

Applying Lemma 5.2, we have $\mu_n \in \mathcal{M}(S^2, f)$ and $(n+n_f) h_{\mu_n}(f) = h_{\widehat{\mu}_n}(f^{n+n_f}) = h_{\operatorname{top}}(F_n)$.

Now we calculate $h_{\mu_n}(f)$ for each $n \in \mathbb{N}$ and show that (5.4) holds.

By Definition 3.23 and Proposition 3.24, we have $h_{\operatorname{top}}(F_n) = \log(\rho(A_n))$, where A_n is the tile matrix of F_n with respect to $\mathcal{C}$ and $\rho(A_n)$ is the spectral radius of A_n. Recall from Definition 3.18 and Remark 3.19 that

$$A_n = A(\mathbf{E}_n) = \begin{bmatrix} N_{\mathfrak{w}\mathfrak{w}} & N_{\mathfrak{b}\mathfrak{w}} \\ N_{\mathfrak{w}\mathfrak{b}} & N_{\mathfrak{b}\mathfrak{b}} \end{bmatrix},$$

where $N_{\mathfrak{c}\mathfrak{c}'} \coloneqq \operatorname{card}\bigl(\bigl\{X \in \mathbf{E}_n : X \in \mathbf{X}_{\mathfrak{c}}^{n+n_f}(f, \mathcal{C}),\, X \subseteq X_{\mathfrak{c}'}^0\bigr\}\bigr)$ for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$. In particular, since $f(\mathcal{C}) \subseteq \mathcal{C}$, by the construction of $\mathbf{E}_n$ and Proposition 3.17 (ii), one has $N_{\mathfrak{b}\mathfrak{c}'} = N_{\mathfrak{w}\mathfrak{c}'}$ and $N_{\mathfrak{c}\mathfrak{b}} + N_{\mathfrak{c}\mathfrak{w}} = \operatorname{card}(\mathbf{E}_n)/2 = 2k^n$ for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$. Then it follows that $\rho(A_n) = 2k^n$. Hence $h_{\operatorname{top}}(F_n) = \log(2k^n)$ and (5.4) holds since

$$h_{\mu_n}(f) = \frac{h_{\operatorname{top}}(F_n)}{n+n_f} = \frac{\log(2k^n)}{n+n_f} \longrightarrow \log k \qquad \text{as } n \to +\infty.$$

Finally, we show that $\{\mu_n\}_{n\in\mathbb{N}}$ converges to δ_p in the weak* topology. It suffices to show that for each $\varphi \in C(S^2)$, $\int \varphi \,\mathrm{d}\mu_n \to \varphi(p)$ as $n \to +\infty$.

We fix a visual metric d that satisfies the Assumptions in Section 4.

Fix arbitrary $\varphi \in C(S^2)$ and $\varepsilon > 0$. Since φ is continuous at p, there exists a number $\delta > 0$ such that for each $x \in S^2$ with $d(x,p) < \delta$, we have $|\varphi(x) - \varphi(p)| < \varepsilon$. By Proposition 3.6 (ii), there exists an integer $N \in \mathbb{N}$ such that for each integer $\ell > N$, we have $\overline{W}^\ell(p) \subseteq B_d(p, \delta)$. For each $n \in \mathbb{N}$ and each $j \in \{0, 1, \dots, n-1\}$, it follows from $\operatorname{supp} \widehat{\mu}_n \subseteq \overline{W}^n(p)$ and (3.7) in Remark 3.4 that $\operatorname{supp} f_*^j \widehat{\mu}_n \subseteq f^j\bigl(\overline{W}^n(p)\bigr) = \overline{W}^{n-j}(p)$. Thus for sufficiently large n, we have $\overline{W}^{n-j}(p) \subseteq B_d(p, \delta)$ for all $j \in \{0, 1, \dots, n-N-1\}$. Then

$$\begin{aligned} \biggl|\varphi(p) - \int \varphi \,\mathrm{d}\mu_n\biggr| &= \biggl|\varphi(p) - \frac{1}{n+n_f} \sum_{j=0}^{n+n_f} \int \varphi \,\mathrm{d} f_*^j \widehat{\mu}_n\biggr| \\ &\leqslant \biggl|\varphi(p) - \frac{1}{n+n_f} \sum_{j=0}^{n-N-1} \int \varphi \,\mathrm{d} f_*^j \widehat{\mu}_n\biggr| + \frac{N+n_f}{n+n_f} \|\varphi\|_\infty \\ &\leqslant \frac{1}{n+n_f} \sum_{j=0}^{n-N-1} \int |\varphi - \varphi(p)| \,\mathrm{d} f_*^j \widehat{\mu}_n + \frac{2(N+n_f)}{n+n_f} \|\varphi\|_\infty \\ &\leqslant \frac{n-N}{n+n_f} \varepsilon + \frac{2(N+n_f)}{n+n_f} \|\varphi\|_\infty \\ &\leqslant 2\varepsilon \end{aligned}$$

for sufficiently large n. This implies $\int \varphi \,\mathrm{d}\mu_n \to \varphi(p)$ as $n \to +\infty$ for each $\varphi \in C(S^2)$.

The proof is complete. □

Based on Proposition 5.4, we now prove the general cases by applying Proposition 5.3.

Recall that a point $x \in S^2$ is a periodic point of $f\colon S^2 \to S^2$ with period $n \in \mathbb{N}$ if $f^n(x) = x$ and $f^i(x) \neq x$ for each $i \in \{1, 2, \dots, n-1\}$.

Theorem 5.5. *Let $f\colon S^2 \to S^2$ be an expanding Thurston map. Suppose that f has a periodic critical point p with period n for some $n \in \mathbb{N}$. Denote $V_n(p) \coloneqq \frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(p)}$. Then there exists a sequence $\{\nu_k\}_{k\in\mathbb{N}}$ of ergodic f-invariant Borel probability measures in $\mathcal{M}(S^2, f)$ such that $\{\nu_k\}_{k\in\mathbb{N}}$ converges to $V_n(p)$ in the weak* topology and satisfies*

$$\lim_{k\to+\infty} h_{\nu_k}(f) = \frac{1}{n}\log\bigl(\deg_{f^n}(p)\bigr) > 0 = h_{V_n(p)}(f). \tag{5.5}$$

In particular, the entropy map of f is not upper semi-continuous at $V_n(p)$.

Proof. Suppose that $p \in S^2$ is a periodic critical point of f with period n for some $n \in \mathbb{N}$.

By Lemma 3.7, we can find a sufficiently high iterate f^m of f that has an f^m-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} f^{nm} = \operatorname{post} f \subseteq \mathcal{C}$. We fix such Jordan curve $\mathcal{C}$ and set $F \coloneqq f^N$ with $N \coloneqq nm$. Thus $F(\mathcal{C}) \subseteq \mathcal{C}$ and p is a fixed critical point of F with $\deg_F(p) = \bigl(\deg_{f^n}(p)\bigr)^m$. Note that F is also an expanding Thurston map.

By Proposition 5.4, there exists a sequence $\{\mu_k\}_{k\in\mathbb{N}}$ of ergodic F-invariant Borel probability measures in $\mathcal{M}(S^2, F)$ such that $\{\mu_k\}_{k\in\mathbb{N}}$ converges to δ_p in the weak* topology and satisfies

$$\lim_{k\to+\infty} h_{\mu_k}(F) = \log\bigl(\deg_F(p)\bigr) > h_{\delta_p}(F) = 0. \tag{5.6}$$

We define $\nu_k \coloneqq \frac{1}{N}\sum_{j=0}^{N-1} f^j_*\mu_k$ for each $k \in \mathbb{N}$. It follows immediately from Lemma 5.2 that $\nu_k \in \mathcal{M}(S^2, f)$ and ν_k is ergodic for f. Note that $\{\nu_k\}_{k\in\mathbb{N}}$ converges to $V_n(p)$ in the weak* topology. Indeed, for each $\varphi \in C(X)$, since $S^f_N\varphi \in C(X)$ and $\{\mu_k\}_{k\in\mathbb{N}}$ converges to δ_p in the weak* topology, we have

$$\int \varphi \,\mathrm{d}\nu_k = \frac{1}{N}\int S^f_N\varphi \,\mathrm{d}\mu_k \longrightarrow \frac{1}{N}\int S^f_N\varphi \,\mathrm{d}\delta_p = \int \varphi \,\mathrm{d}V_n(p) \qquad \text{as } k \to +\infty.$$

Finally, by (5.6) and Lemma 5.2 we have

$$\lim_{k\to+\infty} h_{\nu_k}(f) = \frac{1}{N}\lim_{k\to+\infty} h_{\mu_k}(F) = \frac{1}{N}\log\bigl(\deg_F(p)\bigr) = \frac{1}{n}\log\bigl(\deg_{f^n}(p)\bigr) > \frac{h_{\delta_p}(F)}{N} = h_{V_n(p)}(f) = 0.$$

This completes the proof. □

Proof of Theorem 1.1. By [Li15, Corollary 1.3], if f has no periodic critical points, then the entropy map of f is upper semi-continuous. The opposite direction follows immediately from Theorem 5.5. □

6. Entropy density

This section is devoted to the proof of entropy density of ergodic measures for expanding Thurston maps, with the main result being Theorem 1.2.

We first introduce some notations.

Notations. For $\ell \in \mathbb{N}$ we define

$$C(S^2)^\ell \coloneqq \{\vec{\varphi} = (\varphi_1, \dots, \varphi_\ell) : \varphi_j \in C(S^2) \text{ for each } j \in \{1, \dots, \ell\}\}.$$

For $\vec{\varphi} = (\varphi_1, \dots, \varphi_\ell) \in C(S^2)^\ell$, $\vec{\alpha} = (\alpha_1, \dots, \alpha_\ell) \in \mathbb{R}^\ell$, and $\mu \in \mathcal{P}(S^2)$, the expression $\int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha}$ indicates that $\int \varphi_j \,\mathrm{d}\mu > \alpha_j$ holds for each $j \in \{1, \dots, \ell\}$. The meaning of $\int \vec{\varphi}\,\mathrm{d}\mu \geqslant \vec{\alpha}$ is analogous. We write $S_n\vec{\varphi} \coloneqq (S_n\varphi_1, \dots, S_n\varphi_\ell) \in C(S^2)^\ell$. Put $\|\vec{\alpha}\| \coloneqq \max_{1\leqslant j\leqslant \ell}|\alpha_j|$ and $\|\vec{\varphi}\| \coloneqq \max_{1\leqslant j\leqslant\ell}\|\varphi_j\|_\infty$. For $\varepsilon \in \mathbb{R}$ we use the convention that $\vec{\alpha} + \varepsilon \coloneqq (\alpha_1 + \varepsilon, \dots, \alpha_\ell + \varepsilon) \in \mathbb{R}^\ell$.

Let $f\colon S^2 \to S^2$ be an expanding Thurston map and $\mathcal{C} \subseteq S^2$ be a Jordan curve containing $\operatorname{post} f$. For a real-valued function $\psi : S^2 \to \mathbb{R}$ and an integer $n \in \mathbb{N}$ define

$$D_n(\psi) = D_n^{f,\mathcal{C}}(\psi) \coloneqq \sup_{X^n \in \mathbf{X}^n(f,\mathcal{C})}\ \sup_{x,\,y\in X^n} |S_n\psi(x) - S_n\psi(y)|. \tag{6.1}$$

Note that $D_n(\psi) \leqslant nD_1(\psi)$ holds for every $n \in \mathbb{N}$. For $\ell \in \mathbb{N}$ and $\vec{\varphi} \in C(S^2)^\ell$, we write $D_n(\vec{\varphi}) \coloneqq \max_{1\leqslant j\leqslant \ell} D_n(\varphi_j)$ for each $n \in \mathbb{N}$.

Indeed, for $\varphi \in C(S^2)$ we have $\lim_{n\to+\infty} D_n(\varphi)/n = 0$.

Lemma 6.1. *Let f and $\mathcal{C}$ satisfy the Assumptions in Section 4. Then*

$$\lim_{n\to+\infty} \frac{1}{n} D_n(\varphi) = 0$$

for each $\varphi \in C(S^2)$.

Proof. We fix a visual metric d that satisfies the Assumptions in Section 4.

Fix arbitrary $\varphi \in C(S^2)$ and $\varepsilon > 0$. Since φ is uniformly continuous on S^2, there exists a number $\delta > 0$ such that for each pair of p, $q \in S^2$ that satisfy $d(p,q) < \delta$, we have $|\varphi(p) - \varphi(q)| < \varepsilon$. By Proposition 3.6 (ii), there exists a constant $C \geqslant 1$ such that

$$\operatorname{diam}_d\bigl(X^k\bigr) \leqslant C\Lambda^{-k} \qquad \text{for all } k \in \mathbb{N}_0 \text{ and } X^k \in \mathbf{X}^k(f,\mathcal{C}). \tag{6.2}$$

This implies that there exists $N \in \mathbb{N}$ such that $\operatorname{diam}_d\bigl(X^k\bigr) < \delta$ for all integer $k > N$ and $X^k \in \mathbf{X}^k(f,\mathcal{C})$. Then for each sufficiently large $n \in \mathbb{N}$, each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, and each pair of x, $y \in X^n$, by Proposition 3.2 (i), we have

$$\begin{aligned} |S_n\varphi(x) - S_n\varphi(y)| &\leqslant \sum_{k=0}^{n-N-1} \bigl|\varphi(f^i(x)) - \varphi(f^i(y))\bigr| + \sum_{k=n-N}^{n-1} \bigl|\varphi(f^i(x)) - \varphi(f^i(y))\bigr| \\ &\leqslant (n-N)\varepsilon + N\|\varphi\|_\infty. \end{aligned}$$

It follows that

$$\limsup_{n\to+\infty} \frac{1}{n} D_n(\varphi) \leqslant \lim_{n\to+\infty} \frac{n-N}{n}\varepsilon + \frac{N\|\varphi\|_\infty}{n} = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the proof is complete. □

Let f and $\mathcal{C}$ satisfy the Assumptions in Section 4. We assume in addition that $f(\mathcal{C}) \subseteq \mathcal{C}$. In the following, we set

$$E^\infty \coloneqq \bigcup_{n\in\mathbb{N}_0} f^{-n}(\mathcal{C}). \tag{6.3}$$

Then E^∞ is a Borel set. Proposition 3.2 (ii) implies that E^∞ is equal to the union of all edges. Since every vertex is contained in an edge, the set E^∞ also contains all vertices. Moreover, we have

$$f^{-1}(E^\infty) = E^\infty. \tag{6.4}$$

Indeed, note that $\mathcal{C} \subseteq f^{-1}(\mathcal{C})$ and so

$$f^{-1}(E^\infty) = f^{-1}\Biggl(\bigcup_{n\in\mathbb{N}_0} f^{-n}(\mathcal{C})\Biggr) = \bigcup_{n\in\mathbb{N}_0} f^{-(n+1)}(\mathcal{C}) = \bigcup_{n\in\mathbb{N}} f^{-n}\mathcal{C} = \mathcal{C} \cup \bigcup_{n\in\mathbb{N}} f^{-n}\mathcal{C} = E^\infty.$$

The next lemma approximates ergodic measures with a finite collection of tiles in a particular sense.

Lemma 6.2. *Let f and $\mathcal{C}$ satisfy the Assumptions in Section 4. We assume in addition that $f(\mathcal{C}) \subseteq \mathcal{C}$. Consider $\ell \in \mathbb{N}$ and $\vec{\varphi} \in C(S^2)^\ell$. Then for each ergodic f-invariant measure $\mu \in \mathcal{P}(S^2)$ and each $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for each integer $n \geqslant n_0$, there exists a non-empty subset $\mathbf{T}^n$ of $\mathbf{X}^n(f,\mathcal{C})$ such that*

$$\left|\frac{1}{n}\log\operatorname{card}(\mathbf{T}^n) - h_\mu(f)\right| \leqslant \varepsilon \qquad \textit{and} \tag{6.5}$$

$$\sup_{x\in\bigcup\mathbf{T}^n} \left\|\frac{1}{n}S_n\vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| \leqslant \varepsilon. \tag{6.6}$$

To prove Lemma 6.2, we need the following lemma, which is a generalization of [BM17, Lemma 17.7].

Lemma 6.3. *Let f and $\mathcal{C}$ satisfy the Assumptions in Section 4. We assume in addition that $f(\mathcal{C}) \subseteq \mathcal{C}$. Consider an f-invariant Borel probability measure $\mu \in \mathcal{P}(S^2)$. Then the following statements hold:*

(i) *Suppose that $\mu(E^\infty) = 0$. Then for each $n \in \mathbb{N}$, the set $\mathbf{X}^n(f, \mathcal{C})$ forms a measurable partition for (S^2, μ) and is equivalent to the partition ξ_f^n where $\xi = \mathbf{X}^1(f, \mathcal{C})$. Moreover, $\xi = \mathbf{X}^1(f, \mathcal{C})$ is a generator for (f, μ).*

(ii) *Suppose that $\mu(E^\infty) = 1$ and μ is non-atomic. Then $\mu(\mathcal{C}) = 1$. Additionally, for each $n \in \mathbb{N}$, the set $\mathbf{E}^n(f, \mathcal{C})$ forms a measurable partition for (S^2, μ) and is equivalent to the partition η_f^n where $\eta = \mathbf{E}^1(f, \mathcal{C})$. Moreover, $\eta = \mathbf{E}^1(f, \mathcal{C})$ is a generator for (f, μ).*

(iii) *Suppose that there exists a point $x \in S^2$ such that $\mu(\{x\}) > 0$. Then x is a periodic point of f. Moreover, if we assume in addition that μ is ergodic, then $\mu = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{f^i(x)}$, where n is the period of x.*

Proof. (i) Statement (i) was established in [BM17, Lemma 17.7].

(ii) We first note that it follows immediately from our assumptions that $\mu(\mathcal{C}) = 1$. Indeed, since μ is f-invariant, we have $\mu(f^{-n}(\mathcal{C})) = \mu(\mathcal{C})$ for all $n \in \mathbb{N}_0$. On the other hand, $\mathcal{C} \subseteq f^{-n}(\mathcal{C})$, and so $\mu(f^{-n}(\mathcal{C}) \setminus \mathcal{C}) = 0$. This implies that $\mu(E^\infty \setminus \mathcal{C}) = 0$. So μ is actually concentrated on $\mathcal{C}$.

Since μ is non-atomic, we have $\mu(v) = 0$ for all vertices $v \in \bigcup_{n \in \mathbb{N}_0} \mathbf{V}^n(f, \mathcal{C})$. For each $n \in \mathbb{N}$, since $f(\mathcal{C}) \subseteq \mathcal{C} \subseteq f^{-n}(\mathcal{C})$, the set $\bigcup \mathbf{E}^n(f, \mathcal{C})$ has full measure, and two distinct n-edges have only vertices, i.e., a set of μ-measure zero, in common. Hence $\mathbf{E}^n(f, \mathcal{C})$ is a measurable partition for (S^2, μ).

Fix arbitrary $n \in \mathbb{N}$ and $e \in \mathbf{E}^n(f, \mathcal{C})$. We now show that $\mathbf{E}^n(f, \mathcal{C})$ is equivalent to the partition η_f^n, where $\eta = \mathbf{E}^1(f, \mathcal{C})$. For $i = 1, \dots, n$ there exist unique i-edges $e^i \in \mathbf{E}^i(f, \mathcal{C})$ with $e = e^n \subseteq e^{n-1} \subseteq \cdots \subseteq e^1$. Set $e_i \coloneqq f^{i-1}(e^i)$ for $i = 1, \dots, n$. Then $e_1, \dots, e_n$ are 1-edges. We claim that

$$e = e_1 \cap f^{-1}(e_2) \cap \cdots \cap f^{-(n-1)}(e_n). \tag{6.7}$$

To see this, denote the right hand side in this equation by $\widetilde{e}$. Then it is clear that $e \subseteq \widetilde{e}$. We verify $e = \widetilde{e}$ by inductively showing that for any point $x \in \widetilde{e}$ we have $x \in e^i$ for $i = 1, \dots, n$, and so $x \in e^n = e$.

Indeed, since $\widetilde{e} \subseteq e_1 = e^1$ this is clear for $i = 1$. Suppose $x \in e^{i-1}$ for some $i \in \mathbb{N}$ with $2 \leqslant i \leqslant n$. To complete the inductive step, we have to show $x \in e^i$. Note that $x \in \widetilde{e} \subseteq f^{-(i-1)}(e_i)$ and so $f^{i-1}(x) \in e_i$. By Proposition 3.2 (i), the map $f^{i-1}|_{e^{i-1}}$ is a homeomorphism of e^{i-1} onto the 0-edge $f^{i-1}(e^{i-1})$. Moreover, $x \in e^{i-1}$, $e^i \subseteq e^{i-1}$, and $f^{i-1}(x) \in e_i = f^{i-1}(e^i)$. Hence by injectivity of f^{i-1} on e^{i-1} we have $x \in e^i$ as desired.

Equation (6.7) shows that every element in $\mathbf{E}^n(f, \mathcal{C})$ belongs to η_f^n, where $\eta = \mathbf{E}^1(f, \mathcal{C})$. This implies that the measurable partitions $\mathbf{E}^n(f, \mathcal{C})$ and η_f^n are equivalent (η_f^n may contain additional sets, but they have to be of measure zero).

To establish that $\eta = \mathbf{E}^1(f, \mathcal{C})$ is a generator, let $B \subseteq S^2$ be an arbitrary Borel set and $\varepsilon > 0$. Since the measurable partitions $\mathbf{E}^n(f, \mathcal{C})$ and η_f^n are equivalent for each $n \in \mathbb{N}$, it suffices to show that there exists $k \in \mathbb{N}$ and a union A of k-edges such that $\mu(A \triangle B) < \varepsilon$.

By regularity of μ there exists a compact set $K \subseteq B$ and an open set $U \subseteq S^2$ with $K \subseteq B \subseteq U$ and $\mu(U \setminus K) < \varepsilon$. Since the diameters of edges approach 0 uniformly as their levels become larger, we can choose $k \in \mathbb{N}$ large enough such that every k-edge that meets K is contained in the open neighborhood U of K. Define $K_\mathcal{C} \coloneqq K \cap \mathcal{C}$ and

$$A \coloneqq \bigcup \bigl\{ e^k \in \mathbf{E}^k(f, \mathcal{C}) : e^k \cap K_\mathcal{C} \neq \emptyset \bigr\}.$$

Then $K_\mathcal{C} \subseteq A \subseteq U$. This implies $A \triangle B \subseteq U \setminus K_\mathcal{C}$, and so

$$\mu(A \triangle B) \leqslant \mu(U \setminus K_\mathcal{C}) \leqslant \mu(U \setminus K) + \mu(U \setminus \mathcal{C}) = \mu(U \setminus K) < \varepsilon$$

as desired. The proof of statement (ii) is complete.

(iii) Assume first that there exists a pint $x \in S^2$ such that $\mu(\{x\}) > 0$.

We claim that x is preperiodic. Otherwise, for each pair of $k, \ell \in \mathbb{N}$ with $k \neq \ell$, we have $f^k(x) \neq f^\ell(x)$. We write $x_n \coloneqq f^n(x)$ for each $n \in \mathbb{N}$. Then $x \in f^{-n}(x_n)$ and $\mu(\{x_n\}) = \mu(f^{-n}(x_n)) \geqslant \mu(\{x\})$

for each $n \in \mathbb{N}$ since μ is f-invariant. This implies

$$1 = \mu(S^2) \geqslant \sum_{n=1}^{+\infty} \mu(\{x_n\}) \geqslant \sum_{n=1}^{+\infty} \mu(\{x\}) = +\infty,$$

which is a contradiction. This proves the claim that x is preperiodic.

We now show that x is periodic. Since x is preperiodic, there exist $m \in \mathbb{N}_0$ and $n \in \mathbb{N}$ such that $f^{n+m}(x) = f^m(x)$. Denote $y \coloneqq f^{nm}(x)$. Then $f^n(y) = f^{n+nm}(x) = f^{nm}(x) = y$, i.e., y is a periodic point of f. Since $f^{nm}(x) = f^{nm}(y) = y$ and μ is f-invariant, we have

$$\mu(\{y\}) = \mu(f^{-nm}(y)) \geqslant \mu(\{x\} \cup \{y\}).$$

This implies $x = y$ since $\mu(\{x\}) > 0$. Thus x is periodic.

Finally, we assume in addition that μ is ergodic. Let $n \in \mathbb{N}$ be the period of x. We set

$$\operatorname{Orb}(x) \coloneqq \bigl\{ f^k(x) : k \in \mathbb{N}_0 \bigr\} = \bigl\{ f^k(x) : k \in \{0, \dots, n-1\} \bigr\}$$

and

$$x^\infty \coloneqq \{x\} \cup \bigcup_{k \in \mathbb{N}} f^{-k}(x) = \bigcup_{k \in \mathbb{N}_0} f^{-k}(\operatorname{Orb}(x)).$$

Since $x \in f^{-n}(x)$, we have

$$\begin{aligned} f^{-1}(x^\infty) &= f^{-1}\Biggl(\{x\} \cup \bigcup_{k \in \mathbb{N}} f^{-k}(x) \Biggr) = f^{-1}(x) \cup \bigcup_{k \in \mathbb{N}} f^{-(k+1)}(x) \\ &= \bigcup_{k \in \mathbb{N}} f^{-k}(x) = \{x\} \cup \bigcup_{k \in \mathbb{N}} f^{-k}(x) = x^\infty. \end{aligned}$$

This implies $\mu(x^\infty) = 1$ since μ is ergodic and $\mu(x^\infty) \geqslant \mu(\{x\}) > 0$. Since μ is f-invariant, we have $\mu\bigl(f^{-k}(\operatorname{Orb}(x))\bigr) = \mu(\operatorname{Orb}(x))$ for all $k \in \mathbb{N}_0$. On the other hand, $\operatorname{Orb}(x) \subseteq f^{-k}(\operatorname{Orb}(x))$, and so $\mu\bigl(f^{-k}(\operatorname{Orb}(x)) \setminus \operatorname{Orb}(x)\bigr) = 0$. Thus we have $\mu(x^\infty \setminus \operatorname{Orb}(x)) = 0$. So μ actually concentrates on $\operatorname{Orb}(x)$.

It suffices to show that $\mu\bigl(\bigl\{f^k(x)\bigr\}\bigr) = 1/n$ for each $k \in \{0, \dots, n-1\}$. Indeed, since μ is f-invariant and $f^k(x) \in f^{-1}\bigl(f^{k+1}(x)\bigr)$, we have $\mu\bigl(\bigl\{f^{k+1}(x)\bigr\}\bigr) = \mu\bigl(f^{-1}\bigl(f^{k+1}(x)\bigr)\bigr) \geqslant \mu\bigl(\bigl\{f^k(x)\bigr\}\bigr)$. This implies

$$\mu(\{x\}) \leqslant \mu\bigl(\bigl\{f^k(x)\bigr\}\bigr) \leqslant \mu(\{f^n(x)\}) = \mu(\{x\}).$$

Hence $\mu\bigl(\bigl\{f^k(x)\bigr\}\bigr) = \mu(\{x\}) = 1/n$ for each $k \in \{0, \dots, n-1\}$.

The proof of statement (iii) is complete. □

Now we can prove Lemma 6.2.

Proof of Lemma 6.2. Fix arbitrary $\varepsilon > 0$, $\ell \in \mathbb{N}$, $\vec{\varphi} \in C(S^2)^\ell$, and ergodic f-invariant measure $\mu \in \mathcal{P}(S^2)$. It suffices to show that for every sufficiently large $n \in \mathbb{N}$, there exists a non-empty subset $\mathbf{T}^n$ of $\mathbf{X}^n(f, \mathcal{C})$ such that (6.5) and (6.6) hold. We split the proof into three cases according to the properties of measure μ. Recall the definition of E^∞ in (6.3).

Case 1: $\mu(E^\infty) = 0$.

Since μ is f-invariant and $\mu(E^\infty) = 0$, by Lemma 6.3 (i), $\mathbf{X}^1(f, \mathcal{C})$ is a generator for (f, μ), and for each $n \in \mathbb{N}$ the set $\mathbf{X}^n(f, \mathcal{C})$ forms a measurable partition for (S^2, μ) and is equivalent to the partition ξ_f^n where $\xi = \mathbf{X}^1(f, \mathcal{C})$. Then one can use Birkhoff's ergodic theorem and the Shannon–McMillan–Breiman theorem to show that for every sufficiently large $n \in \mathbb{N}$, there exists a non-empty subset $\mathbf{T}^n$ of $\mathbf{X}^n(f, \mathcal{C})$ such that (6.5) and (6.6) hold.

Case 2: $\mu(E^\infty) > 0$ and μ is non-atomic.

First note that $\mu(E^\infty) = 1$ in this case since μ is ergodic and $f^{-1}(E^\infty) = E^\infty$ (see (6.4)). Then it follows from Lemma 6.3 (ii) that $\mathbf{E}^1(f, \mathcal{C})$ is a generator for (f, μ), and for each $n \in \mathbb{N}$ the set $\mathbf{E}^n(f, \mathcal{C})$ forms a measurable partition for (S^2, μ) and is equivalent to the partition η_f^n where $\eta = \mathbf{E}^1(f, \mathcal{C})$. Similar to Case 1, one can use Birkhoff's ergodic theorem and the Shannon–McMillan–Breiman theorem to show that there exists $N \in \mathbb{N}$ such that for each integer $n \geqslant N$, there exists a non-empty subset $\mathbf{S}^n$ of $\mathbf{E}^n(f, \mathcal{C})$ such that

$$\left| \frac{1}{n} \log \operatorname{card}(\mathbf{S}^n) - h_\mu(f) \right| \leqslant \varepsilon/2 \qquad \text{and} \tag{6.8}$$

$$\sup_{x \in \bigcup \mathbf{S}^n} \left\| \frac{1}{n} S_n \vec{\varphi}(x) - \int \vec{\varphi} \, \mathrm{d}\mu \right\| \leqslant \varepsilon/2. \tag{6.9}$$

For each integer $n \geqslant N$, we set

$$\mathbf{T}^n := \{X^n \in \mathbf{X}^n(f, \mathcal{C}) : e^n \subseteq \partial X^n, \, e^n \in \mathbf{S}^n\}.$$

Then by Proposition 3.2 (iii) and Remark 3.3, we have

$$\operatorname{card}(\mathbf{S}^n) / \operatorname{card}(\operatorname{post} f) \leqslant \operatorname{card}(\mathbf{T}^n) \leqslant 2 \operatorname{card}(\mathbf{S}^n).$$

Combining this with (6.8), we see that (6.5) holds for every sufficiently large $n \in \mathbb{N}$. Noting that $\vec{\varphi} = (\varphi_1, \dots, \varphi_\ell) \in C(S^2)^\ell$, by Lemma 6.1, we have

$$\sup_{X^n \in \mathbf{X}^n(f, \mathcal{C})} \sup_{x, y \in X^n} \left\| \frac{1}{n} S_n \vec{\varphi}(x) - \frac{1}{n} S_n \vec{\varphi}(y) \right\| \longrightarrow 0 \qquad \text{as } n \to +\infty.$$

Therefore, (6.9) implies that (6.6) holds for every sufficiently large $n \in \mathbb{N}$.

Case 3: $\mu(E^\infty) > 0$ and there exists $p \in S^2$ such that $\mu(p) > 0$.

By Lemma 6.3 (iii), p is a periodic point of f and $\mu = V_k(p)$, where k is the period of p. In particular, we have $h_\mu(f) = 0$. Fix arbitrary $n \in \mathbb{N}$. For each $i \in \{1, \dots, k-1\}$, we choose an n-tile $X_i^n \in \mathbf{X}^n(f, \mathcal{C})$ such that $f^i(p) \in X_i^n$. We denote by $\mathbf{T}^n$ the set of those n-tiles. Then $1 \leqslant \operatorname{card}(\mathbf{T}^n) \leqslant k$. This implies (6.5) holds for every sufficiently large $n \in \mathbb{N}$. Since $\lim_{n \to +\infty} \frac{1}{n} S_n \vec{\varphi}(f^i(x)) = k^{-1} S_k \vec{\varphi}(p) = \int \vec{\varphi} \, \mathrm{d}\mu$ for each $i \in \{1, \dots, k-1\}$, by Lemma 6.1, (6.6) holds for every sufficiently large $n \in \mathbb{N}$.

The proof is complete. □

Apart from Lemma 6.2, we need the following lemma to construct strongly primitive subsystems in the proof of Theorem 1.2.

Lemma 6.4. *Let f, $\mathcal{C}$, e^0 satisfy the Assumptions in Section 4. We assume in addition that $f(\mathcal{C}) \subseteq \mathcal{C}$. Consider $\ell \in \mathbb{N}$ and $\vec{\varphi} = (\varphi_1, \dots, \varphi_\ell) \in C(S^2)^\ell$. Let $\mu \in \mathcal{M}(S^2, f)$ be an ergodic measure. Then for each $\varepsilon > 0$, there exists an integer $N \in \mathbb{N}$ such that for each integer $n \geqslant N$ and each color $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, there exists an n-pair $P_\mathfrak{c}^n \in \mathbf{P}^n(f, \mathcal{C}, e^0)$ such that $P_\mathfrak{c}^n \subseteq \operatorname{inte}(X_\mathfrak{c}^0)$ and*

$$\sup_{x \in P_\mathfrak{c}^n} \left\| \frac{1}{n} S_n \vec{\varphi}(x) - \int \vec{\varphi} \, \mathrm{d}\mu \right\| \leqslant \varepsilon.$$

Proof. Let $\varepsilon > 0$ be arbitrary. Since μ is ergodic, it follows from Birkhoff's ergodic theorem that for μ-a.e. $x \in S^2$, $\frac{1}{n} S_n \vec{\varphi}(x) \to \int \vec{\varphi} \, \mathrm{d}\mu$ as $n \to +\infty$. Thus we can fix a point $y \in S^2$ and an integer $n_0 \in \mathbb{N}$ such that for each integer $n \geqslant n_0$,

$$\left\| \frac{1}{n} S_n \vec{\varphi}(y) - \int \vec{\varphi} \, \mathrm{d}\mu \right\| \leqslant \frac{\varepsilon}{2}. \tag{6.10}$$

By Lemma 3.14, there exists an integer $M \in \mathbb{N}$ such that for each color $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, there exists an M-pair $P_\mathfrak{c}^M \in \mathbf{P}^M(f, \mathcal{C}, e^0)$ such that for each integer $n \geqslant M$ and each $x \in P_\mathfrak{c}^M$, we have $U^n(x) \subseteq \operatorname{inte}(X_\mathfrak{c}^0)$.

We fix such an integer $M \in \mathbb{N}$ and the corresponding M-pairs $P_\mathfrak{b}^M$ and $P_\mathfrak{w}^M$ in the following.

Let color $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$ and integer $n \geqslant M + n_0$ be arbitrary. Since $n - M \geqslant n_0$, by (6.10), we have

$$\left\| \frac{1}{n-M} S_{n-M} \vec{\varphi}(y) - \int \vec{\varphi} \, \mathrm{d}\mu \right\| \leqslant \frac{\varepsilon}{2}.$$

Since $f^M(P_{\mathfrak{c}}^M) = S^2$, there exists $x_{\mathfrak{c}} \in P_{\mathfrak{c}}^M$ such that $f^M(x_{\mathfrak{c}}) = y$. Then we have

$$\begin{aligned}\left\|\frac{1}{n}S_n\vec{\varphi}(x_{\mathfrak{c}}) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| &= \left\|\frac{1}{n}S_M\vec{\varphi}(x_{\mathfrak{c}}) + \frac{1}{n}S_{n-M}\vec{\varphi}(y) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| \\ &\leqslant \frac{1}{n}\|S_M\vec{\varphi}\| + \left|\frac{1}{n} - \frac{1}{n-M}\right|\|S_{n-M}\vec{\varphi}\| + \left\|\frac{1}{n-M}S_{n-M}\vec{\varphi}(y) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| \\ &\leqslant \frac{2M}{n}\|\vec{\varphi}\| + \frac{\varepsilon}{2}.\end{aligned}$$

By Definition 3.12, there exists an n-pair $P_{\mathfrak{c}}^n \in \mathbf{P}^n(f, \mathcal{C}, e^0)$ containing $x_{\mathfrak{c}}$. Noting that $x_{\mathfrak{c}} \in P_{\mathfrak{c}}^M$ and $n \geqslant M$, by Lemma 3.14, we get $U^n(x_{\mathfrak{c}}) \subseteq \operatorname{inte}\bigl(X_{\mathfrak{c}}^0\bigr)$. Thus it follows from the definition of $U^n(x_{\mathfrak{c}})$ and $P_{\mathfrak{c}}^n$ that $x_{\mathfrak{c}} \in P_{\mathfrak{c}}^n \subseteq U^n(x_{\mathfrak{c}}) \subseteq \operatorname{inte}\bigl(X_{\mathfrak{c}}^0\bigr)$. Then we have

$$\begin{aligned}\sup_{x\in P_{\mathfrak{c}}^n}\left\|\frac{1}{n}S_n\vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| &\leqslant \sup_{x\in P_{\mathfrak{c}}^n}\left\|\frac{1}{n}S_n\vec{\varphi}(x) - \frac{1}{n}S_n\vec{\varphi}(x_{\mathfrak{c}})\right\| + \left\|\frac{1}{n}S_n\vec{\varphi}(x_{\mathfrak{c}}) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| \\ &\leqslant \frac{2D_n(\vec{\varphi})}{n} + \frac{2M}{n}\|\vec{\varphi}\| + \frac{\varepsilon}{2}.\end{aligned}$$

Therefore, by Lemma 6.1, we can find a sufficiently large integer $N \in \mathbb{N}$ such that for each integer $n \geqslant N$ and each color $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, there exists an n-pair $P_{\mathfrak{c}}^n \in \mathbf{P}^n(f, \mathcal{C}, e^0)$ such that $P_{\mathfrak{c}}^n \subseteq \operatorname{inte}\bigl(X_{\mathfrak{c}}^0\bigr)$ and

$$\sup_{x\in P_{\mathfrak{c}}^n}\left\|\frac{1}{n}S_n\vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu\right\| \leqslant \varepsilon.$$

The proof is complete. □

The following result shows that in order to prove Theorem 1.2, it suffices to prove entropy density of ergodic measures for some iterate of the map f.

Proposition 6.5. *Let (X, d) be a compact metric space and $T\colon X \to X$ be a continuous map. Consider arbitrary $n \in \mathbb{N}$. Then $\mathcal{M}_{\mathrm{erg}}(X, T)$ is entropy-dense in $\mathcal{M}(X, T)$ if and only if $\mathcal{M}_{\mathrm{erg}}(X, T^n)$ is entropy-dense in $\mathcal{M}(X, T^n)$.*

Proof. Fix arbitrary $n \in \mathbb{N}$. Since T^n itself is also a continuous map from X to X, it suffices to prove the "if" part. We assume that for each $\nu \in \mathcal{M}(X, T^n)$, there exists a sequence $\{\nu_k\}_{k\in\mathbb{N}}$ of T^n-invariant ergodic measures of T^n that converges to ν in the weak*-topology with $h_{\nu_k}(T^n) \to h_\nu(T^n)$ as $k \to +\infty$.

Let $\mu \in \mathcal{M}(X, T)$ be arbitrary. Since $\mu \in \mathcal{M}(X, T) \subseteq \mathcal{M}(X, T^n)$, there exists a sequence $\{\nu_k\}_{k\in\mathbb{N}}$ of T^n-invariant ergodic measures of T^n that converges to μ in the weak*-topology with $h_{\nu_k}(T^n) \to h_\mu(T^n)$ as $k \to +\infty$. For each $k \in \mathbb{N}$, we define

$$\mu_k \coloneqq \frac{1}{n}\sum_{i=0}^{n-1} T_*^i \nu_k.$$

Then it follows from Lemma 5.2 that $\mu_k \in \mathcal{M}(X, T)$ is ergodic for T and $nh_{\mu_k}(T) = h_{\nu_k}(T^n)$. This implies $h_{\mu_k}(T) \to h_\mu(T)$ as $k \to +\infty$. Noting that the sequence $\{\mu_k\}_{k\in\mathbb{N}}$ also converges to μ in the weak*-topology, the proof is complete. □

After these preparations, we are ready to prove the main result of this section.

Proof of Theorem 1.2. By Proposition 6.5, it suffices to prove that $\mathcal{M}_{\mathrm{erg}}(S^2, f^i)$ is entropy-dense in $\mathcal{M}(S^2, f^i)$ for some $i \in \mathbb{N}$. Hence by Lemma 3.7, we may assume without loss of generality that there exists a Jordan curve $\mathcal{C} \subseteq S^2$ containing $\operatorname{post} f$ such that $f(\mathcal{C}) \subseteq \mathcal{C}$.

Let $\mu \in \mathcal{M}(S^2, f)$, $\varepsilon > 0$, $\ell \in \mathbb{N}$, and $\vec{\varphi} = (\varphi_1, \dots, \varphi_\ell) \in C(S^2)^\ell$ be arbitrary. By the definition of entropy density (see Subsection 1.1), it suffices to find an ergodic measure $\nu \in \mathcal{M}_{\mathrm{erg}}(S^2, f)$ such that $\bigl\|\int \vec{\varphi}\,\mathrm{d}\nu - \int \vec{\varphi}\,\mathrm{d}\mu\bigr\| \leqslant \varepsilon$ and $|h_\nu(f) - h_\mu(f)| \leqslant \varepsilon$.

By virtue of the Choquet representation theorem (see for example, [KH95, Theorem A.2.10]) and Jacobs' Theorem (see for example, [Wal82, Theorem 8.4]), for every neighborhood Γ of μ in $\mathcal{M}(S^2, f)$

and every $\delta > 0$ there exist $s \in \mathbb{N}$ and ergodic measures $\mu_1, \dots, \mu_s$ and $\rho_1, \dots, \rho_s \in (0,1)$ such that $\rho_1 + \cdots + \rho_s = 1$, and the measure $\bar{\mu} := \rho_1\mu_1 + \cdots + \rho_s\mu_s$ belongs to Γ and satisfies $|h_{\bar{\mu}}(f) - h_\mu(f)| < \delta$. Hence, we may assume without loss of generality that μ is a convex combination of finitely many ergodic measures, i.e.,

$$\mu = \rho_1\mu_1 + \cdots + \rho_s\mu_s \tag{6.11}$$

with $s \in \mathbb{N}$, $\mu_1, \dots, \mu_s \in \mathcal{M}_{\mathrm{erg}}(S^2, f)$, $\rho_1, \dots, \rho_s \in (0,1)$, and $\rho_1 + \cdots + \rho_s = 1$.

We fix a 0-edge $e^0 \in \mathbf{E}^0(f, \mathcal{C})$. By Lemmas 6.2 and 6.4, for each $i \in \{1, \dots, s\}$, there exists $\widetilde{N}_i \in \mathbb{N}$ such that for each integer $n_i \geqslant \widetilde{N}_i$, there exists a non-empty subset $\widetilde{\mathbf{T}}^{n_i}$ of $\mathbf{X}^{n_i}(f, \mathcal{C})$ such that

$$\biggl| \frac{1}{n_i} \log \operatorname{card}\bigl(\widetilde{\mathbf{T}}^{n_i}\bigr) - h_{\mu_i}(f) \biggr| \leqslant \frac{\varepsilon}{4} \qquad \text{and} \tag{6.12}$$

$$\sup_{x \in \bigcup \widetilde{\mathbf{T}}^{n_i}} \biggl\| \frac{1}{n_i} S_{n_i}\vec{\varphi}(x) - \int \vec{\varphi} \,\mathrm{d}\mu_i \biggr\| \leqslant \frac{\varepsilon}{6}, \tag{6.13}$$

and for each color $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, there exists an n_i-pair $P^{n_i}_{\mathfrak{c}} \in \mathbf{P}^{n_i}(f, \mathcal{C}, e^0)$ such that $P^{n_i}_{\mathfrak{c}} \subseteq \operatorname{inte}\bigl(X^0_{\mathfrak{c}}\bigr)$ and

$$\sup_{x \in P^{n_i}_{\mathfrak{c}}} \biggl\| \frac{1}{n_i} S_{n_i}\vec{\varphi}(x) - \int \vec{\varphi} \,\mathrm{d}\mu_i \biggr\| \leqslant \frac{\varepsilon}{6}. \tag{6.14}$$

Fix arbitrary $i \in \{1, \dots, s\}$. For each $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$ we can write $P^{n_i}_{\mathfrak{c}} = X^{n_i}_{\mathfrak{bc}} \cup X^{n_i}_{\mathfrak{wc}}$ for some $X^{n_i}_{\mathfrak{bc}} \in \mathbf{X}^{n_i}_{\mathfrak{b}}(f, \mathcal{C})$ and $X^{n_i}_{\mathfrak{wc}} \in \mathbf{X}^{n_i}_{\mathfrak{w}}(f, \mathcal{C})$. Note that by Proposition 3.2 (i), for each $X^{n_i} \in \mathbf{X}^{n_i}(f, \mathcal{C})$ there exists exactly one n_i-edge $e^{n_i} \subseteq \partial X^{n_i}$ such that $f^{n_i}(e^{n_i}) = e^0$. Then by Remark 3.3, there exist exactly two n_i-tiles in $\mathbf{X}^{n_i}(f, \mathcal{C})$ containing e^{n_i}, one of which is X^{n_i} itself, and we denote by $\widetilde{X}^{n_i}$ the other. Indeed, $X^{n_i} \cup \widetilde{X}^{n_i}$ is an n_i-pair in $\mathbf{P}^{n_i}(f, \mathcal{C}, e^0)$. Now we construct a new subset $\mathbf{T}^{n_i}$ of $\mathbf{X}^{n_i}(f, \mathcal{C})$ from the old one $\widetilde{\mathbf{T}}^{n_i}$ by setting

$$\mathbf{T}^{n_i} := \widetilde{\mathbf{T}}^{n_i} \cup \bigl\{ \widetilde{X}^{n_i} : X^{n_i} \in \widetilde{\mathbf{T}}^{n_i} \bigr\} \cup \bigl\{ X^{n_i}_{\mathfrak{bb}} \bigr\} \cup \bigl\{ X^{n_i}_{\mathfrak{wb}} \bigr\} \cup \bigl\{ X^{n_i}_{\mathfrak{bw}} \bigr\} \cup \bigl\{ X^{n_i}_{\mathfrak{ww}} \bigr\}$$

Then $\bigcup \mathbf{T}^{n_i}$ is actually a union of some pairs in $\mathbf{P}^{n_i}(f, \mathcal{C}, e^0)$ and we have

$$\operatorname{card}\bigl(\widetilde{\mathbf{T}}^{n_i}\bigr) \leqslant \operatorname{card}(\mathbf{T}^{n_i}) \leqslant 2 \operatorname{card}\bigl(\widetilde{\mathbf{T}}^{n_i}\bigr) + 4.$$

Moreover, it follows from (6.13), (6.14), and the structure of pairs that

$$\sup_{x \in \bigcup \mathbf{T}^{n_i}} \biggl\| \frac{1}{n_i} S_{n_i}\vec{\varphi}(x) - \int \vec{\varphi} \,\mathrm{d}\mu_i \biggr\| \leqslant \frac{\varepsilon}{6} + \frac{2D_{n_i}(\vec{\varphi})}{n_i}.$$

Therefore, by (6.12) and Lemma 6.1, we can find a sufficiently large integer $N_i \in \mathbb{N}$ such that for each integer $n_i \geqslant N_i$, there exists a non-empty subset $\mathbf{T}^{n_i}$ of $\mathbf{X}^{n_i}(f, \mathcal{C})$ such that

$$\biggl| \frac{1}{n_i} \log \operatorname{card}(\mathbf{T}^{n_i}) - h_{\mu_i}(f) \biggr| \leqslant \frac{\varepsilon}{2} \qquad \text{and} \tag{6.15}$$

$$\sup_{x \in \bigcup \mathbf{T}^{n_i}} \biggl\| \frac{1}{n_i} S_{n_i}\vec{\varphi}(x) - \int \vec{\varphi} \,\mathrm{d}\mu_i \biggr\| \leqslant \frac{\varepsilon}{3}. \tag{6.16}$$

Moreover, we have $\operatorname{card}\bigl(\mathbf{T}^{n_i} \cap \mathbf{X}^{n_i}_{\mathfrak{b}}(f, \mathcal{C})\bigr) = \operatorname{card}\bigl(\mathbf{T}^{n_i} \cap \mathbf{X}^{n_i}_{\mathfrak{w}}(f, \mathcal{C})\bigr)$, and for each pair of $\mathfrak{c}, \mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$ there exists $X^{n_i}_{\mathfrak{cc}'} \in \mathbf{T}^{n_i}$ such that $f^{n_i}\bigl(X^{n_i}_{\mathfrak{cc}'}\bigr) = X^0_{\mathfrak{c}}$ and $X^{n_i}_{\mathfrak{cc}'} \subseteq X^0_{\mathfrak{c}'}$. In the rest of the proof for each $i \in \{1, \dots, s\}$ we fix an integer $n_i \geqslant N_i$ and a corresponding non-empty subset $\mathbf{T}^{n_i}$ of $\mathbf{X}^{n_i}(f, \mathcal{C})$ obtained from the above construction.

We now introduce some notions that will be used in the rest of the proof. Let $n \in \mathbb{N}$ and $X^n \in \mathbf{X}^n(f, \mathcal{C})$ be arbitrary. Set $Y_{n-j} := f^j(X^n)$ for each $j \in \{0, \dots, n-1\}$. We label the 1-tiles by $X^1_1, \dots, X^1_{2\deg f}$. Then by Proposition 3.8, for each $j \in \{0, \dots, n-1\}$, there exists a unique integer $t_j \in \{1, \dots, 2\deg f\}$ such that $Y_{n-j} \subseteq X^1_{t_j}$. We denote by $w(X^n)$ the n-string $t_0 t_1 \cdots t_{n-1}$ and by $[w(X^n)]$ the n-tile X^n.

Let $k \in \mathbb{N}$ and $Y^k \in \mathbf{X}^k(f,\mathcal{C})$ be arbitrary. If $Y^k \subseteq f^n(X^n)$, then it follows from Proposition 3.2 (i) and [BM17, Lemma 5.17 (i)] that $Z^{n+k} \coloneqq (f^n|_{X^n})^{-1}\bigl(Y^k\bigr) \in \mathbf{X}^{n+k}(f,\mathcal{C})$. One can check that $w\bigl(Z^{n+k}\bigr) = w(X^n)w\bigl(Y^k\bigr)$ in this case. By Definition 3.20 and Remark 3.21, there exists a constant $N \in \mathbb{N}$ such that for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b},\mathfrak{w}\}$, there exists $X^N_{\mathfrak{c}\mathfrak{c}'} \in \mathbf{X}^N_{\mathfrak{c}}(f,\mathcal{C})$ satisfying $X^N_{\mathfrak{c}\mathfrak{c}'} \subseteq X^0_{\mathfrak{c}'}$. We define $\lambda_{\mathfrak{c}\mathfrak{c}'} \coloneqq w\bigl(X^N_{\mathfrak{c}\mathfrak{c}'}\bigr)$ for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b},\mathfrak{w}\}$. Then $[\lambda_{\mathfrak{c}\mathfrak{c}'}] = X^N_{\mathfrak{c}\mathfrak{c}'}$ for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b},\mathfrak{w}\}$. If $f^n(X^n) = X^0_{\mathfrak{c}}$ and $Y^k \subseteq X^0_{\mathfrak{c}'}$ for some $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b},\mathfrak{w}\}$, we define $\lambda\bigl(X^n,Y^k\bigr) \coloneqq \lambda_{\mathfrak{c}\mathfrak{c}'}$. One can check that there exists $Z^{n+k+N} \in \mathbf{X}^{n+k+N}(f,\mathcal{C})$ such that $w\bigl(Z^{n+k+N}\bigr) = w(X^n)\lambda\bigl(X^n,Y^k\bigr)w\bigl(Y^k\bigr)$, $Z^{n+k+N} \subseteq X^n$, and $f^{n+N}\bigl(Z^{n+k+N}\bigr) = Y^k$.

For each $i \in \{1, \dots, s\}$, let $r_i \in \mathbb{N}$ be arbitrary. Denote by M_i the integer $n_ir_i + N(r_i-1)$ and by $\mathbf{T}^{n_i,r_i}$ the non-empty subset of $\mathbf{X}^{M_i}(f,\mathcal{C})$ consisting of M_i-tiles of the form

$$\bigl[w(X_1)\lambda(X_1,X_2)w(X_2)\lambda(X_2,X_3)w(X_3)\cdots\lambda(X_{r_i-1},X_{r_i})w(X_{r_i})\bigr] \tag{6.17}$$

with $X_1, \dots, X_{r_i} \in \mathbf{T}^{n_i}$. Denote by R the integer $sN + \sum_{j=1}^s n_jr_j$ and by $\mathbf{T}$ the non-empty subset of $\mathbf{X}^R(f,\mathcal{C})$ consisting of R-tiles of the form

$$\bigl[w(Y_1)\lambda(Y_1,Y_2)w(Y_2)\lambda(Y_2,Y_3)w(Y_3)\cdots\lambda(Y_{s-1},Y_s)w(Y_s)\lambda(Y_s,Y_1)\bigr] \tag{6.18}$$

with $Y_j \in \mathbf{T}^{n_j,r_j}$ for each $j \in \{1, \dots, s\}$. Note that

$$\operatorname{card}(\mathbf{T}) = \prod_{j=1}^s \bigl(\operatorname{card}\bigl(\mathbf{T}^{n_j}\bigr)\bigr)^{r_j}. \tag{6.19}$$

Enlarging each n_i if necessary, it is possible to choose integers r_i such that the following holds:

$$\frac{\log 2}{R} \leqslant \frac{\varepsilon}{6}, \tag{6.20}$$

$$\sum_{i=1}^s \biggl(h_{\mu_i}(f) + \biggl\lVert \int \vec{\varphi}\,\mathrm{d}\mu_i \biggr\rVert\biggr)\Bigl|\rho_i - \frac{n_ir_i}{R}\Bigr| \leqslant \frac{\varepsilon}{3}, \qquad \text{and} \tag{6.21}$$

$$\frac{1}{R}\sum_{i=1}^s r_i \sup_{\mathfrak{c},\mathfrak{c}'\in\{\mathfrak{b},\mathfrak{w}\}} \sup_{x\in[\lambda_{\mathfrak{c}\mathfrak{c}'}]} \bigl\lVert S_N\vec{\varphi}(x)\bigr\rVert \leqslant \frac{\varepsilon}{3}. \tag{6.22}$$

By our construction of $\mathbf{T}$ and $\mathbf{T}^{n_i}$ for $i \in \{1, \dots, s\}$, we have

$$\operatorname{card}\bigl(\mathbf{T}\cap\mathbf{X}^R_{\mathfrak{b}}(f,\mathcal{C})\bigr) = \operatorname{card}\bigl(\mathbf{T}\cap\mathbf{X}^R_{\mathfrak{w}}(f,\mathcal{C})\bigr) = \operatorname{card}(\mathbf{T})/2 \tag{6.23}$$

and $f^R|_{\bigcup\mathbf{T}}$ is a strongly primitive subsystem of f^R with respect to $\mathcal{C}$. We set $F \coloneqq f^R|_{\bigcup\mathbf{T}}$ and $\widehat{F} \coloneqq F|_\Omega = f^R|_\Omega$, where $\Omega \coloneqq \Omega(F,\mathcal{C})$ is the tile maximal invariant set associated with F with respect to $\mathcal{C}$. Then it follows from Theorem 3.25 that there exists $\widehat{\nu} \in \mathcal{M}(\Omega,\widehat{F}) \subseteq \mathcal{M}\bigl(S^2,f^R\bigr)$ such that $h_{\widehat{\nu}}\bigl(f^R\bigr) = h_{\widehat{\nu}}(\widehat{F}) = P(F,0) = h_{\mathrm{top}}(F)$ (recall Definition 3.23) and $\widehat{\nu}$ is ergodic for $\widehat{F}$. Define

$$\nu \coloneqq \frac{1}{R}\sum_{j=0}^{R-1} f^j_*\widehat{\nu}. \tag{6.24}$$

Noting that $\widehat{\nu}$ is also ergodic for f^R and then applying Lemma 5.2, we deduce that $\nu \in \mathcal{M}(S^2,f)$ is ergodic for f and $Rh_\nu(f) = h_{\widehat{\nu}}(f^R) = h_{\mathrm{top}}(F)$.

We now calculate $h_\nu(f)$. By Definition 3.23 and Proposition 3.24, we have $h_{\mathrm{top}}(F) = \log(\rho(A))$, where A is the tile matrix of F with respect to $\mathcal{C}$ and $\rho(A)$ is the spectral radius of A. Recall from Definition 3.18 and Remark 3.19 that

$$A = A(\mathbf{T}) = \begin{bmatrix} N_{\mathfrak{w}\mathfrak{w}} & N_{\mathfrak{b}\mathfrak{w}} \\ N_{\mathfrak{w}\mathfrak{b}} & N_{\mathfrak{b}\mathfrak{b}} \end{bmatrix},$$

where $N_{\mathfrak{c}\mathfrak{c}'} \coloneqq \operatorname{card}\bigl(\bigl\{X \in \mathbf{T} : X \in \mathbf{X}^R_{\mathfrak{c}}(f,\mathcal{C}),\, X \subseteq X^0_{\mathfrak{c}'}\bigr\}\bigr)$ for each pair of $\mathfrak{c}$, $\mathfrak{c}' \in \{\mathfrak{b},\mathfrak{w}\}$. In particular, since $f(\mathcal{C}) \subseteq \mathcal{C}$, by (6.23) and Proposition 3.17 (ii), one has $N_{\mathfrak{c}\mathfrak{b}} + N_{\mathfrak{c}\mathfrak{w}} = \operatorname{card}(\mathbf{T})/2$ for each $\mathfrak{c} \in \{\mathfrak{b},\mathfrak{w}\}$. Then by some elementary calculations in linear algebra we obtain $\rho(A) = \operatorname{card}(\mathbf{T})/2$. Hence $h_{\mathrm{top}}(F) = \log(\operatorname{card}(\mathbf{T})/2)$ and $h_\nu(f) = (1/R)\log(\operatorname{card}(\mathbf{T})/2)$.

By (6.11), (6.19), (6.15), (6.20), and (6.21), we have

$$\begin{aligned}
|h_\mu(f) - h_\nu(f)| &\leqslant \frac{\log 2}{R} + \sum_{i=0}^{s} \Big|\rho_i h_{\mu_i} - \frac{1}{R} r_i \log\big(\mathrm{card}\big(\mathbf{T}^{n_j}\big)\big)\Big| \\
&\leqslant \frac{\log 2}{R} + \sum_{i=0}^{s} h_{\mu_i}(f)\Big|\rho_i - \frac{n_i r_i}{R}\Big| + \sum_{i=0}^{s} \frac{n_i r_i}{R}\Big|h_{\mu_i} - \frac{1}{n_i}\log\big(\mathrm{card}\big(\mathbf{T}^{n_j}\big)\big)\Big| \\
&\leqslant \frac{\log 2}{R} + \sum_{i=0}^{s} h_{\mu_i}(f)\Big|\rho_i - \frac{n_i r_i}{R}\Big| + \frac{\varepsilon}{2R}\sum_{i=0}^{s} n_i r_i \\
&\leqslant \frac{\varepsilon}{6} + \frac{\varepsilon}{3} + \frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

Recall that $R = sN + \sum_{i=1}^{s} n_i r_i$ and each tile X^R in $\mathbf{T}$ has the form in (6.18), i.e.,

$$X^R = \big[w(Y_1)\lambda(Y_1, Y_2)w(Y_2)\lambda(Y_2, Y_3)w(Y_3)\cdots\lambda(Y_{s-1}, Y_s)w(Y_s)\lambda(Y_s, Y_{s+1})\big]$$

with $Y_{s+1} = Y_1$ and $Y_i \in \mathbf{T}^{n_i, r_i}$ for each $i \in \{1, \ldots, s\}$. Here $\mathbf{T}^{n_i, r_i}$ is a non-empty subset of $\mathbf{X}^{M_i}(f, \mathcal{C})$ with $M_i = n_i r_i + N r_i - N$. By (6.17) and (6.16), for each $i \in \{1, \ldots, s\}$,

$$\begin{aligned}
&\sup_{x \in [w(Y_i)\lambda(Y_i, Y_{i+1})]} \Big\| S_{M_i+N}\vec{\varphi}(x) - R\rho_i \int \vec{\varphi}\,\mathrm{d}\mu_i \Big\| \\
&\quad\leqslant \sup_{x \in [w(Y_i)\lambda(Y_i, Y_{i+1})]} \Big\| S_{M_i+N}\vec{\varphi}(x) - n_i r_i \int \vec{\varphi}\,\mathrm{d}\mu_i \Big\| + R\Big|\rho_i - \frac{n_i r_i}{R}\Big| \Big\|\int \vec{\varphi}\,\mathrm{d}\mu_i\Big\| \\
&\quad\leqslant \frac{\varepsilon}{3} n_i r_i + r_i \sup_{\mathfrak{c}, \mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}} \sup_{x \in [\lambda_{\mathfrak{c}\mathfrak{c}'}]} \|S_N \vec{\varphi}(x)\| + R\Big|\rho_i - \frac{n_i r_i}{R}\Big| \Big\|\int \vec{\varphi}\,\mathrm{d}\mu_i\Big\|.
\end{aligned}$$

Summing this over all $i \in \{1, \ldots, s\}$, dividing the result by R, and then using (6.21) and (6.22) yield

$$\sup_{x \in X^R} \Big\| \frac{1}{R} S_R \vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu \Big\| \leqslant \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

for each $X^R \in \mathbf{T}$. This implies that

$$\sup_{x \in \Omega} \Big\| \frac{1}{R} S_R \vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu \Big\| \leqslant \sup_{x \in \bigcup \mathbf{T}} \Big\| \frac{1}{R} S_R \vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu \Big\| \leqslant \varepsilon.$$

Note that $\operatorname{supp} \widehat{\nu} \subseteq \Omega$. Then it follows from (6.24) that

$$\Big\| \int \vec{\varphi}\,\mathrm{d}\nu - \int \vec{\varphi}\,\mathrm{d}\mu \Big\| = \Big\| \frac{1}{R} \int S_R \vec{\varphi}\,\mathrm{d}\widehat{\nu} - \int \vec{\varphi}\,\mathrm{d}\mu \Big\| \leqslant \varepsilon.$$

This shows that the ergodic measure ν fulfills our requirements (see the beginning of the proof) and completes the proof. □

7. Large deviation principles

7.1. **Level-2 large deviation principles.** In this subsection we review some basic concepts and results from large deviation theory. We refer the reader to [DZ09, Ell12, RAS15] for a systematic and detailed introduction.

Let $\{\xi_n\}_{n \in \mathbb{N}}$ be a sequence of Borel probability measures on a Hausdorff topological space $\mathcal{X}$. We say that $\{\xi_n\}_{n \in \mathbb{N}}$ satisfies a *large deviation principle* in $\mathcal{X}$ if there exists a lower semi-continuous function $I \colon \mathcal{X} \to [0, +\infty]$ such that

$$\liminf_{n \to +\infty} \frac{1}{n} \log \xi_n(\mathcal{G}) \geqslant -\inf_{\mathcal{G}} I \qquad \text{for all open } \mathcal{G} \subseteq \mathcal{X}, \tag{7.1}$$

and

$$\limsup_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{K}) \leqslant -\inf_{\mathcal{K}} I \qquad \text{for all closed } \mathcal{K} \subseteq \mathcal{X}, \tag{7.2}$$

where $\log 0 = -\infty$ and $\inf \emptyset = +\infty$ by convention. Such a function I is called a *rate function*, and we say that I is a *good rate function* if the set $\{x \in \mathcal{X} : I(x) \leqslant \alpha\}$ is compact for every $\alpha \in [0, +\infty)$. If $\mathcal{X}$ is regular, then the rate function I is unique.

A Borel set $\mathcal{A} \subseteq \mathcal{X}$ is called a *I-continuity set* if

$$\inf\{I(x) : x \in \operatorname{int}(\mathcal{A})\} = \inf\{I(x) : x \in \overline{\mathcal{A}}\}.$$

When (7.1) and (7.2) hold then for each I-continuity set $\mathcal{A} \subseteq \mathcal{X}$ the limit $\lim_{n\to+\infty} n^{-1} \log \xi_n(\mathcal{A})$ exists and satisfies

$$\lim_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{A}) = -\inf_{\mathcal{A}} I, \tag{7.3}$$

and we can replace $\mathcal{A}$ by either its interior or its closure. When only (7.1) (resp. (7.2)) is satisfied, we say that the *large deviation lower (resp. upper) bounds* hold with the function I.

We call $x \in \mathcal{X}$ a *minimizer* if $I(x) = 0$ holds. The set of minimizers is a closed set. For a closed subset $\mathcal{K}$ of $\mathcal{X}$ that is disjoint from the set of minimizers, the large deviation principle ensures that $\xi_n(\mathcal{K})$ decays exponentially as $n \to +\infty$. If moreover I is a good rate function, the support of any accumulation point of $\{\xi_n\}_{n\in\mathbb{N}}$ is contained in the set of minimizers. Hence, it is important to determine the set of minimizers. The non-uniqueness of minimizers is referred to as a phase transition. The uniqueness of minimizers implies several strong conclusions.

The following *contraction principle* shows that the large deviation principle transfers nicely through continuous functions.

Theorem 7.1 (Contraction principle [DZ09, Theorem 4.2.1])**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces, and let $g\colon \mathcal{X} \to \mathcal{Y}$ be a continuous map. Consider a sequence $\{\xi_n\}_{n\in\mathbb{N}}$ of Borel probability measures on $\mathcal{X}$ that satisfies a large deviation principle in $\mathcal{X}$ with a good rate function $I\colon \mathcal{X} \to [0, +\infty]$. For each $y \in \mathcal{Y}$, define*

$$J(y) \coloneqq \inf\{I(x) : x \in \mathcal{X},\, y = g(x)\}.$$

Then J is a good rate function on $\mathcal{Y}$, and the sequence $\{g_*(\xi_n)\}_{n\in\mathbb{N}}$ satisfies a large deviation principle in $\mathcal{Y}$ with the rate function $J\colon \mathcal{Y} \to [0, +\infty]$.

The above notations will be applied with $\mathcal{X} = \mathcal{P}(X)$ (for some compact metric space X), $\mathcal{Y} = \mathbb{R}$, and $g = \widehat{\psi}$ for some $\psi \in C(X)$, where $\widehat{\psi}$ is the evaluation map on $\mathcal{P}(X)$ (i.e., $\widehat{\psi}(\mu) = \int \psi \,\mathrm{d}\mu$). In this context, the large deviation principles in $\mathcal{P}(X)$, are usually referred to as "level-2", and the ones in $\mathbb{R}$ (in particular those obtained by contraction) as "level-1".

7.2. **Uniqueness of the minimizer.** In this subsection, we prove that μ_ϕ is the unique minimizer of the rate function I_ϕ defined in (1.4). Recall that we call $\mu \in \mathcal{P}(S^2)$ a minimizer of I_ϕ if $I_\phi(\mu) = 0$.

Definition 7.2. Let X and $\widetilde{X}$ be topological spaces, and $T\colon X \to X$ and $\widetilde{T}\colon \widetilde{X} \to \widetilde{X}$ be continuous maps. We say that T is a *factor* (or *topological factor*) of $\widetilde{T}$ if there exists a surjective continuous map $\pi\colon \widetilde{X} \to X$ such that $\pi \circ \widetilde{T} = T \circ \pi$. Such a map π is called a *semi-conjugacy*.

For an expanding Thurston map, by the results in [DPTUZ21], we have the following proposition, which gives a semi-conjugacy with the one-sided shift map.

Proposition 7.3. *Let f and d satisfy the Assumptions in Section 4. Let $\sigma\colon \Sigma \to \Sigma$ be the one-sided shift map on $\deg f$ symbols. Then there exists a semi-conjugacy $\pi\colon \Sigma \to S^2$ with $\pi \circ \sigma = f \circ \pi$ satisfying the following properties:*

(i) *For each Hölder continuous function $\phi\colon S^2 \to \mathbb{R}$ with respect to the metric d, the function $\phi \circ \pi\colon \Sigma \to \mathbb{R}$ is Hölder continuous with respect to the standard metric on Σ.*

(ii) *For each Hölder continuous function $\phi\colon S^2 \to \mathbb{R}$ with respect to the metric d, $P(f, \phi) = P(\sigma, \phi \circ \pi)$.*

(iii) *Let* $\widetilde{\mu} \in \mathcal{M}(\Sigma, \sigma)$ *be an equilibrium state for the map* σ *and the potential* $\phi \circ \pi$. *Denote* $\mu := \pi_* \widetilde{\mu}$. *Then* $h_\mu(f) = h_{\widetilde{\mu}}(\sigma)$.

Recall that the standard metric on the shift space Σ is given by $\rho(\xi, \eta) := 2^{-\min\{i \in \mathbb{N}_0 : \xi_i \neq \eta_i\}}$ for distinct sequences $\xi = \{\xi_i\}_{i \in \mathbb{N}_0}$ and $\eta = \{\eta_i\}_{i \in \mathbb{N}_0}$ in Σ.

Proposition 7.3 (i) follows immediately from [DPTUZ21, Lemmma 5.4]. Proposition 7.3 (ii) and (iii) was established in the proof of [DPTUZ21, Proposition 5.5].

The following lemma shows that one can "lift" invariant measures by a semi-conjugacy, whose proof is verbatim the same as that of [DPTUZ21, Lemma 4.1].

Lemma 7.4. *Let* X *and* $\widetilde{X}$ *be compact metrizable topological spaces, and* $T\colon X \to X$ *and* $\widetilde{T}\colon \widetilde{X} \to \widetilde{X}$ *be continuous maps. Suppose that* T *is a factor of* $\widetilde{T}$ *and* $\pi\colon \widetilde{T} \to T$ *is a semi-conjugacy with* $\pi \circ \widetilde{T} = T \circ \pi$. *Then for each* $\mu \in \mathcal{M}(X, T)$, *there exists* $\widetilde{\mu} \in \mathcal{M}(\widetilde{X}, \widetilde{T})$ *such that* $\pi_* \widetilde{\mu} = \mu$.

We now prove the uniqueness of the minimizer.

Theorem 7.5. *Let* f, d, ϕ, μ_ϕ *satisfy the Assumptions in Section 4. Then* μ_ϕ *is the unique minimizer of the rate function* I_ϕ *defined in* (1.4).

Proof. It follows immediately from (1.4) and (1.5) that $I_\phi(\mu_\phi) = 0$, i.e., μ_ϕ is a minimizer of I_ϕ. It suffices to show the uniqueness.

Suppose that $\mu_* \in \mathcal{P}(S^2)$ is a minimizer of I_ϕ, i.e., $I_\phi(\mu_*) = 0$. Then by (1.4) and (1.5), there exists a sequence $\{\mu_n\}_{n \in \mathbb{N}}$ of f-invariant Borel probability measures that converges to μ_* in the weak*-topology with $F_\phi(\mu_n) \to 0$ as $n \to +\infty$. In particular, this implies $\mu_* \in \mathcal{M}(S^2, f)$ and $\lim_{n \to +\infty} h_{\mu_n}(f) = P(f, \phi) - \int_{S^2} \phi \, \mathrm{d}\mu_*$ by (1.5).

Let $\sigma\colon \Sigma \to \Sigma$ be the one-sided shift map on $\deg f$ symbols and $\pi\colon \Sigma \to S^2$ be the semi-conjugacy given by Proposition 7.3. Then by Lemma 7.4, there exists a sequence $\{\widetilde{\mu}_n\}_{n \in \mathbb{N}}$ of σ-invariant Borel probability measures on Σ such that $\pi_* \widetilde{\mu}_n = \mu_n$ for each $n \in \mathbb{N}$. Since the space $\mathcal{M}(\Sigma, \sigma)$ is sequentially compact (in the weak*-topology), the sequence $\{\widetilde{\mu}_n\}_{n \in \mathbb{N}}$ has a convergent subsequence. Without loss of generality we may assume that the sequence $\{\widetilde{\mu}_n\}_{n \in \mathbb{N}}$ itself converges to $\widetilde{\mu}_* \in \mathcal{M}(\Sigma, \sigma)$ in the weak*-topology. Then we have $\mu_n = \pi_* \widetilde{\mu}_n \xrightarrow{w^*} \pi_* \widetilde{\mu}_*$ as $n \to +\infty$. This implies $\pi_* \widetilde{\mu}_* = \mu_*$ since $\mu_n \xrightarrow{w^*} \mu_*$ as $n \to +\infty$.

We now show that $\widetilde{\mu}_*$ is an equilibrium state for the shift map σ and the potential $\phi \circ \pi$. Since $\pi_* \widetilde{\mu}_* = \mu_*$, we have $\int_{S^2} \phi \, \mathrm{d}\mu_* = \int_\Sigma \phi \circ \pi \, \mathrm{d}\widetilde{\mu}_*$. Noting that for each $n \in \mathbb{N}$, the dynamical system (S^2, f, μ_n) is a factor of $(\Sigma, \sigma, \widetilde{\mu}_n)$, we have $h_{\mu_n}(f) \leqslant h_{\widetilde{\mu}_n}(\sigma)$ (see for example, [KH95, Proposition 4.3.16 (1)]). For the shift map σ, it is a classical result that the entropy map of σ is upper semi-continuous (see for example, [Wal82, Theorem 8.2]). This implies $h_{\widetilde{\mu}_*}(\sigma) \geqslant \limsup_{n \to +\infty} h_{\widetilde{\mu}_n}(\sigma) \geqslant \limsup_{n \to +\infty} h_{\mu_n}(f)$. Hence by the Variational principle, we have

$$P(\sigma, \phi \circ \pi) \geqslant h_{\widetilde{\mu}_*}(\sigma) + \int_\Sigma \phi \circ \pi \, \mathrm{d}\widetilde{\mu}_* \geqslant \limsup_{n \to +\infty} h_{\mu_n}(f) + \int_{S^2} \phi \, \mathrm{d}\mu_*.$$

Since $\lim_{n \to +\infty} h_{\mu_n}(f) = P(f, \phi) - \int_{S^2} \phi \, \mathrm{d}\mu_*$ (see the beginning of the proof), we deduce that

$$P(\sigma, \phi \circ \pi) \geqslant h_{\widetilde{\mu}_*}(\sigma) + \int_\Sigma \phi \circ \pi \, \mathrm{d}\widetilde{\mu}_* \geqslant P(f, \phi).$$

By Proposition 7.3 (ii), $P(f, \phi) = P(\sigma, \phi \circ \pi)$. Thus, $\widetilde{\mu}_*$ is an equilibrium state for the shift map σ and the potential $\phi \circ \pi$.

Finally, since $\mu_* = \pi_* \widetilde{\mu}_*$ and $\widetilde{\mu}_*$ is an equilibrium state for the shift map σ and the potential $\phi \circ \pi$, it follows from Proposition 7.3 (iii) that $h_{\mu_*}(f) = h_{\widetilde{\mu}_*}(\sigma)$. Therefore, we have

$$h_{\mu_*}(f) + \int_{S^2} \phi \, \mathrm{d}\mu_* = h_{\widetilde{\mu}_*}(\sigma) + \int_\Sigma \phi \circ \pi \, \mathrm{d}\widetilde{\mu}_* = P(\sigma, \phi \circ \pi) = P(f, \phi),$$

i.e., μ_* is an equilibrium state for the map f and the potential ϕ. This implies $\mu_* = \mu_\phi$ by the uniqueness of the equilibrium state (see Theorem 3.10 (i)). The proof is complete. □

7.3. **Characterizations of topological pressures.** In this subsection, we characterize the topological pressure in terms of periodic points and iterated preimages.

Lemma 7.6. *Let f and $\mathcal{C}$ satisfy the Assumptions in Section 4. Let $n_f \in \mathbb{N}$ be the constant from Definition 3.20, which depends only on f and $\mathcal{C}$. Consider arbitrary integer $m \geqslant n_f$, integer $n \in \mathbb{N}_0$, point $x \in S^2$, and n-tile $X^n \in \mathbf{X}^n(f,\mathcal{C})$. Then the following statements hold:*

(i) *If $X^n \subseteq X^0_{\mathfrak{c}}$ for some $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, then there exists a fixed point of f^{n+m} in* $\operatorname{inte}(X^n)$.

(ii) *There exists a preimage of x under f^{n+m} in* $\operatorname{inte}(X^n)$.

Proof. Let integer $m \geqslant n_f$, $n \in \mathbb{N}_0$, and $X^n \in \mathbf{X}^n(f,\mathcal{C})$ be arbitrary. Since $f \in \operatorname{Sub}(f,\mathcal{C})$ is strongly primitive, by Lemma 3.22, for each $\mathfrak{c}' \in \{\mathfrak{b}, \mathfrak{w}\}$ there exists $X^{n+m}_{\mathfrak{c}'} \in \mathbf{X}^{n+m}_{\mathfrak{c}'}(f,\mathcal{C})$ such that $X^{n+m}_{\mathfrak{c}'} \subseteq \operatorname{inte}(X^n)$.

(i) Suppose that $X^n \subseteq X^0_{\mathfrak{c}}$ for some $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$. By Proposition 3.2 (i), $f^{n+m}|_{X^{n+m}_{\mathfrak{c}}}$ is a homeomorphism of $X^{n+m}_{\mathfrak{c}}$ onto $X^0_{\mathfrak{c}}$. Note that $X^{n+m}_{\mathfrak{c}} \subseteq \operatorname{inte}(X^n) \subseteq X^0_{\mathfrak{c}}$. Then by applying Brouwer's Fixed Point Theorem (see for example, [Hat02, Theorem 1.9]) to the inverse of f^{n+m} restricted to $X^{n+m}_{\mathfrak{c}}$, we get a fixed point $x \in X^{n+m}_{\mathfrak{c}} \subseteq \operatorname{inte}(X^n)$ of f^{n+m}.

(ii) Since $X^{n+m}_{\mathfrak{b}} \cup X^{n+m}_{\mathfrak{w}} \subseteq \operatorname{inte}(X^n)$, it follows from Proposition 3.2 (i) that $x \in S^2 = f^{n+m}(X^{n+m}_{\mathfrak{b}}) \cup f^{n+m}(X^{n+m}_{\mathfrak{b}}) \subseteq f^{n+m}(\operatorname{inte}(X^n))$. This implies that there exists $y \in f^{-n-m}(x)$ such that $y \in \operatorname{inte}(X^n)$. ☐

We recall the following characterizations of topological pressure in terms of periodic points and iterated preimages (see [Li15, Propositions 6.8 and 6.7], respectively).

Proposition 7.7 (Z. Li [Li15])**.** *Let f, d, ϕ satisfy the Assumptions in Section 4. Fix an arbitrary sequence $\{w_n\}_{n\in\mathbb{N}}$ of real-valued functions on S^2 satisfying $w_n(x) \in \left[1, \deg_{f^n}(x)\right]$ for each $n \in \mathbb{N}$ and each $x \in S^2$. Then*

$$P(f,\phi) = \lim_{n\to+\infty} \frac{1}{n} \log \sum_{x\in\operatorname{Per}_n(f)} w_n(x) \exp(S_n\phi(x)).$$

Proposition 7.8 (Z. Li [Li15])**.** *Let f, d, ϕ satisfy the Assumptions in Section 4. Then for each sequence $\{x_n\}_{n\in\mathbb{N}}$ in S^2, we have*

$$P(f,\phi) = \lim_{n\to+\infty} \frac{1}{n} \log \sum_{y\in f^{-n}(x_n)} \deg_{f^n}(y) \exp(S_n\phi(y)).$$

If we also assume that f has no periodic critical points, then for an arbitrary sequence $\{w_n\}_{n\in\mathbb{N}}$ of real-valued functions on S^2 satisfying $w_n(x) \in \left[1, \deg_{f^n}(x)\right]$ for each $n \in \mathbb{N}$ and each $x \in S^2$, we have

$$P(f,\phi) = \lim_{n\to+\infty} \frac{1}{n} \log \sum_{y\in f^{-n}(x_n)} w_n(y) \exp(S_n\phi(y)).$$

We now prove a generalization of Proposition 7.8 by removing the assumption on periodic critical points.

Proposition 7.9. *Let f, d, ϕ satisfy the Assumptions in Section 4. Fix an arbitrary sequence $\{w_n\}_{n\in\mathbb{N}}$ of real-valued functions on S^2 satisfying $w_n(x) \in \left[1, \deg_{f^n}(x)\right]$ for each $n \in \mathbb{N}$ and each $x \in S^2$. Then for each sequence $\{x_n\}_{n\in\mathbb{N}}$ in S^2, we have*

$$P(f,\phi) = \lim_{n\to+\infty} \frac{1}{n} \log \sum_{y\in f^{-n}(x_n)} w_n(y) \exp(S_n\phi(y)).$$

Proof. By Proposition 7.8, it suffices to show that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{y\in f^{-n}(x_n)} \deg_{f^n}(y) \exp(S_n\phi)(y) \leqslant \liminf_{m\to+\infty} \frac{1}{m} \sum_{\widehat{y}\in f^{-m}(x_m)} \exp(S_m\phi(\widehat{y})). \tag{7.4}$$

We fix a Jordan curve $\mathcal{C} \subseteq S^2$ that satisfies the Assumptions in Section 4. Suppose that $\phi \in C^{0,\beta}(S^2, d)$ is a real-valued Hölder continuous function with exponent $\beta \in (0,1]$.

Let $N \coloneqq n_f \in \mathbb{N}$ be the constant from Definition 3.20, which depends only on f and $\mathcal{C}$. For each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, it follows from Lemma 7.6 (ii) that there exists a preimage of x_{n+N} under f^{n+N} in $\operatorname{inte}(X^n)$. We fix a preimage of of x_{n+N} under f^{n+N} in $\operatorname{inte}(X^n)$ and denote it by $\widehat{y}(X^n)$. Then for each $n \in \mathbb{N}$, the map $X^n \mapsto \widehat{y}(X^n)$ from $\mathbf{X}^n(f,\mathcal{C})$ to $f^{-n-N}(x_{n+N})$ is injective.

For each $n \in \mathbb{N}$ and each $y \in f^{-n}(x_n)$, let $X^n(y) \in \mathbf{X}^n(f,\mathcal{C})$ be an n-tile that contains y. By Proposition 3.2 (i), for each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, $f^n|_{X^n}$ is a homeomorphism of X^n onto $f^n(X^n)$. This implies that for each integer $n \in \mathbb{N}$, the map $y \mapsto \widehat{y}(X^n(y))$ from $f^{-n}(x_n)$ to $f^{-n-N}(x_{n+N})$ is injective, where $\widehat{y}(X^n(y)) \in \operatorname{inte}(X^n(y))$.

Let $n \in \mathbb{N}$ be arbitrary. Consider $y' \in f^{-n}(x_n) \cap \mathbf{V}^n$, where $\mathbf{V}^n = \mathbf{V}^n(f,\mathcal{C})$ is the set of n-vertices. We set $\mathbf{X}^n(f,\mathcal{C},y') \coloneqq \{X \in \mathbf{X}^n(f,\mathcal{C}) : y' \in X\}$. By Remark 3.4, we have $\overline{W}^n(y') = \bigcup \mathbf{X}^n(f,\mathcal{C},y')$ and $\operatorname{card}(\mathbf{X}^n(f,\mathcal{C},y')) = 2\deg_{f^n}(y')$, where $W^n(y')$ is defined in (3.6) and $\overline{W}^n(y')$ is the closure of $W^n(y')$. Note that $X^n(y) \notin \mathbf{X}^n(f,\mathcal{C},y')$ for every $y \in f^{-n}(x_n) \setminus \mathbf{V}^n(f,\mathcal{C})$.

We now establish (7.4). By the arguments above, for each integer $n \in \mathbb{N}$, we have

$$
\begin{aligned}
&\sum_{y \in f^{-n}(x_n)} \deg_{f^n}(y) \exp(S_n\phi(y)) \\
&\quad\leqslant \sum_{y' \in f^{-n}(x_n)\cap \mathbf{V}^n} \deg_{f^n}(y') \exp(S_n\phi(y')) + \sum_{y \in f^{-n}(x_n)\setminus \mathbf{V}^n} \exp(S_n\phi(y)) \\
&\quad\leqslant \sum_{y' \in f^{-n}(x_n)\cap \mathbf{V}^n} \sum_{X^n \in \mathbf{X}^n(f,\mathcal{C},y')} e^{D_n(\phi) + S_n\phi(\widehat{y}(X^n))} + \sum_{y \in f^{-n}(x_n)\setminus \mathbf{V}^n} e^{D_n(\phi) + S_n\phi(\widehat{y}(X^n(y)))} \\
&\quad\leqslant e^{D_n(\phi)} \sum_{\widehat{y} \in f^{-n-N}(x_{n+N})} \exp(S_n\phi(\widehat{y})) \\
&\quad\leqslant e^{D_n(\phi)} e^{N\|\phi\|_\infty} \sum_{\widehat{y} \in f^{-n-N}(x_{n+N})} \exp(S_{n+N}\phi(\widehat{y})).
\end{aligned}
$$

Then by Lemma 3.9, we get

$$
\frac{1}{n}\log \sum_{y \in f^{-n}(x_n)} \deg_{f^n}(y)\exp(S_n\phi(y)) \leqslant \frac{C}{n} + \frac{1}{n}\log \sum_{\widehat{y} \in f^{-n-N}(x_{n+N})} \exp(S_n\phi(\widehat{y})),
$$

where $C \coloneqq N\|\phi\|_\infty + C_1(\operatorname{diam}_d(S^2))^\beta$ and $C_1 \geqslant 0$ is the constant defined in (3.10) in Lemma 3.9 and depends only on f, $\mathcal{C}$, d, ϕ, and β. Letting $n \to +\infty$ yields the desired inequality. $\square$

7.4. **Large deviation lower bound.** This subsection is devoted to the proof of the lower bound (7.1) for all open sets, with the main result being Proposition 7.10. In Section 7.4.1 we show that the proof of the lower bound can be reduced to the case where the invariant measure in question is ergodic. In Section 7.4.2 we prove lower bounds for certain fundamental open subsets of $\mathcal{P}(S^2)$, where we apply Lemma 6.2 to approximate each ergodic measure with a collection of tiles. Finally, in Section 7.4.3 we establish Proposition 7.10.

Proposition 7.10. *Let f, d, ϕ satisfy the Assumptions in Section 4. Then for each sequence $\{\xi_n\}_{n\in\mathbb{N}} \in \bigl\{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\bigr\}$ (as defined in Theorem 1.3), we have*

$$
\liminf_{n\to+\infty} \frac{1}{n}\log \xi_n(\mathcal{G}) \geqslant -\inf_{\mathcal{G}} I_\phi \qquad \text{for all open } \mathcal{G} \subseteq \mathcal{P}(S^2),
$$

where $I_\phi \colon \mathcal{P}(S^2) \to [0,+\infty]$ is defined in (1.4).

7.4.1. *Reduction to ergodic measures.* A weaker property related to entropy density (defined in Subsection 1.1) is entropy approachablility (see Definition 7.11). Entropy approachablility is a useful property in theories such as multifractal analysis and large deviations in which all invariant measures come into play, in order to reduce one's consideration to ergodic measures only.

Definition 7.11. Let (X, d) be a compact metric space and $T\colon X \to X$ be a continuous map. We say that a measure $\mu \in \mathcal{M}(S^2, f)$ is *entropy-approachable by ergodic measures* if for each $\varepsilon > 0$ and each weak*-open set U containing μ there exists an ergodic measure $\nu \in U \cap \mathcal{M}(X, T)$ such that $h_\nu(T) > h_\mu(T) - \varepsilon$.

Remark 7.12. It is clear that if ergodic measures are entropy-dense, then any invariant measure is entropy-approachable by ergodic measures. One sees that these two notions are equivalent when the entropy map is upper semi-continuous.

It follows immediately from Theorem 1.2 and Remark 7.12 that for expanding Thurston maps, any invariant measure is entropy-approachable by ergodic measures.

Corollary 7.13. *For an expanding Thurston map $f\colon S^2 \to S^2$, any invariant measure $\mu \in \mathcal{M}(S^2, f)$ is entropy-approachable by ergodic measures.*

Remark 7.14. For a continuous map $T\colon X \to X$ on a compact metric space (X, d), it is known that if T has the specification property in the sense of K. Sigmund (see the definition in [Sig74, Section 2]), then any invariant measure is entropy-approachable by ergodic measures (see for example, [EKW94] and [PS05, Theorem 2.1]). In particular, this result applies to expanding Thurston maps since every expanding Thurston map has the specification property (see the proof of [LZ24, Lemma 6.5]).

7.4.2. *Lower bound for fundamental open sets.* We use the notations as introduced in the beginning of Section 6.

We first prove the following result under the additional assumption that there exists an f-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} f \subseteq \mathcal{C}$ and then for the general case.

Proposition 7.15. *Let f, d, ϕ, β satisfy the Assumptions in Section 4. Consider $\ell \in \mathbb{N}$, $\vec{\varphi} \in C(S^2)^\ell$, and $\vec{\alpha} \in \mathbb{R}^\ell$. Let $\mathcal{G} \subseteq \mathcal{P}(S^2)$ be an open set of the form*

$$\mathcal{G} := \left\{ \mu \in \mathcal{P}(S^2) : \int \vec{\varphi} \,\mathrm{d}\mu > \vec{\alpha} \right\}.$$

Then for each $\mu \in \mathcal{G}$ and each sequence $\{\xi_n\}_{n\in\mathbb{N}} \in \big\{ \{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}} \big\}$ (as defined in Theorem 1.3), we have

$$\liminf_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{G}) \geqslant F_\phi(\mu),$$

where $F_\phi\colon \mathcal{P}(S^2) \to [-\infty, 0]$ is defined in (1.5).

Proof of Proposition 7.15 under an additional assumption. We assume in addition that there exists an f-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} f \subseteq \mathcal{C}$.

Let $\mu \in \mathcal{G}$ be arbitrary. We may assume without loss of generality that $\mu \in \mathcal{M}(S^2, f)$ since $F_\phi(\mu) = -\infty$ when $\mu \notin \mathcal{M}(S^2, f)$. Moreover, by virtue of Corollary 7.13 and the definition of F_ϕ, we may assume that μ is ergodic.

Let $\varepsilon > 0$ be such that

$$\int \vec{\varphi} \,\mathrm{d}\mu > \vec{\alpha} + \varepsilon. \tag{7.5}$$

By Lemma 6.2, there exists $n_0 \in \mathbb{N}$ such that for each integer $n \geqslant n_0$, there exists a non-empty subset $\mathbf{T}^n$ of $\mathbf{X}^n(f, \mathcal{C})$ such that

$$\left| \frac{1}{n} \log \operatorname{card}(\mathbf{T}^n) - h_\mu(f) \right| \leqslant \frac{\varepsilon}{2}, \tag{7.6}$$

$$\sup_{x \in \bigcup \mathbf{T}^n} \left\| \frac{1}{n} S_n \vec{\varphi}(x) - \int \vec{\varphi} \,\mathrm{d}\mu \right\| \leqslant \frac{\varepsilon}{2}, \qquad \text{and} \tag{7.7}$$

$$\sup_{x \in \bigcup \mathbf{T}^n} \left| \frac{1}{n} S_n \phi(x) - \int \phi \,\mathrm{d}\mu \right| \leqslant \frac{\varepsilon}{2}. \tag{7.8}$$

We split the rest of the proof into three cases according to the type of the sequence $\{\xi_n\}_{n\in\mathbb{N}}$.

Case 1 (Birkhoff averages): $\xi_n = \Sigma_n = (V_n)_*(\mu_\phi)$ for each $n \in \mathbb{N}$ (recall (1.1)).

For each integer $n \geqslant n_0$, (7.5) and (7.7) yield

$$\Big\{ V_n(x) : x \in \bigcup \mathbf{T}^n \Big\} \subseteq \mathcal{G}. \tag{7.9}$$

Recall from Proposition 3.11 that μ_ϕ is a Gibbs measure with respect to f, $\mathcal{C}$, and ϕ, with the constants $P_{\mu_\phi} = P(f,\phi)$ and $C_{\mu_\phi} \geqslant 1$. Then for each integer $n \geqslant n_0$ and each $X^n \in \mathbf{T}^n$, it follows from (3.11) in Proposition 3.11 and (7.8) that

$$\mu_\phi(X^n) \geqslant C_{\mu_\phi}^{-1} e^{-nP(f,\phi)} \inf_{x \in X^n} \exp(S_n\phi(x)) \geqslant C_{\mu_\phi}^{-1} e^{-nP(f,\phi)} \exp\biggl(n \biggl(\int \phi \,\mathrm{d}\mu - \frac{\varepsilon}{2} \biggr) \biggr).$$

Summing this inequality over all $X^n \in \mathbf{T}^n$ and applying (7.9), Theorem 3.10 (ii), and (7.6), we have

$$\begin{aligned}
\frac{1}{n} \log \Sigma_n(\mathcal{G}) &= \frac{1}{n} \log \mu_\phi(\{ x \in S^2 : V_n(x) \in \mathcal{G} \}) \\
&\geqslant \frac{1}{n} \log \mu_\phi \Bigl(\bigcup \mathbf{T}^n \Bigr) \\
&\geqslant \frac{1}{n} \log \Bigl(\operatorname{card}(\mathbf{T}^n) \inf_{X^n \in \mathbf{T}^n} \mu_\phi(X^n) \Bigr) \\
&\geqslant h_\mu(f) - \frac{\varepsilon}{2} + \int \phi \,\mathrm{d}\mu - \frac{\varepsilon}{2} - P(f,\phi) - \frac{1}{n} \log C_{\mu_\phi} \\
&= F_\phi(\mu) - \varepsilon - \frac{1}{n} \log C_{\mu_\phi}.
\end{aligned}$$

Letting $n \to +\infty$ and then $\varepsilon \to 0$ yields the desired inequality.

Case 2 (Periodic points): $\xi_n = \Omega_n$ for each $n \in \mathbb{N}$ (recall (1.2)).

By Proposition 3.8 and Lemma 7.6 (i), there exists a constant $N \in \mathbb{N}$ depending only on f and $\mathcal{C}$ such that for each $n \in \mathbb{N}_0$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, there exists a fixed point of f^{n+N} in $\operatorname{inte}(X^n)$. For each $n \in \mathbb{N}_0$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, let $p(X^n)$ be a fixed point of f^{n+N} in $\operatorname{inte}(X^n)$. Then the map $X^n \mapsto p(X^n)$ from $\mathbf{X}^n(f,\mathcal{C})$ to $\operatorname{Per}_{n+N}(f)$ is injective.

By (7.5) and (7.7), for each integer $n \geqslant n_0$ and each $x \in \bigcup \mathbf{T}^n$, we have

$$S_{n+N}\vec{\varphi}(x) \geqslant S_n\vec{\varphi}(x) - N\|\vec{\varphi}\| > n\vec{\alpha} + \frac{\varepsilon}{2}n - N\|\vec{\varphi}\| \geqslant (n+N)\vec{\alpha} + \frac{\varepsilon}{2}n - N(\|\vec{\varphi}\| + \|\vec{\alpha}\|).$$

This implies that for each sufficiently large $n \in \mathbb{N}$ and each $x \in \bigcup \mathbf{T}^n$, $(n+N)^{-1} S_{n+N}\vec{\varphi}(x) > \vec{\alpha}$, i.e., $V_{n+N}(x) \in \mathcal{G}$. Therefore, it follows from (7.6) and (7.8) that for each sufficiently large $n \in \mathbb{N}$,

$$\begin{aligned}
\sum_{\substack{p \in \operatorname{Per}_{n+N}(f) \\ V_{n+N}(p) \in \mathcal{G}}} w_{n+N}(p) \exp(S_{n+N}\phi(p)) &\geqslant \sum_{X^n \in \mathbf{T}^n} \exp(S_{n+N}\phi(p(X^n))) \\
&\geqslant \operatorname{card}(\mathbf{T}^n) \inf_{x \in \bigcup \mathbf{T}^n} \exp(S_n\phi(x)) \exp(-N\|\phi\|_\infty) \\
&\geqslant \exp(F_\phi(\mu)n + P(f,\phi)n - \varepsilon n - N\|\phi\|_\infty),
\end{aligned} \tag{7.10}$$

where $\{w_j\}_{j \in \mathbb{N}}$ is an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \bigl[1, \deg_{f^j}(x)\bigr]$ for each $j \in \mathbb{N}$ and each $x \in S^2$. By (1.2), for each $n \in \mathbb{N}$,

$$\begin{aligned}
\frac{1}{n+N} \log \Omega_{n+N}(\mathcal{G}) = &- \frac{1}{n+N} \log \sum_{p' \in \operatorname{Per}_{n+N}(f)} w_{n+N}(p') \exp(S_n\phi(p')) \\
&+ \frac{1}{n+N} \log \sum_{\substack{p \in \operatorname{Per}_{n+N}(f) \\ V_{n+N}(x) \in \mathcal{G}}} w_{n+N}(p) \exp(S_{n+N}\phi(p)).
\end{aligned}$$

Note that as $n \to +\infty$, the first term of the right hand side in the equation above converges to $-P(f,\phi)$ by Proposition 7.7. Combining this with (7.10), we get

$$\liminf_{n\to+\infty} \frac{1}{n} \log \Omega_n(\mathcal{G}) = \liminf_{n\to+\infty} \frac{1}{n+N} \log \Omega_{n+N}(\mathcal{G}) \geqslant F_\phi(\mu) - \varepsilon.$$

Then by letting $\varepsilon \to 0$, the desired inequality follows.

Case 3 (Iterated preimages): $\xi_n = \Omega_n(x_n)$ for each $n \in \mathbb{N}$ (recall (1.3)), where $\{x_n\}_{n\in\mathbb{N}}$ is an arbitrary sequence of points in S^2.

By Lemma 7.6 (ii), there exists a constant $N \in \mathbb{N}$ depending only on f and $\mathcal{C}$ such that for each $n \in \mathbb{N}_0$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, there exists $x \in f^{-n-N}(x_{n+N})$ such that $x \in \operatorname{inte}(X^n)$. For each $n \in \mathbb{N}_0$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, let x_{X^n} be a preimage point of x_{n+N} under f^{n+N} such that $x_{X^n} \in \operatorname{inte}(X^n)$. Then the map $X^n \mapsto x_{X^n}$ from $\mathbf{X}^n(f,\mathcal{C})$ to $f^{-n-N}(x_{n+N})$ is injective.

By the same reasoning as in Case 2, we have $V_{n+N}(x) \in \mathcal{G}$ for each sufficiently large $n \in \mathbb{N}$ and each $x \in \bigcup \mathbf{T}^n$. Similarly, it follows from (7.6) and (7.8) that for each sufficiently large $n \in \mathbb{N}$,

$$\begin{aligned}
\sum_{\substack{y\in f^{-n-N}(x_{n+N})\\ V_{n+N}(p)\in\mathcal{G}}} w_{n+N}(y)\exp(S_{n+N}\phi(y)) &\geqslant \sum_{X^n\in\mathbf{T}^n} \exp(S_{n+N}\phi(x_{X^n})) \\
&\geqslant \operatorname{card}(\mathbf{T}^n) \inf_{x\in\bigcup\mathbf{T}^n} \exp(S_n\phi(x))\exp(-N\|\phi\|_\infty) \\
&\geqslant \exp(F_\phi(\mu)n + P(f,\phi)n - \varepsilon n - N\|\phi\|_\infty),
\end{aligned} \tag{7.11}$$

where $\{w_j\}_{j\in\mathbb{N}}$ is an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \bigl[1, \deg_{f^j}(x)\bigr]$ for each $j \in \mathbb{N}$ and each $x \in S^2$. By (1.3), for each $n \in \mathbb{N}$,

$$\begin{aligned}
\frac{1}{n+N}\log\Omega_{n+N}(x_{n+N})(\mathcal{G}) = &-\frac{1}{n+N}\log \sum_{z\in f^{-n-N}(x_{n+N})} w_{n+N}(z)\exp(S_n\phi(z)) \\
&+\frac{1}{n+N}\log \sum_{\substack{y\in f^{-n-N}(x_{n+N})\\ V_{n+N}(p)\in\mathcal{G}}} w_{n+N}(y)\exp(S_{n+N}\phi(y)).
\end{aligned}$$

Note that as $n \to +\infty$, the first term of the right hand side in the equation above converges to $-P(f,\phi)$ by Proposition 7.9. Combining this with (7.11), we get

$$\liminf_{n\to+\infty} \frac{1}{n} \log \Omega_n(x_n)(\mathcal{G}) = \liminf_{n\to+\infty} \frac{1}{n+N} \log \Omega_{n+N}(x_{n+N})(\mathcal{G}) \geqslant F_\phi(\mu) - \varepsilon.$$

Then by letting $\varepsilon \to 0$, the desired inequality follows.

The proof is complete. □

We now prove the general case.

Proof of Proposition 7.15. Let $\mu \in \mathcal{G}$ be arbitrary. We may assume without loss of generality that $\mu \in \mathcal{M}(S^2,f)$ since $F_\phi(\mu) = -\infty$ when $\mu \notin \mathcal{M}(S^2,f)$.

By Lemma 3.7, we can find a sufficiently high iterate $\widehat{f} \coloneqq f^K$ of f that has an $\widehat{f}$-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post}\widehat{f} = \operatorname{post} f \subseteq \mathcal{C}$. Then $\widehat{f}$ is also an expanding Thurston map.

Denote $\vec{\Phi} \coloneqq S_K^f \vec{\varphi}$ and $\widehat{\phi} \coloneqq S_K^f \phi$. We define $\widehat{F}_{\widehat{\phi}}\colon \mathcal{P}(S^2) \to [-\infty, 0]$ by

$$\widehat{F}_{\widehat{\phi}}(\nu) \coloneqq \begin{cases} h_\nu(\widehat{f}) + \int \widehat{\phi}\,\mathrm{d}\nu - P(\widehat{f},\widehat{\phi}) & \text{if } \nu \in \mathcal{M}(S^2,\widehat{f}); \\ -\infty & \text{if } \nu \in \mathcal{P}(S^2)\setminus\mathcal{M}(S^2,\widehat{f}). \end{cases}$$

Note that $P(\widehat{f},\widehat{\phi}) = KP(f,\phi)$, $h_\mu(\widehat{f}) = Kh_\mu(f)$, and $\int \widehat{\phi}\,\mathrm{d}\mu = K\int\phi\,\mathrm{d}\mu$ (see Subsection 3.1). Then we have $\widehat{F}_{\widehat{\phi}}(\mu) = KF_\phi(\mu)$ since $\mu \in \mathcal{M}(S^2,f) \subseteq \mathcal{M}(S^2,\widehat{f})$. Let $\widehat{\mu}_{\widehat{\phi}}$ be the unique equilibrium state for the map $\widehat{f}$ and the potential $\widehat{\phi}$. Since $P_{\mu_\phi}(\widehat{f},\widehat{\phi}) = KP_{\mu_\phi}(f,\phi) = KP(f,\phi) = P(\widehat{f},\widehat{\phi})$, it follows from the uniqueness of the equilibrium state that $\widehat{\mu}_{\widehat{\phi}} = \mu_\phi$.

Let $\varepsilon > 0$ be such that $\int \vec{\varphi} \, \mathrm{d}\mu > \vec{\alpha} + \varepsilon$. This implies $\int \vec{\Phi} \, \mathrm{d}\mu = K \int \vec{\varphi} \, \mathrm{d}\mu > K\vec{\alpha} + K\varepsilon$. Let $\widehat{\mathcal{G}}_\varepsilon \subseteq \mathcal{P}(S^2)$ be the open set defined by

$$\widehat{\mathcal{G}}_\varepsilon \coloneqq \biggl\{ \nu \in \mathcal{P}(S^2) : \int \vec{\Phi} \, \mathrm{d}\nu > K\vec{\alpha} + K\varepsilon \biggr\}.$$

Then we have $\mu \in \widehat{\mathcal{G}}_\varepsilon$.

We split the proof into three cases according to the type of the sequence $\{\xi_n\}_{n\in\mathbb{N}}$.

Case 1 (Birkhoff averages): $\xi_n = \Sigma_n = (V_n)_*(\mu_\phi)$ for each $n \in \mathbb{N}$ (recall (1.1)).

For each integer $k \in \{0, \dots, K-1\}$ and each integer $m \in \mathbb{N}$ that satisfies $(\|\vec{\alpha}\| + \|\vec{\varphi}\|)/m < \varepsilon$, we have

$$\begin{aligned} \bigl\{ x \in S^2 : S^f_{mK+k} \vec{\varphi}(x) > (mK+k)\vec{\alpha} \bigr\} &\supseteq \bigl\{ x \in S^2 : S^f_{mK} \vec{\varphi}(x) > (mK+k)\vec{\alpha} + k\|\vec{\varphi}\| \bigr\} \\ &\supseteq \bigl\{ x \in S^2 : S^f_{mK} \vec{\varphi}(x) > mK\vec{\alpha} + K(\|\vec{\alpha}\| + \|\vec{\varphi}\|) \bigr\} \\ &\supseteq \bigl\{ x \in S^2 : m^{-1} S^{\widehat{f}}_m \vec{\Phi}(x) > K\vec{\alpha} + K\varepsilon \bigr\}. \end{aligned}$$

For each $n \in \mathbb{N}$, we set $m \coloneqq \lfloor n/K \rfloor$ and write $n = mK + k$ for some integer $k \in \{0, \dots, K-1\}$. Then for each sufficiently large $n \in \mathbb{N}$, we have

$$\begin{aligned} \log \Sigma_n(\mathcal{G}) &= \log \mu_\phi\bigl(\bigl\{ x \in S^2 : S^f_n \vec{\varphi}(x) > n\vec{\alpha} \bigr\} \bigr) \geqslant \log \widehat{\mu}_{\widehat{\phi}}\bigl(\bigl\{ x \in S^2 : m^{-1} S^{\widehat{f}}_m \vec{\Phi}(x) > K\vec{\alpha} + K\varepsilon \bigr\} \bigr) \\ &= \log \widehat{\Sigma}_m(\widehat{\mathcal{G}}_\varepsilon) \geqslant \frac{n}{mK} \log \widehat{\Sigma}_m(\widehat{\mathcal{G}}_\varepsilon), \end{aligned}$$

where $\bigl\{ \widehat{\Sigma}_j \bigr\}_{j\in\mathbb{N}}$ is defined by replacing f with $\widehat{f}$ and ϕ with $\widehat{\phi}$ in the definition of $\{\Sigma_j\}_{j\in\mathbb{N}}$. Since $\widehat{f}$ has an $\widehat{f}$-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} \widehat{f} \subseteq \mathcal{C}$, Proposition 7.15 holds for $\widehat{f}$. Therefore,

$$\liminf_{n\to+\infty} \frac{1}{n} \log \Sigma_n(\mathcal{G}) \geqslant \frac{1}{K} \liminf_{m\to+\infty} \frac{1}{m} \log \widehat{\Sigma}_m(\widehat{\mathcal{G}}_\varepsilon) \geqslant \frac{1}{K} \widehat{F}_{\widehat{\phi}}(\mu) = F_\phi(\mu).$$

Case 2 (Periodic points): $\xi_n = \Omega_n$ for each $n \in \mathbb{N}$ (recall (1.2)).

For each $m \in \mathbb{N}$, it follows from Proposition 3.2 (iv) that $\mathbf{X}^m(\widehat{f}, \mathcal{C}) = \mathbf{X}^{mK}(f, \mathcal{C})$. Since $\widehat{f}(\mathcal{C}) \subseteq \mathcal{C}$, by Proposition 3.8 and Lemma 7.6 (i), there exists a constant $N \in \mathbb{N}$ depending only on f and $\mathcal{C}$ such that for each integer $\ell \geqslant N$, each $m \in \mathbb{N}$, and each $X^{mK} \in \mathbf{X}^{mK}(f, \mathcal{C})$, there exists a fixed point of $f^{mK+\ell}$ in $\operatorname{inte}(X^{mK})$.

For each $m \in \mathbb{N}$, each $k \in \{0, \dots, K-1\}$, and each $\widehat{p} \in \operatorname{Per}_m(\widehat{f})$, let $X^{mK}(\widehat{p}) \in \mathbf{X}^{mK}(f, \mathcal{C})$ be an mK-tile that contains $\widehat{p}$ and let $p(k, \widehat{p})$ be a fixed point of f^{mK+k+N} in $\operatorname{inte}(X^{mK}(\widehat{p}))$. By [Li15, Lemma 6.3], there exists $N_0 \in \mathbb{N}$ such that for each integer $n \geqslant N_0$ and each n-tile $X^n \in \mathbf{X}^n(f, \mathcal{C})$, the number of fixed points of f^n contained in X^n is at most 1. This implies that for each integer $m \geqslant N_0/K$ and each $k \in \{0, \dots, K-1\}$, the map $\widehat{p} \mapsto p(k, \widehat{p})$ from $\operatorname{Per}_m(\widehat{f})$ to $\operatorname{Per}_{mK+k+N}(f)$ is injective.

We claim that there exists $n_0 \in \mathbb{N}$ such that for each integer $m \geqslant n_0$, each $k \in \{0, \dots, K-1\}$, and each $\widehat{p} \in \operatorname{Per}_m(\widehat{f})$ with $\widehat{V}_m(\widehat{p}) \in \widehat{\mathcal{G}}_\varepsilon$, it follows that $V_{mK+k+N}(p(k, \widehat{p})) \in \mathcal{G}$, where we define $\widehat{V}_\ell(x) \coloneqq \frac{1}{\ell} \sum_{i=0}^{\ell-1} \delta_{\widehat{f}^i(x)}$ for each $\ell \in \mathbb{N}$ and each $x \in S^2$. Indeed, by Lemma 6.1, there exists a sufficiently large $n_0 \in \mathbb{N}$ such that for each integer $m \geqslant n_0$,

$$D_{mK}(\vec{\varphi}) + (K+N)(\|\vec{\alpha}\| + \|\vec{\varphi}\|) \leqslant mK\varepsilon.$$

Since $X^{mK}(\widehat{p})$ contains $\widehat{p}$ and $p(k,\widehat{p})$, we have $S^f_{mK}\vec{\varphi}(p(k,\widehat{p})) \geqslant S^f_{mK}\vec{\varphi}(\widehat{p}) - D_{mK}(\vec{\varphi})$. Note that $\widehat{V}_m(\widehat{p}) \in \widehat{\mathcal{G}}_\varepsilon$ means that $m^{-1}S^{\widehat{f}}_m\vec{\Phi}(\widehat{p}) = m^{-1}S^f_{mK}\vec{\varphi}(\widehat{p}) > K\vec{\alpha} + K\varepsilon$. Therefore,

$$\begin{aligned} S^f_{mK+k+N}\vec{\varphi}(p(k,\widehat{p})) &\geqslant S^f_{mK}\vec{\varphi}(p(k,\widehat{p})) - (K+N)\|\vec{\varphi}\| \\ &\geqslant S^f_{mK}\vec{\varphi}(\widehat{p}) - D_{mK}(\vec{\varphi}) - (K+N)\|\vec{\varphi}\| \\ &> mK\vec{\alpha} + mK\varepsilon - D_{mK}(\vec{\varphi}) - (K+N)\|\vec{\varphi}\| \\ &\geqslant (mK+k+N)\vec{\alpha} + mK\varepsilon - D_{mK}(\vec{\varphi}) - (K+N)(\|\vec{\alpha}\| + \|\vec{\varphi}\|) \\ &\geqslant (mK+k+N)\vec{\alpha}. \end{aligned}$$

This implies $V_{mK+k+N}(p(k,\widehat{p})) \in \mathcal{G}$.

We now prove the lower bound.

For each integer $n \geqslant N$, we set $m \coloneqq \lfloor (n-N)/K \rfloor$ and write $n = mK + k + N$ for some integer $k \in \{0, \dots, K-1\}$. By the arguments above, for each sufficiently large $n \in \mathbb{N}$ that satisfies $m \geqslant \max\{N_0/K, n_0\}$, we have

$$\begin{aligned} \sum_{\substack{p\in \mathrm{Per}_n(f)\\ V_n(p)\in\mathcal{G}}} w_n(p)\exp\bigl(S^f_n\phi(p)\bigr) &\geqslant \sum_{\substack{p\in \mathrm{Per}_{mK+k+N}(f)\\ V_{mK+k+N}(p)\in\mathcal{G}}} \exp\bigl(S^f_{mK+k+N}\phi(p)\bigr) \\ &\geqslant \sum_{\substack{\widehat{p}\in \mathrm{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S^f_{mK+k+N}\phi(p(k,\widehat{p}))\bigr), \end{aligned}$$

where $\{w_j\}_{j\in\mathbb{N}}$ is an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \bigl[1, \deg_{f^j}(x)\bigr]$ for each $j \in \mathbb{N}$ and each $x \in S^2$. Then by Lemma 3.9,

$$\begin{aligned} \sum_{\substack{p\in \mathrm{Per}_n(f)\\ V_n(p)\in\mathcal{G}}} w_n(p)\exp\bigl(S^f_n\phi(p)\bigr) &\geqslant e^{-(K+N)\|\phi\|_\infty} \sum_{\substack{\widehat{p}\in \mathrm{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S^f_{mK}\phi(p(k,\widehat{p}))\bigr) \\ &\geqslant e^{-C} \sum_{\substack{\widehat{p}\in \mathrm{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S^f_{mK}\phi(\widehat{p})\bigr) = e^{-C} \sum_{\substack{\widehat{p}\in \mathrm{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S^{\widehat{f}}_m\widehat{\phi}(\widehat{p})\bigr), \end{aligned}$$

where $C \coloneqq (K+N)\|\phi\|_\infty + C_1(\mathrm{diam}_d(S^2))^\beta$ and $C_1 \geqslant 0$ is the constant defined in (3.10) in Lemma 3.9 that depends only on f, $\mathcal{C}$, d, ϕ, and β. Thus by (1.2), we have

$$\begin{aligned} \log\Omega_n(\mathcal{G}) &= \log \sum_{\substack{p\in \mathrm{Per}_n(f)\\ V_n(p)\in\mathcal{G}}} w_n(p)\exp\bigl(S^f_n\phi(p)\bigr) - \log \sum_{p'\in \mathrm{Per}_n(f)} w_n(p')\exp\bigl(S^f_n\phi(p')\bigr) \\ &\geqslant \log \sum_{\substack{\widehat{p}\in \mathrm{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S^{\widehat{f}}_m\widehat{\phi}(\widehat{p})\bigr) - \log \sum_{p'\in \mathrm{Per}_n(f)} w_n(p')\exp\bigl(S^f_n\phi(p')\bigr) - C \\ &= \log\widehat{\Omega}_m(\widehat{\mathcal{G}}_\varepsilon) + \log \sum_{\widehat{p}'\in \mathrm{Per}_m(\widehat{f})} \exp\bigl(S^{\widehat{f}}_m\widehat{\phi}(\widehat{p}')\bigr) - \log \sum_{p'\in \mathrm{Per}_n(f)} w_n(p')\exp\bigl(S^f_n\phi(p')\bigr) - C, \end{aligned}$$

where $\bigl\{\widehat{\Omega}_j\bigr\}_{j\in\mathbb{N}}$ is defined by setting $w_j(x) = 1$ for each $j \in \mathbb{N}$ and each $x \in S^2$ and replacing f with $\widehat{f}$ and ϕ with $\widehat{\phi}$ in the definition of $\{\Omega_j\}_{j\in\mathbb{N}}$ (recall (1.2)). Since $\widehat{f}$ has an $\widehat{f}$-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\mathrm{post}\,\widehat{f} \subseteq \mathcal{C}$, Proposition 7.15 holds for $\widehat{f}$. Therefore, by Proposition 7.7, we get

$$\begin{aligned} \liminf_{n\to+\infty} \frac{1}{n}\log\Omega_n(\mathcal{G}) &\geqslant \frac{1}{K}\liminf_{m\to+\infty}\frac{1}{m}\log\widehat{\Omega}_m(\widehat{\mathcal{G}}_\varepsilon) + \frac{1}{K}P(\widehat{f},\widehat{\phi}) - P(f,\phi) \\ &= \frac{1}{K}\liminf_{m\to+\infty}\frac{1}{m}\log\widehat{\Omega}_m(\widehat{\mathcal{G}}_\varepsilon) \geqslant \frac{1}{K}\widehat{F}_{\widehat{\phi}}(\mu) = F_\phi(\mu). \end{aligned}$$

Case 3 (Iterated preimages): $\xi_n = \Omega_n(x_n)$ for each $n \in \mathbb{N}$ (recall (1.3)), where $\{x_n\}_{n\in\mathbb{N}}$ is an arbitrary sequence of points in S^2.

By Lemma 7.6 (ii), there exists a constant $N \in \mathbb{N}$ depending only on f and $\mathcal{C}$ such that for each integer $\ell \geqslant N$, each $m \in \mathbb{N}$, and each $X^{mK} \in \mathbf{X}^{mK}(f, \mathcal{C})$, there exists a preimage of $x_{mK+\ell}$ under $f^{mK+\ell}$ in $\operatorname{inte}(X^{mK})$.

We fix a point $x_0 \in S^2 \setminus \operatorname{post} f$. Note that $\deg_{f^n}(y) = 1$ for each $n \in \mathbb{N}$ and each $y \in f^{-n}(x_0)$.

For each $m \in \mathbb{N}$, each $k \in \{0, \dots, K-1\}$, and each $\widehat{y} \in \widehat{f}^{-m}(x_0)$, let $X^{mK}(\widehat{y}) \in \mathbf{X}^{mK}(f, \mathcal{C})$ be an mK-tile that contains $\widehat{y}$ and let $y(k, \widehat{y})$ be a preimage of x_{mK+k+N} under f^{mK+k+N} in $\operatorname{inte}(X^{mK}(\widehat{y}))$. By Proposition 3.2 (i), for each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f, \mathcal{C})$, $f^n|_{X^n}$ is a homeomorphism of X^n onto $f^n(X^n)$. This implies that for each $m \in \mathbb{N}$ and each $k \in \{0, \dots, K-1\}$, the map $\widehat{y} \mapsto y(k, \widehat{y})$ from $\widehat{f}^{-m}(x_0)$ to $f^{-mK-k-N}(x_{mK+k+N})$ is injective.

By the same reasoning as in Case 2, there exists $n_0 \in \mathbb{N}$ such that for each integer $m \geqslant n_0$, each $k \in \{0, \dots, K-1\}$, and each $\widehat{y} \in \widehat{f}^{-m}(x_0)$ with $\widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{G}}_\varepsilon$, it follows that $V_{mK+k+N}(y(k, \widehat{y})) \in \mathcal{G}$.

We now prove the lower bound. The proof is essentially the same as in Case 2, and we retain this proof for the convenience of the reader.

For each integer $n \geqslant N$, we set $m := \lfloor (n-N)/K \rfloor$ and write $n = mK + k + N$ for some integer $k \in \{0, \dots, K-1\}$. By the arguments above, for each sufficiently large $n \in \mathbb{N}$ that satisfies $m \geqslant n_0$, we have

$$\begin{aligned}
\sum_{\substack{y \in f^{-n}(x_n) \\ V_n(y) \in \mathcal{G}}} w_n(y) \exp\bigl(S_n^f \phi(y)\bigr) &\geqslant \sum_{\substack{y \in f^{-mK-k-N}(x_{mK+k+N}) \\ V_{mK+k+N}(y) \in \mathcal{G}}} \exp\bigl(S_{mK+k+N}^f \phi(y)\bigr) \\
&\geqslant \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S_{mK+k+N}^f \phi(y(k, \widehat{y}))\bigr),
\end{aligned}$$

where $\{w_j\}_{j\in\mathbb{N}}$ is an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \bigl[1, \deg_{f^j}(x)\bigr]$ for each $j \in \mathbb{N}$ and each $x \in S^2$. Then by Lemma 3.9,

$$\begin{aligned}
\sum_{\substack{y \in f^{-n}(x_n) \\ V_n(y) \in \mathcal{G}}} w_n(y) \exp\bigl(S_n^f \phi(y)\bigr) &\geqslant e^{-(K+N)\|\phi\|_\infty} \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S_{mK}^f \phi(y(k, \widehat{y}))\bigr) \\
&\geqslant e^{-C} \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S_{mK}^f \phi(\widehat{y})\bigr) = e^{-C} \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y})\bigr),
\end{aligned}$$

where the constant C is the same as in Case 2. Thus by (1.3), we have

$$\begin{aligned}
&\log \Omega_n(x_n)(\mathcal{G}) \\
&\quad = \log \sum_{\substack{y \in f^{-n}(x_n) \\ V_n(y) \in \mathcal{G}}} w_n(y) \exp\bigl(S_n^f \phi(y)\bigr) - \log \sum_{y' \in f^{-n}(x_n)} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) \\
&\quad \geqslant \log \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{G}}_\varepsilon}} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y})\bigr) - \log \sum_{y' \in f^{-n}(x_n)} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) - C \\
&\quad = \log \widehat{\Omega}_m(x_0)(\widehat{\mathcal{G}}_\varepsilon) + \log \sum_{\widehat{y}' \in \widehat{f}^{-m}(x_0)} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y}')\bigr) - \log \sum_{y' \in f^{-n}(x_n)} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) - C,
\end{aligned}$$

where $\bigl\{\widehat{\Omega}_j(x_0)\bigr\}_{j\in\mathbb{N}}$ is defined by setting $w_j = \mathbb{1}_{S^2}$ and $x_j = x_0$ for each $j \in \mathbb{N}$ and replacing f with $\widehat{f}$ and ϕ with $\widehat{\phi}$ in the definition of $\{\Omega_j(x_j)\}_{j\in\mathbb{N}}$ (recall (1.3)). Since $\widehat{f}$ has an $\widehat{f}$-invariant Jordan curve

$\mathcal{C} \subseteq S^2$ with post $\widehat{f} \subseteq \mathcal{C}$, Proposition 7.15 holds for $\widehat{f}$. Therefore, by Proposition 7.9, we get

$$\begin{aligned}\liminf_{n\to+\infty} \frac{1}{n} \log \Omega_n(x_n)(\mathcal{G}) &\geqslant \frac{1}{K} \liminf_{m\to+\infty} \frac{1}{m} \log \widehat{\Omega}_m(x_0)(\widehat{\mathcal{G}}_\varepsilon) + \frac{1}{K} P(\widehat{f}, \widehat{\phi}) - P(f,\phi) \\ &= \frac{1}{K} \liminf_{m\to+\infty} \frac{1}{m} \log \widehat{\Omega}_m(x_0)(\widehat{\mathcal{G}}_\varepsilon) \geqslant \frac{1}{K} \widehat{F}_{\widehat{\phi}}(\mu) = F_\phi(\mu).\end{aligned}$$

The proof is complete. □

7.4.3. *End of proof of the lower bound.*

Proof of Proposition 7.10. Let $\mathcal{G}$ be a non-empty open subset of $\mathcal{P}(S^2)$. Since subsets of $\mathcal{P}(S^2)$ of the form $\big\{\mu \in \mathcal{P}(S^2) : \int \vec{\varphi}\, \mathrm{d}\mu > \vec{\alpha}\big\}$ with $\ell \in \mathbb{N}$, $\vec{\varphi} \in C(S^2)^\ell$, $\vec{\alpha} \in \mathbb{R}^\ell$ constitute a base of the weak*-topology of $\mathcal{P}(S^2)$, we can write $\mathcal{G}$ as a union $\mathcal{G} = \bigcup_\lambda \mathcal{G}_\lambda$ of sets of this form. For each $\mathcal{G}_\lambda$, it follows from Proposition 7.15 that

$$\liminf_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{G}_\lambda) \geqslant \sup_{\mathcal{G}_\lambda} F_\phi,$$

for each sequence $\{\xi_n\}_{n\in\mathbb{N}} \in \big\{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\big\}$. Then by Remark 1.4, we get

$$\liminf_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{G}) \geqslant \sup_\lambda \sup_{\mathcal{G}_\lambda} F_\phi = \sup_{\mathcal{G}} F_\phi = -\inf_{\mathcal{G}} I_\phi$$

and complete the proof. □

7.5. **Large deviation upper bound.** In this subsection, we prove the upper bound (7.2) for all closed sets, with the main result being Proposition 7.16. Based on a preliminary result in Section 7.5.1, we prove upper bounds for certain fundamental closed subsets of $\mathcal{P}(S^2)$ in Section 7.5.2. Finally, in Section 7.5.3 we establish Proposition 7.16.

Proposition 7.16. *Let f, d, ϕ satisfy the Assumptions in Section 4. Then for each sequence $\{\xi_n\}_{n\in\mathbb{N}} \in \big\{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\big\}$ (as defined in Theorem 1.3), we have*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{K}) \leqslant -\inf_{\mathcal{K}} I_\phi \qquad \text{for all closed } \mathcal{K} \subseteq \mathcal{P}(S^2),$$

where $I_\phi \colon \mathcal{P}(S^2) \to [0,+\infty]$ is defined in (1.4).

7.5.1. *Construction of suitable invariant measures.* We use the notations as introduced in the beginning of Section 6.

Definition 7.17. Let f, $\mathcal{C}$, e^0 satisfy the Assumptions in Section 4. Consider $\ell \in \mathbb{N}$ and $\vec{\varphi} \in C(S^2)^\ell$. For each integer $n \in \mathbb{N}$ and each $\vec{\alpha} \in \mathbb{R}^\ell$ we define

$$\mathbf{P}^n(\vec{\alpha}) \coloneqq \big\{P^n \in \mathbf{P}^n(f,\mathcal{C},e^0) : \text{there exists } x \in P^n \text{ such that } n^{-1} S_n\vec{\varphi}(x) \geqslant \vec{\alpha}\big\}.$$

Here $\mathbf{P}^n(f,\mathcal{C},e^0)$ is the set of n-pairs (recall Definition 3.12).

Lemma 7.18. *Let f, $\mathcal{C}$, e^0 satisfy the Assumptions in Section 4. Consider $\ell \in \mathbb{N}$, $\vec{\varphi} \in C(S^2)^\ell$, and $\vec{\alpha} \in \mathbb{R}^\ell$. Then for each $n \in \mathbb{N}$ and each $x \in \bigcup \mathbf{P}^n(\vec{\alpha})$ we have $S_n\vec{\varphi}(x) \geqslant n\vec{\alpha} - 2D_n(\vec{\varphi})$. Here $D_n(\vec{\varphi})$ is defined in* (6.1).

Proof. Let $n \in \mathbb{N}$ and $P^n = X^n_{\mathfrak{b}} \cup X^n_{\mathfrak{w}} \in \mathbf{P}^n(\vec{\alpha})$ be arbitrary. It suffices to show that $S_n\vec{\varphi}(x) \geqslant n\vec{\alpha} - 2D_n(\vec{\varphi})$ for each $x \in P^n$. By Definition 3.12, there exists $e^n \in \mathbf{E}^n(f,\mathcal{C})$ with $f^n(e^n) = e^0$ such that $e^n \subseteq X^n_{\mathfrak{b}} \cap X^n_{\mathfrak{w}}$. We fix an arbitrary point $x_e \in e^n$. By the definition of $\mathbf{P}^n(\vec{\alpha})$, there exists $x_0 \in P^n = X^n_{\mathfrak{b}} \cup X^n_{\mathfrak{w}}$ such that $\frac{1}{n} S_n\vec{\varphi}(x_0) \geqslant \vec{\alpha}$. Since $x_e \in X^n_{\mathfrak{b}} \cap X^n_{\mathfrak{w}}$, we have $S_n\vec{\varphi}(x_e) \geqslant S_n\vec{\varphi}(x_0) - D_n(\vec{\varphi})$. Then for each $x \in P^n = X^n_{\mathfrak{b}} \cup X^n_{\mathfrak{w}}$,

$$S_n\vec{\varphi}(x) \geqslant S_n\vec{\varphi}(x_e) - D_n(\vec{\varphi}) \geqslant S_n\vec{\varphi}(x_0) - 2D_n(\vec{\varphi}) \geqslant n\vec{\alpha} - 2D_n(\vec{\varphi}).$$

The proof is complete. □

The following lemma is analog to [LSZ25, Proposition 7.15]. The proof is essentially the same, and we retain this proof for the convenience of the reader.

Proposition 7.19. *Let f, $\mathcal{C}$, d, ϕ, β, μ_ϕ, e^0 satisfy the Assumptions in Section 4. We assume in addition that $f(\mathcal{C}) \subseteq \mathcal{C}$. Consider $n \in \mathbb{N}$, $\ell \in \mathbb{N}$, $\vec{\varphi} \in C(S^2)^\ell$, and $\vec{\alpha} \in \mathbb{R}^\ell$. Suppose that for each $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$ there exists $P^n_{\mathfrak{c}} \in \mathbf{P}^n(\vec{\alpha})$ such that $P^n_{\mathfrak{c}} \subseteq \operatorname{inte}(X^0_{\mathfrak{c}})$. Then there exists a measure $\mu \in \mathcal{M}(S^2, f)$ such that*

$$\int \vec{\varphi} \,\mathrm{d}\mu \geqslant \vec{\alpha} - \frac{2D_n(\vec{\varphi})}{n} \qquad \text{and} \qquad \mu_\phi\Bigl(\bigcup \mathbf{P}^n(\vec{\alpha})\Bigr) \leqslant C \exp((P_\mu(f,\phi) - P(f,\phi))n),$$

where $D_n(\vec{\varphi})$ is defined in (6.1) *and $C = 2C_{\mu_\phi} \exp\bigl(C_1(\operatorname{diam}_d(S^2))^\beta\bigr)$. Here C_{μ_ϕ} is the constant from Proposition 3.11 and $C_1 \geqslant 0$ is the constant defined in* (3.10) *in Lemma 3.9.*

Proof. Denote $P^n(\vec{\alpha}) := \bigcup \mathbf{P}^n(\vec{\alpha})$. Note that the subsystem $F := f^n|_{P^n(\vec{\alpha})} \in \operatorname{Sub}(f^n, \mathcal{C})$ is strongly primitive (recall Definition 3.20). We set $F_\Omega := F|_\Omega$, where $\Omega := \Omega(F, \mathcal{C})$ is the tile maximal invariant set associated with F with respect to $\mathcal{C}$. Then it follows from Propositions 3.17 (i) and [LSZ25, Proposition 5.20 (ii)] that $F(\Omega) \subseteq \Omega$ and $\Omega \setminus \mathcal{C} \neq \emptyset$.

Let $y_0 \in \Omega \setminus \mathcal{C}$ be arbitrary. By [LSZ25, Theorem 1.1 and Proposition 6.20], we have

$$\sup_{\nu \in \mathcal{M}(\Omega, F_\Omega)} \biggl\{ h_\nu(F_\Omega) + \int S_n\phi \,\mathrm{d}\nu \biggr\} = \lim_{m \to +\infty} \frac{1}{m} \log \sum_{x \in F_\Omega^{-m}(y_0)} \exp\biggl(\sum_{k=0}^{m-1} S_n\phi\bigl(f^{nk}(x)\bigr) \biggr). \tag{7.12}$$

For the summand inside the logarithm in (7.12), we have

$$\sum_{x \in F_\Omega^{-m}(y_0)} \exp\biggl(\sum_{k=0}^{m-1} S_n\phi\bigl(f^{nk}(x)\bigr) \biggr) = \prod_{i=0}^{m-1} \sum_{y_{i+1} \in F_\Omega^{-1}(y_i)} \exp(S_n\phi(y_{i+1})). \tag{7.13}$$

Claim. For each point $y \in \Omega \setminus \mathcal{C}$, we have $\operatorname{card}\bigl(F_\Omega^{-1}(y)\bigr) = \operatorname{card}(\mathbf{P}^n(\vec{\alpha}))$, and each n-pair $P^n \in \mathbf{P}^n(\vec{\alpha})$ contains exactly one preimage $x \in F_\Omega^{-1}(y)$, which satisfies $x \in \Omega \setminus \mathcal{C}$.

To establish this Claim, we consider an arbitrary point $y \in \Omega \setminus \mathcal{C}$. Without loss of generality we may assume that $y \in \operatorname{inte}\bigl(X^0_{\mathfrak{b}}\bigr)$. Then by Proposition 3.2, we have $\operatorname{card}(f^{-n}(y)) = (\deg f)^n = \operatorname{card}(\mathbf{X}^n_{\mathfrak{b}})$, and each black n-tile $X^n_{\mathfrak{b}} \in \mathbf{X}^n_{\mathfrak{b}}$ contains exactly one preimage $x \in f^{-n}(y)$, which satisfies $x \in \operatorname{inte}(X^n_{\mathfrak{b}})$. Thus each n-pair $P^n \in \mathbf{P}^n(\vec{\alpha})$ contains exactly one preimage $x \in f^{-n}(y) \cap P^n(\vec{\alpha})$, which satisfies $x \in \operatorname{inte}(P^n)$, and we have

$$\operatorname{card}(f^{-n}(y) \cap P^n(\vec{\alpha})) = \operatorname{card}(\mathbf{P}^n(\vec{\alpha})).$$

Let preimage $x \in f^{-n}(y) \cap P^n(\vec{\alpha})$ be arbitrary. Noting that $f^{-n}(y) \cap P^n(\vec{\alpha}) = \bigl(f^n|_{P^n(\vec{\alpha})}\bigr)^{-1}(y) = F^{-1}(y)$ and $y \in \Omega \setminus \mathcal{C}$, by Proposition 3.17 (iii), we have $x \in \Omega \setminus \mathcal{C}$. Since $F_\Omega^{-1}(y) = f^{-n}(y) \cap \Omega = f^{-n}(y) \cap P^n(\vec{\alpha})$, the claim follows.

By the claim, we know that all the preimages y_i in the summation in (7.13) belong to $\Omega \setminus \mathcal{C}$. Moreover, for each point $y \in \Omega \setminus \mathcal{C}$, every n-pair $P^n \in \mathbf{P}^n(\vec{\alpha})$ contains exactly one preimage $x \in F_\Omega^{-1}(y)$, and every preimage $x \in F_\Omega^{-1}(y)$ is contained in a unique n-pair $P^n \in \mathbf{P}^n(\vec{\alpha})$. Thus by (7.13) we get the first two inequalities of the following:

$$\begin{aligned}\sum_{x\in F_\Omega^{-m}(y_0)} \exp\biggl(\sum_{k=0}^{m-1} S_n\phi\bigl(f^{nk}(x)\bigr)\biggr) &\geqslant \biggl(\inf_{y\in\Omega\setminus\mathcal{C}} \sum_{x\in F_\Omega^{-1}(y)} e^{S_n\phi(x)}\biggr)^m \\ &\geqslant \biggl(\sum_{P^n\in\mathbf{P}^n(\vec{\alpha})} \inf_{x\in P^n} e^{S_n\phi(x)}\biggr)^m \\ &\geqslant \biggl(\frac{e^{P(f,\phi)n}}{2C_{\mu_\phi} e^{C_1(\operatorname{diam}_d(S^2))^\beta}} \sum_{P^n\in\mathbf{P}^n(\vec{\alpha})} \mu_\phi(P^n)\biggr)^m \\ &= \biggl(\frac{e^{P(f,\phi)n}}{2C_{\mu_\phi} e^{C_1(\operatorname{diam}_d(S^2))^\beta}} \mu_\phi(P^n(\vec{\alpha}))\biggr)^m.\end{aligned}$$

The last inequality follows from [LSZ25, Lemmma 7.14 (i)] and the last equality follows from Lemma 3.13 and Theorem 3.10 (ii), where $C_1 \geqslant 0$ is the constant defined in (3.10) in Lemma 3.9. Taking logarithms of both sides, dividing by m, and plugging the result into the previous inequality, we get

$$\begin{aligned}&\lim_{m\to+\infty} \frac{1}{m} \log\biggl(\sum_{x\in F_\Omega^{-m}(y_0)} \exp\biggl(\sum_{k=0}^{m-1} S_n\phi\bigl(f^{nk}(x)\bigr)\biggr)\biggr) \\ &\qquad\geqslant \log\bigl(\mu_\phi(P^n(\vec{\alpha}))\bigr) + nP(f,\phi) - \bigl(C_1(\operatorname{diam}_d(S^2))^\beta + \log(2C_{\mu_\phi})\bigr).\end{aligned}$$

Plugging this inequality into (7.12) yields
(7.14)

$$\sup_{\nu\in\mathcal{M}(\Omega,F_\Omega)} \biggl\{h_\nu(F_\Omega) + \int S_n\phi\,\mathrm{d}\nu\biggr\} \geqslant \log\bigl(\mu_\phi(P^n(\vec{\alpha}))\bigr) + P(f,\phi)n - \bigl(C_1(\operatorname{diam}_d(S^2))^\beta + \log(2C_{\mu_\phi})\bigr).$$

By Theorem 3.25, there exists an equilibrium state $\widehat{\mu} \in \mathcal{M}(\Omega, F_\Omega) \subseteq \mathcal{M}(S^2, f^n)$ that attains the supremum in (7.14). Denote $\mu \coloneqq \frac{1}{n}\sum_{i=0}^{n-1} f_*^i\widehat{\mu}$. Then $\mu \in \mathcal{M}(S^2, f)$ and we have

$$\int \phi\,\mathrm{d}\mu = \frac{1}{n}\int \sum_{i=0}^{n-1} \phi\,\mathrm{d}f_*^i\widehat{\mu} = \frac{1}{n}\int \sum_{i=0}^{n-1} \phi\circ f^i\,\mathrm{d}\widehat{\mu} = \frac{1}{n}\int S_n\phi\,\mathrm{d}\widehat{\mu}.$$

By Lemma 7.18, we have $S_n\vec{\varphi}(x) \geqslant n\vec{\alpha} - 2D_n(\vec{\varphi})$ for each $x \in P^n(\vec{\alpha})$. Noting that $\operatorname{supp}\widehat{\mu} \subseteq \Omega \subseteq P^n(\vec{\alpha})$, we get

$$\int \vec{\varphi}\,\mathrm{d}\mu = \frac{1}{n}\int S_n\vec{\varphi}\,\mathrm{d}\widehat{\mu} \geqslant \vec{\alpha} - \frac{2D_n(\vec{\varphi})}{n}.$$

By Lemma 5.2, we have $nh_\mu(f) = h_{\widehat{\mu}}(f^n) = h_{\widehat{\mu}}(F_\Omega)$. Then

$$\begin{aligned}n\biggl(h_\mu(f) + \int\phi\,\mathrm{d}\mu\biggr) &= h_{\widehat{\mu}}(F_\Omega) + \int S_n\phi\,\mathrm{d}\widehat{\mu} \\ &\geqslant \log\bigl(\mu_\phi(P^n(\vec{\alpha}))\bigr) + nP(f,\phi) - \bigl(C_1(\operatorname{diam}_d(S^2))^\beta + \log(2C_{\mu_\phi})\bigr),\end{aligned}$$

i.e., $\log(\mu_\phi(P^n(\vec{\alpha}))) \leqslant n(P_\mu(f,\phi) - P(f,\phi)) + C_1(\operatorname{diam}_d(S^2))^\beta + \log(2C_{\mu_\phi})$. The proof is complete. □

7.5.2. *Upper bound for fundamental closed sets.* We first prove the following result under the additional assumption that there exists an f-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} f \subseteq \mathcal{C}$ and then for the general case.

Proposition 7.20. *Let f, $\mathcal{C}$, d, ϕ, β, μ_ϕ satisfy the Assumptions in Section 4. Consider $\ell \in \mathbb{N}$, $\vec{\varphi} \in C(S^2)^\ell$, and $\vec{\alpha} \in \mathbb{R}^\ell$. Let $\mathcal{K} \subseteq \mathcal{P}(S^2)$ be a non-empty closed set of the form*

$$\mathcal{K} \coloneqq \biggl\{\mu \in \mathcal{P}(S^2) : \int \vec{\varphi}\,\mathrm{d}\mu \geqslant \vec{\alpha}\biggr\}.$$

Then for each $\varepsilon > 0$ *and each sequence* $\{\xi_n\}_{n\in\mathbb{N}} \in \bigl\{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\bigr\}$ *(as defined in Theorem 1.3), we have*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \xi_n(\mathcal{K}) \leqslant \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha} - \varepsilon\biggr\}, \tag{7.15}$$

where $F_\phi\colon \mathcal{P}(S^2) \to [-\infty, 0]$ *is defined in* (1.5).

Proof of Proposition 7.20 under an additional assumption. We assume in addition that there exists an f-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post} f \subseteq \mathcal{C}$.

Let $\varepsilon > 0$ and $\{\xi_n\}_{n\in\mathbb{N}} \in \bigl\{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\bigr\}$ be arbitrary. We may assume without loss of generality that the set $\{n \in \mathbb{N} : \xi_n(\mathcal{K}) > 0\}$ is unbounded, because otherwise (7.15) holds trivially. Then it follows from the definition of $\{\Sigma_n\}_{n\in\mathbb{N}}$, $\{\Omega_n\}_{n\in\mathbb{N}}$, and $\{\Omega_n(x_n)\}_{n\in\mathbb{N}}$ that for each $n_0 \in \mathbb{N}$ there exists an integer $n \geqslant n_0$ and a point $x \in S^2$ such that $S_n\vec{\varphi}(x) \geqslant n\vec{\alpha}$.

We first show that there exists an ergodic measure $\mu_0 \in \mathcal{M}(S^2, f)$ such that $\int \vec{\varphi}\,\mathrm{d}\mu_0 \geqslant \vec{\alpha} - \varepsilon/4$. Let $n_f \in \mathbb{N}$ be the constant from Definition 3.20, which depends only on f and $\mathcal{C}$. Then by Proposition 3.8 and Lemma 7.6 (i), for each $n \in \mathbb{N}_0$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, there exists a fixed point of f^{n+n_f} in $\operatorname{inte}(X^n)$. By Lemma 6.1, there exists a sufficiently large $n_0 \in \mathbb{N}$ such that for each integer $n \geqslant n_0$,

$$\frac{n_f(\|\vec{\alpha}\| + \|\vec{\varphi}\|) + D_n(\vec{\varphi})}{n + n_f} \leqslant \frac{\varepsilon}{4}.$$

Then by the argument in the beginning of the proof, there exists an integer $n \geqslant n_0$ and a point $x \in S^2$ such that $S_n\vec{\varphi}(x) \geqslant n\vec{\alpha}$. We pick an n-tile $X^n \in \mathbf{X}^n(f,\mathcal{C})$ such that $x \in X^n$. Thus there exists a fixed point $p \in X^n$ of f^{n+n_f}. Noting that $S_n\vec{\varphi}(p) \geqslant S_n\vec{\varphi}(x) - D_n(\vec{\varphi})$, we have

$$\begin{aligned} S_{n+n_f}\vec{\varphi}(p) &\geqslant S_n\vec{\varphi}(p) - n_f\|\vec{\varphi}\| \geqslant S_n\vec{\varphi}(x) - n_f\|\vec{\varphi}\| - D_n(\vec{\varphi}) \\ &\geqslant n\vec{\alpha} - n_f\|\vec{\varphi}\| - D_n(\vec{\varphi}) \geqslant (n+n_f)\vec{\alpha} - n_f\|\vec{\alpha}\| - n_f\|\vec{\varphi}\| - D_n(\vec{\varphi}). \end{aligned}$$

This implies

$$\int \vec{\varphi}\,\mathrm{d}V_{n+n_f}(p) = \frac{1}{n+n_f} S_{n+n_f}\vec{\varphi}(p) \geqslant \vec{\alpha} - \frac{n_f(\|\vec{\alpha}\| + \|\vec{\varphi}\|) + D_n(\vec{\varphi})}{n+n_f} \geqslant \vec{\alpha} - \frac{\varepsilon}{4}.$$

Set $\mu_0 \coloneqq V_{n+n_f}(p)$. Then $\mu_0 \in \mathcal{M}(S^2, f)$ is an ergodic measure with $\int \vec{\varphi}\,\mathrm{d}\mu_0 \geqslant \vec{\alpha} - \varepsilon/4$.

Fix an arbitrary 0-edge $e^0 \in \mathbf{E}^0(f,\mathcal{C})$. By Lemma 6.4, there exists $N \in \mathbb{N}$ such that for each integer $n \geqslant N$ and each $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, there exists $P^n_{\mathfrak{c}} \in \mathbf{P}^n(f, \mathcal{C}, e^0)$ such that $P^n_{\mathfrak{c}} \subseteq \operatorname{inte}\bigl(X^0_{\mathfrak{c}}\bigr)$ and

$$\sup_{x\in P^n_{\mathfrak{c}}} \biggl\| \frac{1}{n} S_n\vec{\varphi}(x) - \int \vec{\varphi}\,\mathrm{d}\mu_0 \biggr\| \leqslant \frac{\varepsilon}{4}.$$

Then for each $x \in P^n_{\mathfrak{c}}$,

$$\frac{1}{n} S_n\vec{\varphi}(x) \geqslant \int \vec{\varphi}\,\mathrm{d}\mu_0 - \frac{\varepsilon}{4} \geqslant \vec{\alpha} - \frac{\varepsilon}{2}.$$

This implies that for each integer $n \geqslant N$ and each $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$, there exists $P^n_{\mathfrak{c}} \in \mathbf{P}^n(\vec{\alpha} - \varepsilon/2)$ such that $P^n_{\mathfrak{c}} \subseteq \operatorname{inte}\bigl(X^0_{\mathfrak{c}}\bigr)$. Therefore, it follows from Proposition 7.19 that for each integer $n \geqslant N$, there exists a measure $\mu_n \in \mathcal{M}(S^2, f)$ such that

$$\int \vec{\varphi}\,\mathrm{d}\mu_n \geqslant \vec{\alpha} - \frac{\varepsilon}{2} - \frac{2D_n(\vec{\varphi})}{n} \qquad \text{and} \qquad \mu_\phi(P^n(\vec{\alpha} - \varepsilon/2)) \leqslant C \exp(F_\phi(\mu_n) n), \tag{7.16}$$

where $C = 2C_{\mu_\phi} \exp\bigl(C_1(\operatorname{diam}_d(S^2))^\beta\bigr)$. Here C_{μ_ϕ} is the constant from Proposition 3.11 and $C_1 \geqslant 0$ is the constant defined in (3.10) in Lemma 3.9 that depends only on f, $\mathcal{C}$, d, ϕ, and β.

We split the rest of the proof into three cases according to the type of the sequence $\{\xi_n\}_{n\in\mathbb{N}}$.

Case 1 (Birkhoff averages): $\xi_n = \Sigma_n = (V_n)_*(\mu_\phi)$ for each $n \in \mathbb{N}$ (recall (1.1)).

For each integer $n \geqslant N$, since $\{x \in S^2 : V_n(x) \in \mathcal{K}\} \subseteq P^n(\vec{\alpha} - \varepsilon/2)$, we have

$$\log \Sigma_n(\mathcal{K}) = \log \mu_\phi(\{x \in S^2 : V_n(x) \in \mathcal{K}\}) \leqslant \log \mu_\phi(P^n(\vec{\alpha} - 2^{-1}\varepsilon)).$$

Thus by (7.16) and Lemma 6.1, for each sufficiently large integer $n \geqslant N$ that satisfies $2D_n(\vec{\varphi})/n < \varepsilon/2$, we have

$$\frac{1}{n}\log \Sigma_n(\mathcal{K}) \leqslant F_\phi(\mu_n) + \frac{\log C}{n} \leqslant \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha} - \varepsilon\biggr\} + \frac{\log C}{n}.$$

Then by letting $n \to +\infty$, the desired inequality follows.

Case 2 (Periodic points): $\xi_n = \Omega_n$ for each $n \in \mathbb{N}$ (recall (1.2)).

For each $n \in \mathbb{N}$ and each $p \in \mathrm{Per}_n(f)$, let $X^n(p) \in \mathbf{X}^n(f, \mathcal{C})$ be an n-tile that contains p. In particular, if $p \in \mathrm{Per}_n(f)$ satisfies $V_n(p) \in \mathcal{K}$, then $X^n(p) \subseteq P^n(\vec{\alpha}) \subseteq P^n(\vec{\alpha} - \varepsilon/2)$. By [Li15, Lemma 6.3], there exists $N_0 \in \mathbb{N}$ such that for each integer $n \geqslant N_0$ and each n-tile $X^n \in \mathbf{X}^n(f, \mathcal{C})$, the number of fixed points of f^n contained in X^n is at most 1. This implies that for each integer $n \geqslant N_0$, the map $p \mapsto X^n(p)$ from $\mathrm{Per}_n(f)$ to $\mathbf{X}^n(f, \mathcal{C})$ is injective.

Let $n \in \mathbb{N}$ be arbitrary. Consider $p' \in \mathrm{Per}_n(f) \cap \mathbf{V}^n$, where $\mathbf{V}^n = \mathbf{V}^n(f, \mathcal{C})$ is the set of n-vertices. We set $\mathbf{X}^n(f, \mathcal{C}, p') \coloneqq \{X \in \mathbf{X}^n(f, \mathcal{C}) : p' \in X\}$. By Remark 3.4, we have $\overline{W}^n(p') = \bigcup \mathbf{X}^n(f, \mathcal{C}, p')$ and $\mathrm{card}(\mathbf{X}^n(f, \mathcal{C}, p')) = 2\deg_{f^n}(p')$, where $W^n(p')$ is defined in (3.6) and $\overline{W}^n(p')$ is the closure of $W^n(p')$. In particular, if $p' \in \mathrm{Per}_n(f)$ satisfies $V_n(p) \in \mathcal{K}$, then $\overline{W}^n(p') = \bigcup \mathbf{X}^n(f, \mathcal{C}, p') \subseteq P^n(\vec{\alpha}) \subseteq P^n(\vec{\alpha} - \varepsilon/2)$. Moreover, if $n \geqslant N_0$, then $X^n(p) \notin \mathbf{X}^n(f, \mathcal{C}, p')$ for every $p \in \mathrm{Per}_n(f) \setminus \mathbf{V}^n(f, \mathcal{C})$.

We are now ready to establish the desired upper bound. Let $\{w_j\}_{j\in\mathbb{N}}$ be an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \bigl[1, \deg_{f^j}(x)\bigr]$ for each $j \in \mathbb{N}$ and each $x \in S^2$. For each integer $n \geqslant N_0$, we have

$$\begin{aligned}
\sum_{\substack{p \in \mathrm{Per}_n(f) \\ V_n(p) \in \mathcal{K}}} w_n(p) e^{S_n\phi(p)} &\leqslant \sum_{\substack{p' \in \mathrm{Per}_n(f) \cap \mathbf{V}^n \\ V_n(p') \in \mathcal{K}}} \deg_{f^n}(p') e^{S_n\phi(p')} + \sum_{\substack{p \in \mathrm{Per}_n(f) \setminus \mathbf{V}^n \\ V_n(p) \in \mathcal{K}}} e^{S_n\phi(p)} \\
&\leqslant \sum_{\substack{p' \in \mathrm{Per}_n(f) \cap \mathbf{V}^n \\ V_n(p') \in \mathcal{K}}} \sum_{X^n \in \mathbf{X}^n(f,\mathcal{C},p')} e^{S_n\phi(X^n)} + \sum_{\substack{p \in \mathrm{Per}_n(f) \setminus \mathbf{V}^n \\ V_n(p) \in \mathcal{K}}} e^{S_n\phi(X^n(p))} \\
&\leqslant \sum_{\substack{X^n \subseteq P^n(\vec{\alpha} - \varepsilon/2) \\ X^n \in \mathbf{X}^n(f,\mathcal{C})}} e^{S_n\phi(X^n)},
\end{aligned}$$

where we write $S_n\phi(X^n) \coloneqq \sup_{x \in X^n} S_n\phi(x)$ for each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f, \mathcal{C})$. Then it follows from Proposition 3.11 and Theorem 3.10 (ii) that for each integer $n \geqslant N_0$,

$$\sum_{\substack{p \in \mathrm{Per}_n(f) \\ V_n(p) \in \mathcal{K}}} w_n(p) e^{S_n\phi(p)} \leqslant C_{\mu_\phi} e^{nP(f,\phi)} \sum_{\substack{X^n \subseteq P^n(\vec{\alpha} - \varepsilon/2) \\ X^n \in \mathbf{X}^n(f,\mathcal{C})}} \mu_\phi(X^n) = C_{\mu_\phi} e^{nP(f,\phi)} \mu_\phi(P^n(\vec{\alpha} - \varepsilon/2)).$$

By (7.16) and Lemma 6.1, for each sufficiently large integer $n \geqslant \max\{N_0, N\}$ that satisfies $2D_n(\vec{\varphi})/n < \varepsilon/2$,

$$\frac{1}{n}\log \mu_\phi(P^n(\vec{\alpha} - \varepsilon/2)) \leqslant F_\phi(\mu_n) + \frac{\log C}{n} \leqslant \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha} - \varepsilon\biggr\} + \frac{\log C}{n},$$

and therefore

$$\begin{aligned}
\frac{1}{n}\log \Omega_n(\mathcal{K}) \leqslant{}& -\frac{1}{n}\log \sum_{p \in \mathrm{Per}_n(f)} w_n(p)\exp(S_n\phi(p)) \\
&+ \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha} - \varepsilon\biggr\} + P(f,\phi) + \frac{\log(CC_{\mu_\phi})}{n}.
\end{aligned}$$

Note that as $n \to +\infty$, the first term of the right hand side in the equation above converges to $-P(f, \phi)$ by Proposition 7.7. Therefore, by letting $n \to +\infty$, the desired inequality follows.

Case 3 (Iterated preimages): $\xi_n = \Omega_n(x_n)$ for each $n \in \mathbb{N}$ (recall (1.3)), where $\{x_n\}_{n\in\mathbb{N}}$ is an arbitrary sequence of points in S^2.

For each $n \in \mathbb{N}$ and each $y \in f^{-n}(x_n)$, let $X^n(y) \in \mathbf{X}^n(f,\mathcal{C})$ be an n-tile that contains y. In particular, if $y \in f^{-n}(x_n)$ satisfies $V_n(y) \in \mathcal{K}$, then $X^n(y) \subseteq P^n(\vec{\alpha}) \subseteq P^n(\vec{\alpha}-\varepsilon/2)$. By Proposition 3.2 (i), for each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f,\mathcal{C})$, $f^n|_{X^n}$ is a homeomorphism of X^n onto $f^n(X^n)$. This implies that for each integer $n \in \mathbb{N}$ the map $y \mapsto X^n(y)$ from $f^{-n}(x_n)$ to $\mathbf{X}^n(f,\mathcal{C})$ is injective.

Let $\{w_j\}_{j\in\mathbb{N}}$ be an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \bigl[1, \deg_{f^j}(x)\bigr]$ for each $j \in \mathbb{N}$ and each $x \in S^2$. By similar arguments as in Case 2, for each sufficiently large integer $n \geqslant N$ that satisfies $2D_n(\vec{\varphi})/n < \varepsilon/2$,

$$\frac{1}{n}\log \sum_{\substack{y\in f^{-n}(x_n)\\ V_n(y)\in\mathcal{K}}} w_n(y)e^{S_n\phi(y)} \leqslant \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha}-\varepsilon\biggr\} + P(f,\phi) + \frac{\log(CC_{\mu_\phi})}{n},$$

and therefore

$$\begin{aligned}\frac{1}{n}\log \Omega_n(x_n)(\mathcal{K}) \leqslant{}& -\frac{1}{n}\log \sum_{y\in f^{-n}(x_n)} w_n(y)\exp(S_n\phi(y)) \\ &+ \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha}-\varepsilon\biggr\} + P(f,\phi) + \frac{\log(CC_{\mu_\phi})}{n}.\end{aligned}$$

Note that as $n \to +\infty$, the first term of the right hand side in the equation above converges to $-P(f,\phi)$ by Proposition 7.9. Therefore, by letting $n \to +\infty$, the desired inequality follows.

The proof is complete. □

We now prove the general case.

Proof of Proposition 7.20. By Lemma 3.7, we can find a sufficiently high iterate $\widehat{f} := f^K$ of f that has an $\widehat{f}$-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post}\widehat{f} = \operatorname{post} f \subseteq \mathcal{C}$. Then $\widehat{f}$ is also an expanding Thurston map.

Denote $\vec{\Phi} := S_K^f\vec{\varphi}$ and $\widehat{\phi} := S_K^f\phi$. We define $\widehat{F}_{\widehat{\phi}} \colon \mathcal{P}(S^2) \to [-\infty, 0]$ by

$$\widehat{F}_{\widehat{\phi}}(\nu) := \begin{cases} h_\nu(\widehat{f}) + \int \widehat{\phi}\,\mathrm{d}\nu - P(\widehat{f},\widehat{\phi}) & \text{if } \nu \in \mathcal{M}(S^2,\widehat{f}); \\ -\infty & \text{if } \nu \in \mathcal{P}(S^2)\setminus\mathcal{M}(S^2,\widehat{f}). \end{cases}$$

Recall from Subsection 3.1 that $P(\widehat{f},\widehat{\phi}) = KP(f,\phi)$, $h_\mu(\widehat{f}) = Kh_\mu(f)$, and $\int\widehat{\phi}\,\mathrm{d}\mu = K\int\phi\,\mathrm{d}\mu$ for $\mu \in \mathcal{M}(S^2,f)$. Then for each $\mu \in \mathcal{M}(S^2,f)$, we have $\widehat{F}_{\widehat{\phi}}(\mu) = KF_\phi(\mu)$ since $\mathcal{M}(S^2,f) \subseteq \mathcal{M}(S^2,\widehat{f})$. Since $P_{\mu_\phi}(\widehat{f},\widehat{\phi}) = KP_{\mu_\phi}(f,\phi) = KP(f,\phi) = P(\widehat{f},\widehat{\phi})$, it follows from Theorem 3.10 (i) that μ_ϕ is the unique equilibrium state for $\widehat{f}$ and $\vec{\Phi}$.

For each $\delta > 0$, let $\widehat{\mathcal{K}}_\delta \subseteq \mathcal{P}(S^2)$ be the closed set defined by

$$\widehat{\mathcal{K}}_\delta := \biggl\{\mu \in \mathcal{P}(S^2) : \int \vec{\Phi}\,\mathrm{d}\mu \geqslant K\vec{\alpha} - K\delta\biggr\}.$$

Let $\varepsilon > 0$ be arbitrary. We claim that

$$\sup\biggl\{\widehat{F}_{\widehat{\phi}}(\widehat{\mu}) : \widehat{\mu} \in \mathcal{P}(S^2),\ \int\vec{\Phi}\,\mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon\biggr\} = K\sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int\vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha}-\varepsilon\biggr\}. \tag{7.17}$$

By the definitions of F_ϕ and $\widehat{F}_{\widehat{\phi}}$, it suffices to show that

$$\sup\biggl\{\widehat{F}_{\widehat{\phi}}(\widehat{\mu}) : \widehat{\mu} \in \mathcal{M}(S^2,\widehat{f}),\ \int\vec{\Phi}\,\mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon\biggr\} = K\sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{M}(S^2,f),\ \int\vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha}-\varepsilon\biggr\}.$$

To see this, we consider arbitrary $\widehat{\mu} \in \mathcal{M}(S^2,\widehat{f})$ satisfying $\int\vec{\Phi}\,\mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon$. Define $\mu := \frac{1}{K}\sum_{j=0}^{K-1} f_*^j\widehat{\mu}$. Then $\int\vec{\varphi}\,\mathrm{d}\mu = \frac{1}{K}\int\vec{\Phi}\,\mathrm{d}\widehat{\mu} > \vec{\alpha}-\varepsilon$ and it follows from Lemma 5.2 that $\mu \in \mathcal{M}(S^2,f)$

and $h_{\widehat{\mu}}(\widehat{f}) = K h_\mu(f)$. Thus we have $\widehat{F}_{\widehat{\phi}}(\widehat{\mu}) = K F_\phi(\mu)$ and

$$\sup\biggl\{\widehat{F}_{\widehat{\phi}}(\widehat{\mu}) : \widehat{\mu} \in \mathcal{M}(S^2, \widehat{f}),\ \int \vec{\Phi}\,\mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon\biggr\} \leqslant K \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{M}(S^2, f),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha} - \varepsilon\biggr\}.$$

The other direction follows immediately from the facts that $\mathcal{M}(S^2, f) \subseteq \mathcal{M}(S^2, \widehat{f})$ and $KF_\phi(\mu) = \widehat{F}_{\widehat{\phi}}(\mu)$ for each $\mu \in \mathcal{M}(S^2, f)$.

We split the proof into three cases according to the type of the sequence $\{\xi_n\}_{n\in\mathbb{N}}$.

Case 1 (Birkhoff averages): $\xi_n = \Sigma_n = (V_n)_*(\mu_\phi)$ for each $n \in \mathbb{N}$ (recall (1.1)).

For each integer $k \in \{0, \dots, K-1\}$ and each integer $m \in \mathbb{N}$ that satisfies $(\|\vec{\alpha}\| + \|\vec{\varphi}\|)/m \leqslant \varepsilon/2$, we have

$$\begin{aligned}\bigl\{x \in S^2 : S^f_{mK-k}\vec{\varphi}(x) \geqslant (mK-k)\vec{\alpha}\bigr\} &\subseteq \bigl\{x \in S^2 : S^f_{mK}\vec{\varphi}(x) \geqslant (mK-k)\vec{\alpha} - k\|\vec{\varphi}\|\bigr\}\\ &\subseteq \bigl\{x \in S^2 : S^f_{mK}\vec{\varphi}(x) \geqslant mK\vec{\alpha} - K(\|\vec{\alpha}\| + \|\vec{\varphi}\|)\bigr\}\\ &\subseteq \bigl\{x \in S^2 : m^{-1} S^{\widehat{f}}_m \vec{\Phi}(x) \geqslant K\vec{\alpha} - K\varepsilon/2\bigr\}.\end{aligned}$$

For each $n \in \mathbb{N}$, we can write $n = mK - k$ for some integer $k \in \{0, \dots, K-1\}$ and $m \in \mathbb{N}$. Then for each sufficiently large $n \in \mathbb{N}$, we have

$$\begin{aligned}\log \Sigma_n(\mathcal{K}) &= \log \mu_\phi\bigl(\bigl\{x \in S^2 : S^f_n\vec{\varphi}(x) \geqslant n\vec{\alpha}\bigr\}\bigr) \leqslant \log \widehat{\mu}_{\widehat{\phi}}\bigl(\bigl\{x \in S^2 : m^{-1} S^{\widehat{f}}_m \vec{\Phi}(x) \geqslant K\vec{\alpha} - K\varepsilon/2\bigr\}\bigr)\\ &= \log \widehat{\Sigma}_m(\widehat{\mathcal{K}}_{\varepsilon/2}) \leqslant \frac{n}{mK} \log \widehat{\Sigma}_m(\widehat{\mathcal{K}}_{\varepsilon/2}),\end{aligned}$$

where $m = \lceil n/K \rceil$ and $\bigl\{\widehat{\Sigma}_j\bigr\}_{j\in\mathbb{N}}$ is defined by replacing f with $\widehat{f}$ and ϕ with $\widehat{\phi}$ in the definition of $\{\Sigma_j\}_{j\in\mathbb{N}}$. Since $\widehat{f}$ has an $\widehat{f}$-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with $\operatorname{post}\widehat{f} \subseteq \mathcal{C}$, Proposition 7.20 holds for $\widehat{f}$. Therefore, by (7.17), we get

$$\begin{aligned}\limsup_{n\to+\infty} \frac{\log \Sigma_n(\mathcal{K})}{n} &\leqslant \frac{1}{K} \limsup_{m\to+\infty} \frac{\log \widehat{\Sigma}_m(\widehat{\mathcal{K}}_{\varepsilon/2})}{m} \leqslant \frac{1}{K} \sup\biggl\{\widehat{F}_{\widehat{\phi}}(\widehat{\mu}) : \widehat{\mu} \in \mathcal{P}(S^2),\ \int \vec{\Phi}\,\mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon\biggr\}\\ &= \sup\biggl\{F_\phi(\mu) : \mu \in \mathcal{P}(S^2),\ \int \vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha} - \varepsilon\biggr\}.\end{aligned}$$

Case 2 (Periodic points): $\xi_n = \Omega_n$ for each $n \in \mathbb{N}$ (recall (1.2)).

Let $N := n_f \in \mathbb{N}$ be the constant from Definition 3.20, which depends only on f and $\mathcal{C}$. For each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f, \mathcal{C})$, it follows from Lemma 3.22 that there exists $X^{\ell K} \in \mathbf{X}^{\ell K}(f, \mathcal{C})$ such that $X^{\ell K} \subseteq \operatorname{inte}(X^n)$, where $\ell := \lceil (n+N)/K \rceil$. Since $\widehat{f}(\mathcal{C}) \subseteq \mathcal{C}$, it follows from Propositions 3.2 (iv) and 3.8 that $X^{\ell K} \subseteq X^0_{\mathfrak{c}}$ for some $\mathfrak{c} \in \{\mathfrak{b}, \mathfrak{w}\}$. Define $m := \ell + \lceil N/K \rceil$. Then by Lemma 7.6 (i), there exists a fixed point of f^{mK} in $\operatorname{inte}(X^{\ell K}) \subseteq \operatorname{inte}(X^n)$.

For each $n \in \mathbb{N}$ and each $X^n \in \mathbf{X}^n(f, \mathcal{C})$, we fix a fixed point of f^{mK} in $\operatorname{inte}(X^n)$ and denote it by $\widehat{p}(X^n)$, where $m = \lceil (n+N)/K \rceil + \lceil N/K \rceil$. Then for each $n \in \mathbb{N}$, the map $X^n \mapsto \widehat{p}(X^n)$ from $\mathbf{X}^n(f, \mathcal{C})$ to $\operatorname{Per}_m(\widehat{f})$ is injective.

For each $n \in \mathbb{N}$ and each $p \in \operatorname{Per}_n(f)$, let $X^n(p) \in \mathbf{X}^n(f, \mathcal{C})$ be an n-tile that contains p. By [Li15, Lemma 6.3], there exists $N_0 \in \mathbb{N}$ such that for each integer $n \geqslant N_0$ and each n-tile $X^n \in \mathbf{X}^n(f, \mathcal{C})$, the number of fixed points of f^n contained in X^n is at most 1. This implies that for each integer $n \geqslant N_0$, the map $p \mapsto X^n(p)$ from $\operatorname{Per}_n(f)$ to $\mathbf{X}^n(f, \mathcal{C})$ is injective. Therefore, for each integer $n \geqslant N_0$, the map $p \to \widehat{p}(X^n(p))$ from $\operatorname{Per}_n(f)$ to $\operatorname{Per}_m(\widehat{f})$ is injective, where $m = \lceil (n+N)/K \rceil + \lceil N/K \rceil$ and $\widehat{p}(X^n(p)) \in \operatorname{inte}(X^n(p))$.

We claim that there exists $n_0 \in \mathbb{N}$ such that for each integer $n \geqslant n_0$, each $p \in \operatorname{Per}_n(f)$ with $V_n(p) \in \mathcal{K}$, and each $X^n \in \mathbf{X}^n(f, \mathcal{C})$ with $p \in X^n$, it follows that $\widehat{V}_m(\widehat{p}(X^n)) \in \widehat{\mathcal{K}}_{\varepsilon/2}$, where $m = \lceil (n+N)/K \rceil + \lceil N/K \rceil$ and $\widehat{V}_\ell(x) := \frac{1}{\ell} \sum_{i=0}^{\ell-1} \delta_{\widehat{f}^i(x)}$ for each $\ell \in \mathbb{N}$ and each $x \in S^2$. Indeed, by Lemma 6.1, there exists a sufficiently large $n_0 \in \mathbb{N}$ such that for each integer $n \geqslant n_0$,

$$D_n(\vec{\varphi}) + 2(K+N)(\|\vec{\alpha}\| + \|\vec{\varphi}\|) \leqslant n\varepsilon/2.$$

Since X^n contains p and $\widehat{p}(X^n)$, we have $S_n^f\vec{\varphi}(\widehat{p}(X^n)) \geqslant S_n^f\vec{\varphi}(p) - D_n(\vec{\varphi})$. Set $m := \lceil (n+N)/K \rceil + \lceil N/K \rceil$. Note that $0 \leqslant mK - n \leqslant 2(N+K)$ and $V_n(p) \in \mathcal{K}$ means that $S_n^f\vec{\varphi}(p) \geqslant n\vec{\alpha}$. Therefore,

$$
\begin{aligned}
S_m^{\widehat{f}}\vec{\Phi}(\widehat{p}(X^n)) = S_{mK}^f\vec{\varphi}(\widehat{p}(X^n)) &\geqslant S_n^f\vec{\varphi}(\widehat{p}(X^n)) - 2(K+N)\|\vec{\varphi}\| \\
&\geqslant S_n^f\vec{\varphi}(\widehat{p}) - D_n(\vec{\varphi}) - 2(K+N)\|\vec{\varphi}\| \\
&\geqslant n\vec{\alpha} - D_n(\vec{\varphi}) - 2(K+N)\|\vec{\varphi}\| \\
&\geqslant mK\vec{\alpha} - D_n(\vec{\varphi}) - 2(K+N)(\|\vec{\alpha}\| + \|\vec{\varphi}\|) \\
&\geqslant mK\vec{\alpha} - n\varepsilon/2 \\
&\geqslant mK\vec{\alpha} - mK\varepsilon/2.
\end{aligned}
$$

This implies $V_m(\widehat{p}(X^n)) \in \widehat{\mathcal{K}}_{\varepsilon/2}$.

Let $n \in \mathbb{N}$ be arbitrary. Consider $p' \in \operatorname{Per}_n(f) \cap \mathbf{V}^n$, where $\mathbf{V}^n = \mathbf{V}^n(f,\mathcal{C})$ is the set of n-vertices. We set $\mathbf{X}^n(f,\mathcal{C},p') := \{X \in \mathbf{X}^n(f,\mathcal{C}) : p' \in X\}$. By Remark 3.4, we have $\overline{W}^n(p') = \bigcup \mathbf{X}^n(f,\mathcal{C},p')$ and $\operatorname{card}(\mathbf{X}^n(f,\mathcal{C},p')) = 2\deg_{f^n}(p')$, where $W^n(p')$ is defined in (3.6) and $\overline{W}^n(p')$ is the closure of $W^n(p')$. Note that if $n \geqslant N_0$, then $X^n(p) \notin \mathbf{X}^n(f,\mathcal{C},p')$ for every $p \in \operatorname{Per}_n(f) \setminus \mathbf{V}^n(f,\mathcal{C})$.

We are now ready to establish the desired upper bound. Let $\{w_j\}_{j\in\mathbb{N}}$ be an arbitrary sequence of real-valued functions on S^2 with $w_j(x) \in \big[1, \deg_{f^j}(x)\big]$ for each $j \in \mathbb{N}$ and each $x \in S^2$.

For each integer $n \in \mathbb{N}$, we set $m := \lceil (n+N)/K \rceil + \lceil N/K \rceil$. Note that $0 \leqslant mK - n \leqslant 2(N+K)$. By the arguments above, for each integer $n \geqslant \max\{N_0, n_0\}$, we have

$$
\begin{aligned}
&\sum_{\substack{p\in\operatorname{Per}_n(f)\\ V_n(p)\in\mathcal{K}}} w_n(p)\exp\big(S_n^f\phi(p)\big) \\
&\quad\leqslant \sum_{\substack{p'\in\operatorname{Per}_n(f)\cap\mathbf{V}^n\\ V_n(p')\in\mathcal{K}}} \deg_{f^n}(p')\exp\big(S_n^f\phi(p')\big) + \sum_{\substack{p\in\operatorname{Per}_n(f)\setminus\mathbf{V}^n\\ V_n(p)\in\mathcal{K}}} \exp\big(S_n^f\phi(p)\big) \\
&\quad\leqslant \sum_{\substack{p'\in\operatorname{Per}_n(f)\cap\mathbf{V}^n\\ V_n(p')\in\mathcal{K}}} \sum_{X^n\in\mathbf{X}^n(f,\mathcal{C},p')} e^{D_n(\phi)+S_n^f\phi(\widehat{p}(X^n))} + \sum_{\substack{p\in\operatorname{Per}_n(f)\setminus\mathbf{V}^n\\ V_n(p)\in\mathcal{K}}} e^{D_n(\phi)+S_n^f\phi(\widehat{p}(X^n(p)))} \\
&\quad\leqslant e^{D_n(\phi)} \sum_{\substack{\widehat{p}\in\operatorname{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\big(S_n^f\phi(\widehat{p})\big) \\
&\quad\leqslant e^{D_n(\phi)} e^{2(N+K)\|\phi\|_\infty} \sum_{\substack{\widehat{p}\in\operatorname{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\big(S_m^{\widehat{f}}\widehat{\phi}(\widehat{p})\big).
\end{aligned}
$$

Then by Lemma 3.9,

$$
\sum_{\substack{p\in\operatorname{Per}_n(f)\\ V_n(p)\in\mathcal{K}}} w_n(p)\exp\big(S_n^f\phi(p)\big) \leqslant e^C \sum_{\substack{\widehat{p}\in\operatorname{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\big(S_m^{\widehat{f}}\widehat{\phi}(\widehat{p})\big),
$$

where $C := 2(N+K)\|\phi\|_\infty + C_1(\operatorname{diam}_d(S^2))^\beta$ and $C_1 \geqslant 0$ is the constant defined in (3.10) in Lemma 3.9 that depends only on f, $\mathcal{C}$, d, ϕ, and β. Thus by (1.2), we have

$$
\begin{aligned}
\log \Omega_n(\mathcal{K}) &= \log \sum_{\substack{p\in \operatorname{Per}_n(f)\\ V_n(p)\in\mathcal{K}}} w_n(p)\exp\bigl(S_n^f\phi(p)\bigr) - \log \sum_{p'\in\operatorname{Per}_n(f)} w_n(p')\exp\bigl(S_n^f\phi(p')\bigr)\\
&\leqslant \log \sum_{\substack{\widehat{p}\in \operatorname{Per}_m(\widehat{f})\\ \widehat{V}_m(\widehat{p})\in\widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\bigl(S_m^{\widehat{f}}\widehat{\phi}(\widehat{p})\bigr) - \log \sum_{p'\in\operatorname{Per}_n(f)} w_n(p')\exp\bigl(S_n^f\phi(p')\bigr) + C\\
&= \log \widehat{\Omega}_m(\widehat{\mathcal{K}}_{\varepsilon/2}) + \log \sum_{\widehat{p}'\in\operatorname{Per}_m(\widehat{f})} \exp\bigl(S_m^{\widehat{f}}\widehat{\phi}(\widehat{p}')\bigr) - \log \sum_{p'\in\operatorname{Per}_n(f)} w_n(p')\exp\bigl(S_n^f\phi(p')\bigr) + C,
\end{aligned}
$$

where $\bigl\{\widehat{\Omega}_j\bigr\}_{j\in\mathbb{N}}$ is defined by setting $w_j(x)=1$ for each $j\in\mathbb{N}$ and each $x\in S^2$ and replacing f with $\widehat{f}$ and ϕ with $\widehat{\phi}$ in the definition of $\{\Omega_j\}_{j\in\mathbb{N}}$ (recall (1.2)). Since $\widehat{f}$ has an $\widehat{f}$-invariant Jordan curve $\mathcal{C}\subseteq S^2$ with $\operatorname{post}\widehat{f}\subseteq\mathcal{C}$, Proposition 7.20 holds for $\widehat{f}$. Therefore, it follows from Proposition 7.7 and (7.17) that

$$
\begin{aligned}
\limsup_{n\to+\infty}\frac{1}{n}\log\Omega_n(\mathcal{K}) &\leqslant \frac{1}{K}\limsup_{m\to+\infty}\frac{1}{m}\log\widehat{\Omega}_m(\widehat{\mathcal{K}}_{\varepsilon/2}) + \frac{1}{K}P(\widehat{f},\widehat{\phi}) - P(f,\phi)\\
&\leqslant \frac{1}{K}\sup\biggl\{\widehat{F}_{\widehat{\phi}}(\widehat{\mu}) : \widehat{\mu}\in\mathcal{P}(S^2),\ \int\vec{\Phi}\,\mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon\biggr\}\\
&= \sup\biggl\{F_\phi(\mu) : \mu\in\mathcal{P}(S^2),\ \int\vec{\varphi}\,\mathrm{d}\mu > \vec{\alpha}-\varepsilon\biggr\}.
\end{aligned}
$$

Case 3 (Iterated preimages): $\xi_n = \Omega_n(x_n)$ for each $n\in\mathbb{N}$ (recall (1.3)), where $\{x_n\}_{n\in\mathbb{N}}$ is an arbitrary sequence of points in S^2.

We fix a point $x_0\in S^2$.

Let $N := n_f \in\mathbb{N}$ be the constant from Definition 3.20, which depends only on f and $\mathcal{C}$. For each $n\in\mathbb{N}$ and each $X^n\in\mathbf{X}^n(f,\mathcal{C})$, it follows from Lemma 7.6 (ii) that there exists a preimage of x_0 under f^{mK} in $\operatorname{inte}(X^n)$, where $m := \lceil (n+N)/K\rceil$. We fix a preimage of of x_0 under f^{mK} in $\operatorname{inte}(X^n)$ and denote it by $\widehat{y}(X^n)$. Then for each $n\in\mathbb{N}$, the map $X^n\mapsto\widehat{y}(X^n)$ from $\mathbf{X}^n(f,\mathcal{C})$ to $\widehat{f}^{-m}(x_0)$ is injective.

For each $n\in\mathbb{N}$ and each $y\in f^{-n}(x_n)$, let $X^n(y)\in\mathbf{X}^n(f,\mathcal{C})$ be an n-tile that contains y. By Proposition 3.2 (i), for each $n\in\mathbb{N}$ and each $X^n\in\mathbf{X}^n(f,\mathcal{C})$, $f^n|_{X^n}$ is a homeomorphism of X^n onto $f^n(X^n)$. This implies that for each integer $n\in\mathbb{N}$, the map $y\mapsto\widehat{y}(X^n(y))$ from $f^{-n}(x_n)$ to $\widehat{f}^{-m}(x_0)$ is injective, where $m=\lceil(n+N)/K\rceil$ and $\widehat{y}(X^n(y))\in\operatorname{inte}(X^n(y))$.

By similar arguments as in Case 2, there exists $n_0\in\mathbb{N}$ such that for each integer $n\geqslant n_0$, each $y\in f^{-n}(x_n)$ with $V_n(y)\in\mathcal{K}$, and each $X^n\in\mathbf{X}^n(f,\mathcal{C})$ with $y\in X^n$, it follows that $\widehat{V}_m(\widehat{y}(X^n))\in\widehat{\mathcal{K}}_{\varepsilon/2}$, where $m=\lceil(n+N)/K\rceil$.

Let $n\in\mathbb{N}$ be arbitrary. Consider $y'\in f^{-n}(x_n)\cap\mathbf{V}^n$, where $\mathbf{V}^n=\mathbf{V}^n(f,\mathcal{C})$ is the set of n-vertices. We set $\mathbf{X}^n(f,\mathcal{C},y') := \{X\in\mathbf{X}^n(f,\mathcal{C}) : y'\in X\}$. By Remark 3.4, we have $\overline{W}^n(y') = \bigcup\mathbf{X}^n(f,\mathcal{C},y')$ and $\operatorname{card}(\mathbf{X}^n(f,\mathcal{C},y')) = 2\deg_{f^n}(y')$, where $W^n(y')$ is defined in (3.6) and $\overline{W}^n(y')$ is the closure of $W^n(y')$. Note that $X^n(y)\notin\mathbf{X}^n(f,\mathcal{C},y')$ for every $y\in f^{-n}(x_n)\setminus\mathbf{V}^n(f,\mathcal{C})$.

We now prove the upper bound. The proof is essentially the same as in Case 2, and we retain this proof for the convenience of the reader.

Let $\{w_j\}_{j\in\mathbb{N}}$ be an arbitrary sequence of real-valued functions on S^2 with $w_j(x)\in\bigl[1,\deg_{f^j}(x)\bigr]$ for each $j\in\mathbb{N}$ and each $x\in S^2$. For each integer $n\in\mathbb{N}$, we set $m:=\lceil(n+N)/K\rceil$. Note that

$0 \leqslant mK - n \leqslant N + K$. By the arguments above, for each integer $n \geqslant n_0$, we have

$$\begin{aligned}
&\sum_{\substack{y \in f^{-n}(x_n) \\ V_n(y) \in \mathcal{K}}} w_n(y) \exp\bigl(S_n^f \phi(y)\bigr) \\
&\quad \leqslant \sum_{\substack{y' \in f^{-n}(x_n) \cap \mathbf{V}^n \\ V_n(y') \in \mathcal{K}}} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) + \sum_{\substack{y \in f^{-n}(x_n) \setminus \mathbf{V}^n \\ V_n(y) \in \mathcal{K}}} \exp\bigl(S_n^f \phi(y)\bigr) \\
&\quad \leqslant \sum_{\substack{y' \in f^{-n}(x_n) \cap \mathbf{V}^n \\ V_n(y') \in \mathcal{K}}} \sum_{X^n \in \mathbf{X}^n(f, \mathcal{C}, y')} e^{D_n(\phi) + S_n^f \phi(\widehat{y}(X^n))} + \sum_{\substack{y \in f^{-n}(x_n) \setminus \mathbf{V}^n \\ V_n(y) \in \mathcal{K}}} e^{D_n(\phi) + S_n^f \phi(\widehat{y}(X^n(y)))} \\
&\quad \leqslant e^{D_n(\phi)} \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\bigl(S_n^f \phi(\widehat{y})\bigr) \\
&\quad \leqslant e^{D_n(\phi)} e^{2(N+K)\|\phi\|_\infty} \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y})\bigr).
\end{aligned}$$

Then by Lemma 3.9,

$$\sum_{\substack{y \in f^{-n}(x_n) \\ V_n(y) \in \mathcal{K}}} w_n(y) \exp\bigl(S_n^f \phi(y)\bigr) \leqslant e^C \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y})\bigr),$$

where the constant C is the same as in Case 2. Thus by (1.3), we have

$$\begin{aligned}
&\log \Omega_n(x_n)(\mathcal{K}) \\
&\quad = \log \sum_{\substack{y \in f^{-n}(x_n) \\ V_n(y) \in \mathcal{K}}} w_n(y) \exp\bigl(S_n^f \phi(y)\bigr) - \log \sum_{y' \in f^{-n}(x_n)} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) \\
&\quad \leqslant \log \sum_{\substack{\widehat{y} \in \widehat{f}^{-m}(x_0) \\ \widehat{V}_m(\widehat{y}) \in \widehat{\mathcal{K}}_{\varepsilon/2}}} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y})\bigr) - \log \sum_{y' \in f^{-n}(x_n)} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) + C \\
&\quad = \log \widehat{\Omega}_m(x_0)(\widehat{\mathcal{K}}_{\varepsilon/2}) + \log \sum_{\widehat{y}' \in \widehat{f}^{-m}(x_0)} \exp\bigl(S_m^{\widehat{f}} \widehat{\phi}(\widehat{y}')\bigr) - \log \sum_{y' \in f^{-n}(x_n)} w_n(y') \exp\bigl(S_n^f \phi(y')\bigr) + C,
\end{aligned}$$

where $\bigl\{\widehat{\Omega}_j(x_0)\bigr\}_{j \in \mathbb{N}}$ is defined by setting $w_j = \mathbb{1}_{S^2}$ and $x_j = x_0$ for each $j \in \mathbb{N}$ and replacing f with $\widehat{f}$ and ϕ with $\widehat{\phi}$ in the definition of $\{\Omega_j(x_j)\}_{j \in \mathbb{N}}$ (recall (1.3)). Since $\widehat{f}$ has an $\widehat{f}$-invariant Jordan curve $\mathcal{C} \subseteq S^2$ with post $\widehat{f} \subseteq \mathcal{C}$, Proposition 7.20 holds for $\widehat{f}$. Therefore, it follows from Proposition 7.9 and (7.17) that

$$\begin{aligned}
\limsup_{n \to +\infty} \frac{1}{n} \log \Omega_n(x_n)(\mathcal{K}) &\leqslant \frac{1}{K} \limsup_{m \to +\infty} \frac{1}{m} \log \widehat{\Omega}_m(x_0)(\widehat{\mathcal{K}}_{\varepsilon/2}) + \frac{1}{K} P(\widehat{f}, \widehat{\phi}) - P(f, \phi) \\
&\leqslant \frac{1}{K} \sup \biggl\{ \widehat{F}_{\widehat{\phi}}(\widehat{\mu}) : \widehat{\mu} \in \mathcal{P}(S^2), \int \vec{\Phi} \, \mathrm{d}\widehat{\mu} > K\vec{\alpha} - K\varepsilon \biggr\} \\
&= \sup \biggl\{ F_\phi(\mu) : \mu \in \mathcal{P}(S^2), \int \vec{\varphi} \, \mathrm{d}\mu > \vec{\alpha} - \varepsilon \biggr\}.
\end{aligned}$$

The proof is complete. □

7.5.3. *End of proof of the upper bound.*

Proof of Proposition 7.16. Let $\mathcal{K}$ be a closed subset of $\mathcal{P}(S^2)$. Let $\mathcal{G} \subseteq \mathcal{P}(S^2)$ be an open set containing $\mathcal{K}$. Since $\mathcal{P}(S^2)$ is metrizable and compact in the weak*-topology (see for example, [Wal82, Theorems 6.4 and 6.5]) and $\mathcal{K}$ is compact, we can choose $\varepsilon > 0$ and finitely many closed sets $\mathcal{K}_1, \dots, \mathcal{K}_s$ of the form $\mathcal{K}_j = \{\mu \in \mathcal{P}(S^2) : \int \vec{\varphi}_j \,\mathrm{d}\mu \geqslant \vec{\alpha}_j\}$ with $\ell_j \in \mathbb{N}$, $\vec{\varphi}_j \in C(S^2)^{\ell_j}$, and $\vec{\alpha}_j \in \mathbb{R}^{\ell_j}$, so that $\mathcal{K} \subseteq \bigcup_{j=1}^s \mathcal{K}_j \subseteq \bigcup_{j=1}^s \mathcal{K}_j(\varepsilon) \subseteq \mathcal{G}$, where $\mathcal{K}_j(\varepsilon) \coloneqq \{\mu \in \mathcal{P}(S^2) : \int \vec{\varphi}_j \,\mathrm{d}\mu > \vec{\alpha}_j - \varepsilon\}$. For each $j \in \{1, \dots, s\}$, it follows from Proposition 7.20 that

$$\limsup_{n \to +\infty} \frac{1}{n} \log \xi_n(\mathcal{K}_j) \leqslant \sup_{\mathcal{K}_j(\varepsilon)} F_\phi.$$

for each sequence $\{\xi_n\}_{n\in\mathbb{N}} \in \{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\}$. Hence,

$$\limsup_{n \to +\infty} \frac{\log \xi_n(\mathcal{K})}{n} \leqslant \limsup_{n \to +\infty} \frac{1}{n} \log \xi_n\left(\bigcup_{j=1}^s \mathcal{K}_j\right) \leqslant \max_{1 \leqslant j \leqslant s} \limsup_{n \to +\infty} \frac{\log \xi_n(\mathcal{K}_j)}{n} \leqslant \max_{1 \leqslant j \leqslant s} \sup_{\mathcal{K}_j(\varepsilon)} F_\phi \leqslant \sup_{\mathcal{G}} F_\phi.$$

Since $\mathcal{G}$ is an arbitrary open set containing $\mathcal{K}$, it follows from Remark 1.4 that

$$\limsup_{n \to +\infty} \frac{1}{n} \log \xi_n(\mathcal{K}) \leqslant \inf_{\mathcal{G} \supseteq \mathcal{K}} \sup_{\mathcal{G}} F_\phi = \inf_{\mathcal{G} \supseteq \mathcal{K}} \sup_{\mathcal{G}} (-I_\phi) = -\inf_{\mathcal{K}} I_\phi,$$

where the last equality is due to the lower semi-continuity of I_ϕ. □

7.6. **Proof of large deviation principles.** In this subsection, we finish the proof of Theorem 1.3 and its corollaries, Corollaries 1.5 and 1.6.

We record the following well-known lemma, sometimes known as the *Portmanteau Theorem*, and refer the reader to [Bil13, Theorem 2.1] for a proof.

Lemma 7.21. *Let (X, d) be a compact metric space, and μ and μ_i, for $i \in \mathbb{N}$, be Borel probability measures on X. Then the following statements are equivalent:*

(i) $\mu_i \xrightarrow{w^*} \mu$ *as* $i \to +\infty$;
(ii) $\limsup_{i \to +\infty} \mu_i(E) \leqslant \mu(E)$ *for each closed set* $E \subseteq X$;
(iii) $\liminf_{i \to +\infty} \mu_i(G) \geqslant \mu(G)$ *for each open set* $G \subseteq X$;
(iv) $\lim_{i \to +\infty} \mu_i(B) \leqslant \mu(B)$ *for each Borel set* $B \subseteq X$ *with* $\mu(\partial B) = 0$.

Proof of Theorem 1.3. We fix arbitrary sequence $\{\xi_n\}_{n\in\mathbb{N}} \in \{\{\Sigma_n\}_{n\in\mathbb{N}}, \{\Omega_n\}_{n\in\mathbb{N}}, \{\Omega_n(x_n)\}_{n\in\mathbb{N}}\}$.

By Propositions 7.10 and 7.16, $\{\xi_n\}_{n\in\mathbb{N}}$ satisfies a large deviation principle with the rate function I_ϕ as defined in (1.4).

By Theorem 7.5, μ_ϕ is the unique minimizer of the rate function I_ϕ.

To prove that $\{\xi_n\}_{n\in\mathbb{N}}$ converges to δ_{μ_ϕ} in the weak* topology, by Lemma 7.21 (i) and (ii), it suffices to show that $\limsup_{n\to+\infty} \xi_n(\mathcal{K}) \leqslant 0$ for each closed set $\mathcal{K} \subseteq \mathcal{P}(S^2) \setminus \{\mu_\phi\}$. Let $\mathcal{K}$ be a closed set in $\mathcal{P}(S^2)$ with $\mu_\phi \notin \mathcal{K}$. Indeed, since I_ϕ is lower semi-continuous, non-negative, and it vanishes precisely on $\{\mu_\phi\}$ by Theorem 7.5, the infimum of I_ϕ on $\mathcal{K}$ is attained at some point of $\mathcal{K}$, and thus $\inf_{\mathcal{K}} I_\phi > 0$. Therefore, it follows immediately from the large deviation upper bounds that

$$\limsup_{n \to +\infty} \frac{1}{n} \log \xi_n(\mathcal{K}) \leqslant -\inf_{\mathcal{K}} I_\phi < 0$$

and $\lim_{n\to+\infty} \xi_n(\mathcal{K}) = 0$.

To prove the last statement of the theorem, let $\mathcal{G} \subseteq \mathcal{P}(S^2)$ be a convex and open set containing an invariant measure μ'. Since the rate function I_ϕ is lower semi-continuous, and since it takes finite values precisely on the compact set $\mathcal{M}(S^2, f)$ by (1.4), there exists $\mu \in \overline{\mathcal{G}} \cap \mathcal{M}(S^2, f)$ such that

$I_\phi(\mu) = \inf_{\overline{\mathcal{G}}} I_\phi$. For each $t \in (0,1)$, put $\mu_t := (1-t)\mu + t\mu'$, and note that $\mu_t \in \mathcal{M}(S^2, f)$ and $\mu_t \in \mathcal{G}$ (see for example, [Sch71, 1.1, p. 38]). Since I_ϕ is convex (recall Remark 1.4), we have

$$\inf_{\mathcal{G}} I_\phi \leqslant \liminf_{t\to 0} I_\phi(\mu_t) \leqslant I_\phi(\mu) = \inf_{\overline{\mathcal{G}}} I_\phi.$$

This shows that $\inf_{\mathcal{G}} I_\phi = \inf_{\overline{\mathcal{G}}} I_\phi$. Hence, $\mathcal{G}$ is a I_ϕ-continuity set and the last assertion of the theorem follows immediately from (7.3) in Subsection 7.1.

The proof is complete. □

We now prove the corollaries of Theorem 1.3, as stated in Corollaries 1.5 and 1.6.

Proof of Corollary 1.5. Fix $\mu \in \mathcal{M}(S^2, f)$ and a convex local basis G_μ at μ. We show that (1.7) in Corollary 1.5 holds. Since the rate function I_ϕ is lower semi-continuous (recall Remark 1.4), we get

$$-I_\phi(\mu) = \inf_{\mathcal{G}\in G_\mu} \sup_{\mathcal{G}} (-I_\phi) = \inf_{\mathcal{G}\in G_\mu} \bigl(-\inf_{\mathcal{G}} I_\phi\bigr).$$

Then it follows from (1.6) in Theorem 1.3 that

$$\begin{aligned} -I_\phi(\mu) &= \inf_{\mathcal{G}\in G_\mu} \biggl\{ \lim_{n\to+\infty} \frac{1}{n} \log \mu_\phi(\{x \in S^2 : V_n(x) \in \mathcal{G}\}) \biggr\} \\ &= \inf_{\mathcal{G}\in G_\mu} \biggl\{ \lim_{n\to+\infty} \frac{1}{n} \log \sum_{p\in \operatorname{Per}_n(f), V_n(p)\in\mathcal{G}} \frac{w_n(p)\exp(S_n\phi(p))}{Z_n(\phi)} \biggr\} \\ &= \inf_{\mathcal{G}\in G_\mu} \biggl\{ \lim_{n\to+\infty} \frac{1}{n} \log \sum_{y\in f^{-n}(x_n), V_n(y)\in\mathcal{G}} \frac{w_n(y)\exp(S_n\phi(y))}{Z'_n(\phi)} \biggr\}, \end{aligned}$$

where we write $Z_n(\phi) := \sum_{x\in\operatorname{Per}_n(f)} w_n(x)\exp(S_n\phi(x))$ and $Z'_n(\phi) := \sum_{y\in f^{-n}(x_n)} w_n(y)\exp(S_n\phi(y))$. Note that by Propositions 7.7 and 7.9 we have $P(f,\phi) = \lim_{n\to+\infty} \frac{1}{n}\log Z_n(\phi) = \lim_{n\to+\infty} \frac{1}{n}\log Z'_n(\phi)$. Thus (1.7) holds. □

Proof of Corollary 1.6. The first assertion follows immediately from Theorems 1.3 and 7.1.

To establish (1.9), we consider an arbitrary interval $K \subseteq \mathbb{R}$ that intersects (c_ψ, d_ψ). Note that the rate function J defined by (1.8) is bounded on $[c_\psi, d_\psi]$ and constantly equal to $+\infty$ on $\mathbb{R} \setminus [c_\psi, d_\psi]$. Furthermore, it follows from the convexity of I_ϕ that J is convex on $\mathbb{R}$, and therefore continuous on (c_ψ, d_ψ). This implies that $\inf_{\operatorname{int}(K)} J = \inf_{\overline{K}} J$ since $K \cap (c_\psi, d_\psi) \neq \emptyset$. Thus (1.9) follows from (7.1) and (7.2).

Note that (1.10) follows immediately from the definitions of I_ϕ and F_ϕ (see (1.4) and (1.5)). We now assume that ψ is Hölder continuous with respect to the visual metric d. Consider arbitrary $\alpha \in (c_\psi, d_\psi)$ and $\varepsilon > 0$. To show that $J(\alpha) = \widetilde{J}(\alpha)$, we define $\mathcal{G}_\varepsilon := \bigcup_{t\in(\alpha-\varepsilon,\alpha+\varepsilon)} \mathcal{K}(t)$, where

$$\mathcal{K}(t) := \biggl\{ \nu \in \mathcal{P}(S^2) : \int \psi \,\mathrm{d}\nu = t \biggr\}.$$

Then for each $\mu \in \mathcal{K}(\alpha)$, the open set $\mathcal{G}_\varepsilon \subseteq \mathcal{P}(S^2)$ contains μ, and it follows from (1.4) that

$$I_\phi(\mu) = -\inf_{\mathcal{G}\ni\mu} \sup_{\mathcal{G}} F_\phi = \sup_{\mathcal{G}\ni\mu} \inf_{\mathcal{G}} (-F_\phi) \geqslant \inf_{\mathcal{G}_\varepsilon} (-F_\phi) = \inf_{t\in(\alpha-\varepsilon,\alpha+\varepsilon)} \inf_{\mathcal{K}(t)} (-F_\phi) = \inf_{t\in(\alpha-\varepsilon,\alpha+\varepsilon)} \widetilde{J}(t).$$

Mimicking the proof of [LSZ25, Proposition 7.1 (iii)], we can show that the function $\widetilde{J}$ defined in (1.10) is continuous on (c_ψ, d_ψ) (although [LSZ25, Proposition 7.1 (iii)] additionally assumes that $\psi = \phi$, the proof for the general case is verbatim the same). This implies that $\inf_{t\in(\alpha-\varepsilon,\alpha+\varepsilon)} \widetilde{J}(t) \to \widetilde{J}(\alpha)$ as $\varepsilon \to 0$. Thus we conclude that

$$J(\alpha) = \inf_{\mu\in\mathcal{K}(\alpha)} I_\phi(\mu) \geqslant \widetilde{J}(\alpha) \geqslant J(\alpha).$$

This completes the proof. □

7.7. **Equidistribution with respect to the equilibrium state.** We finish this section with an equidistribution result, as a consequence of level-2 large deviation principles.

Proof of Theorem 1.7. Recall that $V_n(x) = \frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i(x)}$ for $x \in S^2$ and $n \in \mathbb{N}$ as defined in (1.1). For each $n \in \mathbb{N}$ and each open set $\mathcal{G} \subseteq \mathcal{P}(S^2)$, we write

$$Z_n^+(\mathcal{G}) \coloneqq \sum_{y\in f^{-n}(x_n), V_n(y)\in\mathcal{G}} w_n(y) e^{S_n\phi(y)} \quad \text{and} \quad Z_n^-(\mathcal{G}) \coloneqq \sum_{y\in f^{-n}(x_n), V_n(y)\notin\mathcal{G}} w_n(y) e^{S_n\phi(y)}.$$

Let G_{μ_ϕ} be a convex local basis of $\mathcal{P}(S^2)$ at μ_ϕ. We fix an arbitrary convex open set $\mathcal{G} \in G_{\mu_\phi}$.

Recall that μ_ϕ is the unique minimizer of the rate function I_ϕ by Theorem 7.5. Then it follows from Corollary 1.5 that for each $\mu \in \mathcal{P}(S^2) \setminus \{\mu_\phi\}$, there exist numbers $a_\mu < P(f,\phi)$ and $N_\mu \in \mathbb{N}$ and an open neighborhood $\mathcal{G}_\mu \subseteq \mathcal{P}(S^2) \setminus \{\mu_\phi\}$ containing μ such that for each $n \geqslant N_\mu$,

$$Z_n^+(\mathcal{G}_\mu) \leqslant \exp(na_\mu). \tag{7.18}$$

Since $\mathcal{P}(S^2)$ is compact in the weak* topology by Alaoglu's theorem, so is $\mathcal{P}(S^2)\setminus\mathcal{G}$. Thus there exists a finite set $\{\mu_i : i \in I\} \subseteq \mathcal{P}(S^2)\setminus\mathcal{G}$ (where I is a finite index set) such that

$$\mathcal{P}(S^2)\setminus\mathcal{G} \subseteq \bigcup_{i\in I}\mathcal{G}_{\mu_i}. \tag{7.19}$$

Set $a \coloneqq \max\{a_{\mu_i} : i \in I\}$. Note that $a < P(f,\phi)$. Applying Corollary 1.5 with $\mu = \mu_\phi$ and noting that $I_\phi(\mu_\phi) = 0$ (recall (1.4) and (1.5) in Theorem 1.3), we get

$$P(f,\phi) \leqslant \varliminf_{n\to+\infty}\frac{1}{n}\log Z_n^+(\mathcal{G}). \tag{7.20}$$

Combining (7.20) with Proposition 7.9, we get that the equality holds in (7.20). So there exist numbers $b \in (a, P(f,\phi))$ and $N \geqslant \max\{N_{\mu_i} : i \in I\}$ such that for each integer $n \geqslant N$,

$$Z_n^+(\mathcal{G}) \geqslant \exp(nb). \tag{7.21}$$

We claim that every subsequential limit of $\{\nu_n\}_{n\in\mathbb{N}}$ in the weak* topology lies in the closure $\overline{\mathcal{G}}$ of $\mathcal{G}$. Assuming that the claim holds, then since $\mathcal{G} \in G_{\mu_\phi}$ is arbitrary, we get that any subsequential limit of $\{\nu_n\}_{n\in\mathbb{N}}$ in the weak* topology is μ_ϕ, i.e., $\nu_n \xrightarrow{w^*} \mu_\phi$ as $n \to +\infty$.

We now prove the claim. We first observe that for each $n \in \mathbb{N}$,

$$\begin{aligned}\nu_n &= \sum_{y\in f^{-n}(x_n)} \frac{w_n(y)\exp(S_n\phi(y))}{Z_n^+(\mathcal{G}) + Z_n^-(\mathcal{G})} V_n(y) \\ &= \frac{Z_n^+(\mathcal{G})}{Z_n^+(\mathcal{G}) + Z_n^-(\mathcal{G})}\nu_n' + \sum_{y\in f^{-n}(x_n), V_n(y)\notin\mathcal{G}} \frac{w_n(y)\exp(S_n\phi(y))}{Z_n^+(\mathcal{G}) + Z_n^-(\mathcal{G})} V_n(y),\end{aligned}$$

where

$$\nu_n' \coloneqq \sum_{y\in f^{-n}(x_n), V_n(y)\in\mathcal{G}} \frac{w_n(y)\exp(S_n\phi(y))}{Z_n^+(\mathcal{G})} V_n(y).$$

Note that since $a < b$, it follows from (7.19), (7.18), and (7.21) that

$$0 \leqslant \varlimsup_{n\to+\infty}\frac{Z_n^-(\mathcal{G})}{Z_n^+(\mathcal{G})} \leqslant \varlimsup_{n\to+\infty}\frac{\sum_{i\in I} Z_n^+(\mathcal{G}_\mu)}{Z_n^+(\mathcal{G})} \leqslant \varlimsup_{n\to+\infty}\frac{\operatorname{card}(I)\exp(na)}{\exp(nb)} = 0.$$

So $\lim_{n\to+\infty}\frac{Z_n^+(\mathcal{G})}{Z_n^+(\mathcal{G})+Z_n^-(\mathcal{G})} = 1$, and that the total variation

$$\begin{aligned}&\Bigg\|\sum_{y\in f^{-n}(x_n), V_n(y)\notin\mathcal{G}} \frac{w_n(y)\exp(S_n\phi(y))}{Z_n^+(\mathcal{G}) + Z_n^-(\mathcal{G})} V_n(y)\Bigg\| \\ &\qquad \leqslant \frac{\sum_{y\in f^{-n}(x_n), V_n(y)\notin\mathcal{G}} w_n(y)\exp(S_n\phi(y))\|V_n(y)\|}{Z_n^+(\mathcal{G}) + Z_n^-(\mathcal{G})} \leqslant \frac{Z_n^-(\mathcal{G})}{Z_n^+(\mathcal{G}) + Z_n^-(\mathcal{G})} \longrightarrow 0\end{aligned}$$

as $n \to +\infty$. Thus a measure is a subsequential limit of $\{\nu_n\}_{n\in\mathbb{N}}$ if and only if it is a subsequential limit of $\{\nu'_n\}_{n\in\mathbb{N}}$. Note that for each $n \in \mathbb{N}$, ν'_n is a convex combination of measures in the convex set $\mathcal{G}$, so $\nu'_n \in \mathcal{G}$. Hence each subsequential limit of $\{\nu_n\}_{n\in\mathbb{N}}$ lies in the closure $\overline{\mathcal{G}}$ of $\mathcal{G}$. The proof of the claim is complete now.

By similar arguments as in the proof of the convergence of $\{\nu_n\}_{n\in\mathbb{N}}$ above, with Proposition 7.9 replaced by Proposition 7.7, we get that $\mu_n \xrightarrow{w^*} \mu_\phi$ as $n \to +\infty$. □

References

[Bil13] Billingsley, P., *Convergence of probability measures*, John Wiley & Sons, New York, 2013.

[BLL] Bonk, M., Li, W., and Li, Z., "A dynamical version of the Kapovich–Kleiner conjecture", In preparation.

[BM10] Bonk, M. and Meyer, D., *Expanding Thurston Maps*, 2010, arXiv: 1009.3647v1 [math.DS].

[BM17] Bonk, M. and Meyer, D., *Expanding Thurston Maps*, vol. 225, Amer. Math. Soc., Providence, RI, 2017.

[Buz14] Buzzi, J., surface diffeomorphisms with no maximal entropy measure, *Ergodic Theory and Dynamical Systems* 34 (2014), 1770–1793.

[BCS22] Buzzi, J., Crovisier, S., and Sarig, O., Continuity properties of Lyapunov exponents for surface diffeomorphisms, *Invent. Math.* 230 (2022), 767–849.

[Can94] Cannon, J. W., The combinatorial Riemann mapping theorem, *Acta Math.* 173 (1994), 155–234.

[CRLT19] Chung, Y. M., Rivera-Letelier, J., and Takahasi, H., Large deviation principle in one-dimensional dynamics, *Invent. Math.* 218 (2019), 853–888.

[CRL11] Comman, H. and Rivera-Letelier, J., Large deviation principles for non-uniformly hyperbolic rational maps, *Ergodic Theory Dynam. Systems* 31 (2011), 321–349.

[DPTUZ21] Das, T., Przytycki, F., Tiozzo, G., Urbański, M., and Zdunik, A., Thermodynamic formalism for coarse expanding dynamical systems, *Comm. Math. Phys.* 384 (2021), 165–199.

[DZ09] Dembo, A. and Zeitouni, O., *Large deviations techniques and applications*, vol. 38, Springer, New York, 2009.

[DH93] Douady, A. and Hubbard, J. H., A proof of Thurston's topological characterization of rational functions, *Acta Math.* 171 (1993), 263–297.

[EKW94] Eizenberg, A., Kifer, Y., and Weiss, B., Large deviations for $\mathbb{Z}^d$-actions, *Comm. Math. Phys.* 164 (1994), 433–454.

[Ell12] Ellis, R. S., *Entropy, large deviations, and statistical mechanics*, vol. 271, Springer, New York, 2012.

[Fol99] Folland, G. B., *Real analysis*, 2nd. ed., Wiley, New York, 1999.

[FO88] Follmer, H. and Orey, S., Large deviations for the empirical field of a Gibbs measure, *Ann. Probab.* 16 (1988), 961–977.

[HP09] Haïssinsky, P. and Pilgrim, K. M., Coarse expanding conformal dynamics, *Astérisque* (2009).

[Hat02] Hatcher, A., *Algebraic Topology*, Cambridge Univ. Press, Cambridge, 2002.

[IJ15] Iommi, G. and Jordan, T., Multifractal analysis of Birkhoff averages for countable Markov maps, *Ergodic Theory Dynam. Systems* 35 (2015), 2559–2586.

[ITV22] Iommi, G., Todd, M., and Velozo, A., Escape of entropy for countable Markov shifts, *Adv. Math.* 405 (2022), 108507.

[JMU05] Jenkinson, O., Mauldin, R. D., and Urbański, M, Zero temperature limits of Gibbs-equilibrium states for countable alphabet subshifts of finite type, *J. Stat. Phys.* 119 (2005), 765–776.

[KH95] Katok, A. and Hasselblatt, B., *Introduction to the Modern Theory of Dynamical Systems*, Cambridge Univ. Press, Cambridge, 1995.

[Kif90] Kifer, Y., Large deviations in dynamical systems and stochastic processes, *Trans. Amer. Math. Soc.* 321 (1990), 505–524.

[Li15] Li, Z., Weak expansion properties and large deviation principles for expanding Thurston maps, *Adv. Math.* 285 (2015), 515–567.

[Li17] Li, Z., *Ergodic Theory of Expanding Thurston Maps*, Atlantis Series in Dynamical Systems, Springer, 2017.

[Li18] Li, Z., Equilibrium states for expanding Thurston maps, *Comm. Math. Phys.* 357 (2018), 811–872.

[LS24] Li, Z. and Shi, X., *Thermodynamic formalism for subsystems of expanding Thurston maps II*, 2024, arXiv: `2404.07247 [math.DS]`.

[LSZ25] Li, Z., Shi, X., and Zhang, Y., Thermodynamic formalism for subsystems of expanding Thurston maps and large deviations asymptotics, *Proc. Lond. Math. Soc. (3)* 130 (2025), e70019.

[LZ24] Li, Z. and Zhang, Y., Ground states and periodic orbits for expanding Thurston maps, *Math. Ann.* (2024).

[Lyu83] Lyubich, M. Yu., Entropy properties of rational endomorphisms of the Riemann sphere, *Ergodic Theory Dynam. Systems* 3 (1983), 351–385.

[Mis73] Misiurewicz, M., Diffeomorphism without any measure with maximal entropy, *Bull. Acad. Pol. Sci.* 21 (1973), 903–910.

[Mis76] Misiurewicz, M., Topological conditional entropy, *Studia Math.* 55 (1976), 175–200.

[New89] Newhouse, S. E., Continuity properties of entropy, *Ann. of Math. (2)* 129 (1989), 215–235.

[PS05] Pfister, C.-E. and Sullivan, W. G., Large deviations estimates for dynamical systems without the specification property. Application to the β-shifts, *Nonlinearity* 18 (2005), 237–261.

[Prz85] Przytycki, F., Hausdorff dimension of harmonic measure on the boundary of an attractive basin for a holomorphic map, *Invent. Math.* 80 (1985), 161–179.

[RAS15] Rassoul-Agha, F. and Seppäläinen, T., *A course on large deviations with an introduction to Gibbs measures*, vol. 162, Amer. Math. Soc., Providence, RI, 2015.

[Sch71] Schaefer, H. H., *Topological vector spaces*, Springer, New York, 1971.

[Sig74] Sigmund, K., On dynamical systems with the specification property, *Trans. Amer. Math. Soc.* 190 (1974), 285–299.

[Tak19] Takahasi, H., Large deviation principles for countable Markov shifts, *Trans. Amer. Math. Soc.* 372 (2019), 7831–7855.

[Tak20] Takahasi, H., Entropy-approachability for transitive Markov shifts over infinite alphabet, *Proc. Amer. Math. Soc.* 148 (2020), 3847–3857.

[Wal82] Walters, P., *An Introduction to Ergodic Theory*, vol. 79, Springer, New York–Berlin, 1982.

[Yam09] Yamamoto, K., On the weaker forms of the specification property and their applications, *Proc. Amer. Math. Soc.* 137 (2009), 3807–3814.

[Yom87] Yomdin, Y., Volume growth and entropy, *Israel J. Math.* 57 (1987), 285–300.

[You90] Young, L.-S., Large deviations in dynamical systems, *Trans. Amer. Math. Soc.* 318 (1990), 525–543.

Zhiqiang Li, School of Mathematical Sciences & Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China

Email address: `zli@math.pku.edu.cn`

Xianghui Shi, Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China

Email address: `xhshi@pku.edu.cn`