THERMODYNAMIC FORMALISM FOR CORRESPONDENCES

XIAORAN LI, ZHIQIANG LI, AND YIWEI ZHANG

Abstract. In this work, we investigate the Variational Principle and develop a thermodynamic formalism for correspondences. We define the measure-theoretic entropy for transition probability kernels and topological pressure for correspondences. Based on these two notions, we establish the following results:

The Variational Principle holds and equilibrium states exist for continuous potential functions, provided that the correspondence satisfies some expansion property called forward expansiveness. If, in addition, the correspondence satisfies the specification property and the potential function is Bowen summable, then the equilibrium state is unique. On the other hand, for a distance-expanding, open, strongly transitive correspondence and a Hölder continuous potential function, there exists a unique equilibrium state and the backward orbits are equidistributed. Furthermore, we investigate the Variational Principle for general correspondences.

In conformal dynamics, we establish the Variational Principle for the Lee–Lyubich–Markorov–Mazor–Mukherjee anti-holomorphic correspondences, which are matings of some anti-holomorphic rational maps with anti-Hecke groups and not forward expansive. We also show a Ruelle–Perron–Frobenius Theorem for a family of hyperbolic holomorphic correspondences of the form $\mathbf{f}_c(z) = z^{q/p} + c$.

Contents

2020 *Mathematics Subject Classification.* Primary: 37F05; Secondary: 54C60, 37D35, 37C30, 37D20.

Key words and phrases. thermodynamic formalism, correspondence, variational principle, equilibrium state, Ruelle operator, anti-holomorphic.

Z. Li and X. Li were partially supported by NSFC Nos. 12101017, 12090010, 12090015, and BJNSF No. 1214021. X. Li was also supported by Peking University Funding Nos. 7101303303 and 6201001891. Y. Zhang was partially supported by NSFC Nos. 12161141002, 11871262, and 12271432.

1. Introduction

Correspondences. A *correspondence*[1] T on a compact metric space X is a map from X to the set consisting of all non-empty closed subsets of X with the property that the set $\{(x_1, x_2) \in X^2 : x_2 \in T(x_1)\}$ is closed in X^2. As a natural generalization of (single-valued) continuous maps, correspondences appear abundantly in control theory [Po21], differential games [Pe93], mathematical economics and game

[1]There exist notions in the literature related to correspondences, such as upper semi-continuous set-valued functions in [KT17], set-valued maps in [RT18], and closed relations in [MA99]. Our notion of correspondence coincides with the first one but differs slightly from the other two.

theory [CPMP08], qualitative physics and viability [For88], and continuous selections [Mic56a, Mic56b, Mic57]. To quote from the monograph of J. P. Aubin and H. Frankowska [AF09]: "Who needs set-valued analysis? Everyone, we are tempted to say." Indeed, they listed eight famous examples of correspondences studied by J. Hadamard, J. von Neumann, K. Kuratowski, E. Michael, T. Ważewski, V. V. Filippov, and many other mathematicians, ranging from eight different mathematical subjects mentioned in [AF09, Introduction].

To the best of our knowledge, the studies on correspondences originated from dealing with ill-posed problems (for partial differential equations) in the sense of J. Hadamard [Ha02]. Here by ill-posed problems, we mean that the existence of a solution or the uniqueness of the solution fails for some choice of data. This was indeed noticed during the first three decades of the 20th century by founders of "Functional Calculus", such as P. Painlevé, F. Hausdorff, G. Bouligand, and K. Kuratowski to quote only a few. In his remarkable book Topologie [Ku66], K. Kuratowski gave set-valued maps their proper status. Since then, the study of correspondences (known as set-valued analysis) has been increasing rapidly. Many fundamental concepts of single-valued analysis such as limits, differentiation, integral, and fixed point theorems have been adapted to the set-valued realm, see the monograph [AF09] and references therein.

Other than general correspondences, the study of holomorphic and anti-holomorphic correspondences attracts its independent interests in complex dynamics.

The study of (anti-)holomorphic correspondences dates back at least to P. Fatou [Fa29]. Indeed, P. Fatou observed similarities between limit sets of Kleinian groups and Julia sets of rational maps in the 1920s, and proposed the following question [Fa29]:

"L'analogie remarquée entre les ensembles de points limites des groupes Kleinéens-et ceux qui sont constitués par les frontières des régions de convergence des itérées d'une fonction rationnelle ne paraît d'ailleurs pas fortuite et il serait probablement possible d'en faire la syntèse dans une théorie générale des groupes discontinus des substitutions algrébriques."

About the analogy between Kleinian groups and rational maps, D. Sullivan discovered deep connections between the iteration theory of rational maps and the theory of Kleinian groups (see [Su85, Introduction]), which became known as *Sullivan's dictionary*. Since then, there have been considerable efforts to draw direct connections between these two branches of conformal dynamics. See for example, the works of S. Bullett and C. Penrose [BP94], C. T. McMullen [Mc95, Mc96], M. Yu. Lyubich and Y. Minsky [LM97], P. Haïssinsky and K. M. Pilgrim [HP09], M. Bonk and D. Meyer [BM10, BM17], M. Mj and S. Mukherjee [MM23], and references therein.

Our current work is partially motivated by the interest of the community including M. Bonk, D. Meyer, S. Rohde, etc., to extend Sullivan's dictionary to some fractals arising from probability theory, hoping to transplant key analytic tools and techniques to such settings, and partially motivated by the recent works of S. Lee, M. Yu. Lyubich, N. G. Makarov, S. Mukherjee, etc., on certain anti-holomorphic correspondences which will be discussed below.

Apart from the analogy from Sullivan's dictionary, to answer Fatou's question, we need to "naturally" combine the dynamics of a rational map with that of a Kleinian group. Matings between Kleinian groups and rational maps developed by S. Bullett, C. Penrose, L. Lomonaco, P. Haïssinsky, and M. Freiberger could combine some Kleinian groups and some rational maps in the category of holomorphic correspondences, see [BP94, BL20, BL22, BL23, BF05, BH07, Bu00]. In detail, in [BP94, BH07], the authors established the framework of matings for quadratic polynomials with the modular group as well as some Kleinian groups abstractly isomorphic to the modular group. In [BL20, BL22, BL23], they studied a family of quadratic correspondences $\mathcal{F}_a$. Specifically, when a belongs to what they called the *modular Mandelbrot set*, the correspondence $\mathcal{F}_a$ is a mating between the modular group and some quadratic rational map $P_A(z) \coloneqq z + \frac{1}{z} + A$. Moreover, the modular Mandelbrot set is homeomorphic to the classical Mandelbrot set. In [BF05, Bu00], they discussed matings between some polynomial-like maps and Hecke groups.

Motivated by the study of the dynamics of Schwarz reflection maps associated to quadrature domains, S. Lee, M. Yu. Lyubich, N. G. Makarov, and S. Mukherjee investigated matings between such reflection maps and a discrete group abstractly isomorphic to the modular group, see [LLMM21]. Such matings are anti-holomorphic correspondences. Later, M. Yu. Lyubich, J. Mazor, and S. Mukherjee constructed a family of anti-holomorphic correspondences for Schwarz reflection maps associated to quadrature domains and gave two criteria that ensure that such anti-holomorphic correspondences are matings between Schwarz reflections and anti-Hecke groups, see [LMM23].

Apart from matings, holomorphic correspondences, uniting rational maps, Kleinian groups, and matings, also attracted the attention of S. Bullett and C. Penrose. To extend Sullivan's dictionary to include holomorphic correspondences, they studied the general theory of holomorphic correspondences in [BP01]. Specifically, they formulated a formal definition for holomorphic correspondences and generalized the notions of regular sets, limit sets, and Julia sets for holomorphic correspondences.

As part of attempts to study the dynamics of holomorphic correspondences, as well as to investigate the density of hyperbolicity and structural stability in the category of holomorphic correspondences, C. Siqueira and D. Smania studied a specific family of holomorphic correspondences $\mathbf{f}_c(z) = z^{q/p} + c$ in [Siq15, SS17, Siq22, Siq23]. In these papers, C. Siqueira and D. Smania generalized the notion of Julia sets[2] for $\mathbf{f}_c$, discussed the hyperbolicity for such holomorphic correspondences, established some geometric rigidity results for the Julia sets, and gave an upper bound of the Hausdorff dimension of the Julia sets.

However, among all the works cited above about holomorphic and anti-holomorphic correspondences, few focused on ergodic theory for these correspondences.

On the other hand, ergodic theory for general correspondences has also attracted interest recently and there have been some loosely connected but individually valuable

[2]C. Siqueira and D. Smania's version of Julia sets for holomorphic correspondences is different from S. Bullett and C. Penrose's version in [BP01, Section 3.2].

works on this topic. For example, from the topological aspect, Poincaré's recurrence theorem was investigated by J. P. Aubin, H. Frankowska, and A. Lasota in [AFL91]; various concepts of topological entropy, and their upper and lower bounds were established in [KT17]; expansiveness was discussed in [Wi70, PV17]. In addition, from the measure-theoretic aspect, several characterizations of invariant measures are systematically investigated in [MA99]; Perron–Frobenius operators and approximations of invariant measures are studied in [Mil95]. Moreover, in the setting of holomorphic correspondences, T. C. Dinh, L. Kaufmann, and H. Wu [Wu20, DKW20] studied some canonical probability measures under the dynamics of some holomorphic correspondences on Riemann surfaces. V. M. Parra [Par23a, Par23b] studied the equidistributions for the matings discussed in [BP94] and proved that a version of entropy (see [VS22]) of the equidistributions equals to the topological entropy (see [KT17]) of the matings. However, systematic studies on invariant measures are still under development, which motivates us to study the thermodynamic formalism for correspondences.

Thermodynamic formalism for single-valued maps. Thermodynamic formalism, inspired by statistical mechanics and created by Ya. G. Sinai, R. Bowen, D. Ruelle, and others around the early 1970s [Do68, Sin72, Bow75, Ru78], is a mechanism to produce invariant measures with nice properties and prescribed Jacobian functions.

To be more precise, for a continuous (single-valued) map $f\colon X \to X$ on a compact metric space (X, d), and a continuous function $\varphi\colon X \to \mathbb{R}$ (called a potential), we can consider the associated topological pressure $P(f, \varphi)$ as a weighted version of the topological entropy $h_{\mathrm{top}}(f)$. The *Variational Principle* identifies $P(f, \varphi)$ with the supremum of its measure-theoretic counterpart, the measure-theoretic pressure $P_\mu(f, \varphi) \coloneqq h_\mu(f) + \int_X \varphi \,\mathrm{d}\mu$, (where $h_\mu(f)$ is the measure-theoretic entropy), over all invariant Borel probability measures μ [Bow75, Wa76]. A measure that attains the supremum is called an *equilibrium state* for the given map and potential. In particular, when the potential φ is (cohomologous to) a constant function, the equilibrium state is called a *measure of maximal entropy*. The studies on the existence and uniqueness of equilibrium states (or measures of maximal entropy), as well as their ergodic and statistical properties such as supporting sets and equidistributions, have been the main motivation for much research in ergodic theory.

The theory of thermodynamic formalism for f with strong forms of hyperbolicity has been systematically studied. For example, it is well-known that if f is forward expansive, then an equilibrium state exists. Moreover, we have the *Ruelle–Perron–Frobenius Theorems* (see also Propositions 7.10 and 7.15 for more detailed statements), which describes the equilibrium states for more regular potentials, see for example [RT18, Theorem 2.1] and [PU10, Chapter 5].

Briefly speaking, if f is forward expansive and has the specification property and φ is Bowen summable, or if f is open, topologically transitive, and distance-expanding and φ is Hölder continuous, then the equilibrium state exists and is unique, and also has some Gibbs property.

One active direction for investigation in thermodynamic formalism nowadays is to extend the Ruelle–Perron–Frobenius Theorem beyond the scope of uniform hyperbolicity. Our Theorem E on the Lee–Lyubich–Markorov–Mazor–Mukherjee anti-holomorphic correspondence can be seen as such an attempt in the setting of correspondences.

Thermodynamic formalism for correspondences. In the present work, we systematically develop a thermodynamic formalism for correspondences. We will address the following aims:

(i) Formulate definitions of *measure-theoretic entropy of transition probability kernels* and *topological pressure* for correspondences;
(ii) Establish a Variational Principle for some correspondences;
(iii) Establish the existence of equilibrium states and obtain a *Ruelle–Perron–Frobenius Theorem* for correspondences with some strong expansion properties.

Statement of main results. Our main results consist of four parts: a Variational Principle and existence of equilibrium states, a thermodynamic formalism for equilibrium states, a lower bound for the topological pressure, and applications to holomorphic and anti-holomorphic correspondences.

Entropy functions and Variational Principle. We start by defining the measure-theoretic entropy for transition probability kernels and the topological pressure for correspondences.

Roughly speaking, a transition probability kernel $\mathcal{Q}$ on a compact metric space (X, d) assigns each $x \in X$ a Borel probability measure $\mathcal{Q}_x$ on X. We define the measure-theoretic entropy $h_\mu(\mathcal{Q})$ (see Definition 5.22) for a transition probability kernel $\mathcal{Q}$ with respect to a $\mathcal{Q}$-invariant (see Definition 5.10) probability measure μ. The potential function is defined on the set $\mathcal{O}_2(T) := \{(x_1, x_2) \in X^2 : x_2 \in T(x_1)\}$ equipped with the metric d_2 given by $d_2((x_1, x_2), (y_1, y_2)) := \max\{d(x_1, y_1), d(x_2, y_2)\}$ for all $(x_1, x_2), (y_1, y_2) \in \mathcal{O}_2(T)$. The topological pressure $P(T, \phi)$ is defined for a correspondence T and a continuous potential function ϕ. Then, to connect the transition probability kernel and correspondences, we give a relation between them.

We say that a transition probability kernel $\mathcal{Q}$ on X is *supported by* a correspondence T if the measure $\mathcal{Q}_x$ is supported on the closed set $T(x)$ for every $x \in X$.

We conjecture the following Variational Principle to hold.

Conjecture (Variational Principle for correspondences)**.** *Let T be in a suitable class of correspondences on a compact metric space (X, d) and $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$ be a sufficiently regular function, then*

$$P(T, \phi) = \sup_{\mathcal{Q}, \mu} \left\{ h_\mu(\mathcal{Q}) + \int_X \int_{T(x_1)} \phi(x_1, x_2) \, \mathrm{d}\mathcal{Q}_{x_1}(x_2) \, \mathrm{d}\mu(x_1) \right\}, \tag{1.1}$$

where $\mathcal{Q}$ ranges over all transition probability kernels on X supported by T and μ ranges over all $\mathcal{Q}$-invariant probability measures on X.

This conjecture naturally generalizes the classical Variational Principle for single-valued maps. Specifically, the topological pressure $P(T,\phi)$ for the correspondence T generalizes the classical topological pressure for continuous maps, the measure-theoretic entropy $h_\mu(\mathcal{Q})$ of the transition probability kernel $\mathcal{Q}$ with respect to the $\mathcal{Q}$-invariant measure μ generalizes the classical measure-theoretic entropy of a measure-preserving endomorphism, and the integral in (1.1) corresponds to the potential energy in the classical statement of Variational Principle, see Appendix B for details.

Currently, no one has established any version of the Variational Principle for correspondences to our knowledge. In this work, we establish a version of the Variational Principle for correspondences with some expansion properties, see Theorem A. Moreover, we have not found any counterexample to our conjecture.

If a transition probability kernel $\mathcal{Q}$ on X supported by T and a $\mathcal{Q}$-invariant Borel probability measure μ on X satisfy the equality (1.1), then we call the pair $(\mu,\mathcal{Q})$ an *equilibrium state* for the correspondence T and the potential function ϕ. Moreover, if $\phi\equiv 0$, we call $(\mu,\mathcal{Q})$ a *measure of maximal entropy* for T.

Variational Principle and the existence of equilibrium states. We establish the Variational Principle and the existence of equilibrium states for forward expansive correspondences.

A correspondence is *forward expansive*, if, roughly speaking, every pair of distinct forward orbits $(x_1,x_2,\dots),(y_1,y_2,\dots)$ consists of a pair of corresponding entries x_k and y_k with at least a specific distance apart (see Definition 6.1).

The statement of the Variational Principle and the existence of the equilibrium state for forward expansive correspondence is as follows:

Theorem A. Let (X,d) be a compact metric space, T be a forward expansive correspondence on X, and $\phi\colon \mathcal{O}_2(T)\to\mathbb{R}$ be a continuous function. Then the following statements hold:

(i) The Variational Principle holds:
$$P(T,\phi)=\sup_{\mathcal{Q},\mu}\biggl\{h_\mu(\mathcal{Q})+\int_X\int_{T(x_1)}\phi(x_1,x_2)\,\mathrm{d}\mathcal{Q}_{x_1}(x_2)\,\mathrm{d}\mu(x_1)\biggr\}\in\mathbb{R},$$
where $\mathcal{Q}$ ranges over all transition probability kernels on X supported by T and μ ranges over all $\mathcal{Q}$-invariant Borel probability measure on X.

(ii) There exists an equilibrium state $(\mu,\mathcal{Q})$ for the correspondence T and the potential ϕ.

The proof of this theorem occupies from Subsection 6.2 to Subsection 6.5 and is the most technical part of this work.

Thermodynamic formalism and equidistribution. We introduce various properties for correspondences and potential functions and then give two versions of thermodynamic formalism in Section 7.

Let T be a correspondence on a compact metric space X. The metric space $(\mathcal{O}_\omega(T),d_\omega)$ is given in (2.3) and (2.2). If $\mathcal{Q}$ is a transition probability kernel on

X supported by T and μ is a $\mathcal{Q}$-invariant Borel probability measure on X, then we denote $\mu\mathcal{Q}^{\omega}|_T$ a probability measure on $\mathcal{O}_{\omega}(T)$ given in Remark 6.14.

In the first version, we assume that T has the specification property (see Definition 7.1) and ϕ is Bowen summable (see Definition 7.3), then consequently the Variational Principle holds, the equilibrium state exists and is unique in an appropriate sense, and the unique equilibrium state can be obtained by investigating the (classical) Ruelle operator $\mathcal{L}_{\widetilde{\phi}}$ and $\mathcal{L}^*_{\widetilde{\phi}}$ (see (7.5) and (7.6)).

Theorem B. Let (X,d) be a compact metric space, T be a forward expansive correspondence with the specification property, and $\phi\colon \mathcal{O}_2(T)\to\mathbb{R}$ be a Bowen summable continuous function. Then the Variational Principle (1.1) holds and there exists an equilibrium state $(\mu_\phi,\mathcal{Q})$ for the correspondence T and the potential ϕ, i.e., there exist a transition probability kernel $\mathcal{Q}$ on X supported by T and a $\mathcal{Q}$-invariant Borel probability measure μ_ϕ on X such that the following equality holds:

$$P(T,\phi)=h_{\mu_\phi}(\mathcal{Q})+\int_X\int_{T(x_1)}\phi(x_1,x_2)\,\mathrm{d}\mathcal{Q}_{x_1}(x_2)\,\mathrm{d}\mu_\phi(x_1). \tag{1.2}$$

Moreover, the equilibrium state $(\mu_\phi,\mathcal{Q})$ is unique in the sense that the measure μ_ϕ is unique and that if there are two equilibrium states $(\mu_\phi,\mathcal{Q})$ and $(\mu_\phi,\mathcal{Q}')$, then for μ_ϕ-almost every $x\in X$ and all $A\in\mathscr{B}(X)$, the equality $\mathcal{Q}_x(A)=\mathcal{Q}'_x(A)$ holds.

Furthermore, the equilibrium state $(\mu_\phi,\mathcal{Q})$ can be obtained in the following way:

(i) There is a Borel probability measure m_ϕ on X and a transition probability kernel $\mathcal{Q}$ on X supported by T such that $m_\phi\mathcal{Q}^{\omega}|_T$ is an eigenvector of $\mathcal{L}^*_{\widetilde{\phi}}$.

(ii) There is a Borel measurable function $u_\phi\in L^1(m_\phi)$ such that $\mathcal{L}_{\widetilde{\phi}}(\widetilde{u}_\phi)=\lambda\widetilde{u}_\phi$, where $\lambda=\exp\bigl(P\bigl(\sigma,\widetilde{\phi}\bigr)\bigr)=\exp(P(T,\phi))$ and $\widetilde{u}_\phi\colon\mathcal{O}_\omega(T)\to\mathbb{R}$ is the bounded Borel measurable function induced by u_ϕ in the following way:

$$\widetilde{u}_\phi(x_1,x_2,\dots)\coloneqq u_\phi(x_1).$$

(iii) Set $\mu_\phi\coloneqq u_\phi m_\phi$, then $(\mu_\phi,\mathcal{Q})$ is the equilibrium state for the correspondence T and the potential ϕ.

In the second version, we assume that T is distance-expanding (see Definition 7.5), open (see Definition 7.11), and strongly transitive (see Definition 7.12) and ϕ is Hölder continuous, then the Variational Principle holds, the equilibrium state exists and is unique in an appropriate sense, and the unique equilibrium state can be obtained from the Ruelle operator and has the equidistribution property.

Theorem C. Let T be a open, strongly transitive, distance-expanding correspondence on X and $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ be a Hölder continuous function. Then the Variational Principle (1.1) holds and there exists an equilibrium state $(\mu_\phi,\mathcal{Q})$ for the correspondence T and the potential ϕ. Moreover, the equilibrium state $(\mu_\phi,\mathcal{Q})$ is unique in the sense that the measure μ_ϕ is unique and if there are two equilibrium states $(\mu_\phi,\mathcal{Q})$ and $(\mu_\phi,\mathcal{Q}')$, then for μ_ϕ-almost every $x\in X$ and all $A\in\mathscr{B}(X)$, the equality $\mathcal{Q}_x(A)=\mathcal{Q}'_x(A)$ holds.

Furthermore, the equilibrium state $(\mu_\phi,\mathcal{Q})$ can be obtained in the following way:

(i) There is a Borel probability measure m_ϕ on X and a transition probability kernel $\mathcal{Q}$ on X supported by T such that $m_\phi \mathcal{Q}^\omega|_T$ is an eigenvector of $\mathcal{L}^*_{\widetilde{\phi}}$.

(ii) There is a Borel measurable function $u_\phi \in L^1(m_\phi)$ such that $\mathcal{L}_{\widetilde{\phi}}(\widetilde{u}_\phi) = \lambda \widetilde{u}_\phi$, where $\lambda = \exp\bigl(P\bigl(\sigma, \widetilde{\phi}\bigr)\bigr) = \exp(P(T,\phi))$ and $\widetilde{u}_\phi \colon \mathcal{O}_\omega(T) \to \mathbb{R}$ is the bounded Borel measurable function induced by u_ϕ in the following way:

$$\widetilde{u}_\phi(x_1, x_2, \dots) \coloneqq u_\phi(x_1).$$

If, moreover, T is continuous (Definition 4.3), then u_ϕ is continuous.

(iii) Set $\mu_\phi \coloneqq u_\phi m_\phi$, then $(\mu_\phi, \mathcal{Q})$ is the equilibrium state for the correspondence T and the potential ϕ.

In addition, the backward orbits under T are equidistributed with respect to the measure μ_ϕ. More precisely, if we denote

$$\mathcal{O}_{-n}(x) \coloneqq \bigl\{(y_0, y_1, \dots, y_n) \in X^{n+1} : y_n = x,\, y_k \in T(y_{k-1}) \text{ for each } k \in \{1, \dots, n\}\bigr\},$$

$$Z_n(x) \coloneqq \sum_{(y_0,\dots,y_n)\in\mathcal{O}_{-n}(x)} \exp\biggl(\sum_{i=0}^{n-1} \phi(y_i, y_{i+1})\biggr),$$

then the following statements hold:

(a) For each $x \in X$, the following sequence of Borel probability measures on X

$$\frac{1}{Z_n(x)} \sum_{(y_0,\dots,y_n)\in\mathcal{O}_{-n}(x)} \frac{\sum_{j=0}^{n} \delta_{y_j}}{n+1} \exp\biggl(\sum_{i=0}^{n-1} \phi(y_i, y_{i+1})\biggr),\ n \in \mathbb{N},$$

converges to μ_ϕ in the weak* topology as n tends to $+\infty$.

(b) If, moreover, T is topologically exact (Definition 7.13), then for each $x \in X$, the following sequence of Borel probability measures on X

$$\frac{1}{Z_n(x)} \sum_{(y_0,\dots,y_n)\in\mathcal{O}_{-n}(x)} \delta_{y_0} \exp\biggl(\sum_{i=0}^{n-1} \phi(y_i, y_{i+1})\biggr),\ n \in \mathbb{N},$$

converges to m_ϕ in the weak* topology as n tends to $+\infty$.

For general correspondences and continuous potential function, we establish the following result.

Theorem D. Let T be a correspondence on a compact metric space (X,d) and $\phi \colon \mathcal{O}_2(T) \to \mathbb{R}$ be a continuous function. Then the following statements hold:

(i) There exists a transition probability kernel $\mathcal{Q}$ on X supported by T and a $\mathcal{Q}$-invariant Borel probability measure μ on X.

(ii) We have

$$P(T,\phi) \geqslant \sup_{\mathcal{Q},\mu} \biggl\{ h_\mu(\mathcal{Q}) + \int_X \int_{T(x_1)} \phi(x_1, x_2)\, \mathrm{d}\mathcal{Q}_{x_1}(x_2)\, \mathrm{d}\mu(x_1) \biggr\}, \tag{1.3}$$

where $\mathcal{Q}$ ranges over all transition probability kernels on X supported by T and μ ranges over all $\mathcal{Q}$-invariant Borel probability measures on X.

Applications to holomorphic or anti-holomorphic correspondences. Our main motivation for the investigation in this work comes from the following two classes of correspondences in the complex dynamics, especially the first one. For more details, see Section 3.

Lee–Lyubich–Markorov–Mazor–Mukherjee anti-holomorphic correspondence.

Let $d \in \mathbb{N}$ and $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a rational map of degree $d+1$ that is univalent on the open unit disk $\mathbb{D}$. Set $\eta(z) \coloneqq 1/\overline{z}$, the reflection map on the unit circle.

The Lee–Lyubich–Markorov–Mazor–Mukherjee anti-holomorphic correspondence $\mathfrak{C}^*$ is defined as follows:

$$\mathfrak{C}^*(z) \coloneqq \left\{ w \in \widehat{\mathbb{C}} : \frac{f(w) - f(\eta(z))}{w - \eta(z)} = 0 \right\} \tag{1.4}$$

for all $z \in \widehat{\mathbb{C}}$. See [LMM23, Section 2] for more details. We note that is not $\mathfrak{C}^*$ forward expansive.

The correspondence $\mathfrak{C}^*$ can be divided into two independent parts in the sense of Proposition 3.1. The dynamics of $\mathfrak{C}^*$ is equivalent to the action of an abstract anti-Hecke group on one part in some cases (see [LMM23, Propositions 2.15 and 2.19]), and is related to an anti-holomorphic map on another part, see Proposition 3.6 for details. Note that $\mathfrak{C}^*$ is not forward expansive.

We establish a version of the Variational Principle for the correspondence $\mathfrak{C}^*$ as follows.

Theorem E. Let $\mathfrak{C}^*$ be the correspondence given above and $\phi\colon \mathcal{O}_2(\mathfrak{C}^*) \to \mathbb{R}$ be a continuous function. Then we have

$$P(\mathfrak{C}^*, \phi) = \sup_{\mathcal{Q},\mu} \left\{ h_\mu(\mathcal{Q}) + \int_{\widehat{\mathbb{C}}} \int_{\mathfrak{C}^*(x_1)} \phi(x_1, x_2) \,\mathrm{d}\mathcal{Q}_{x_1}(x_2) \,\mathrm{d}\mu(x_1) \right\}, \tag{1.5}$$

where $\mathcal{Q}$ ranges over all transition probability kernels on $\widehat{\mathbb{C}}$ supported by $\mathfrak{C}^*$ and μ ranges over all $\mathcal{Q}$-invariant probability measures on $\widehat{\mathbb{C}}$.

A family of hyperbolic holomorphic correspondences.

Next, we consider a family of holomorphic correspondences studied in [Siq15, SS17, Siq22, Siq23]. Fix p, $q \in \mathbb{N}$ satisfying $p < q$. Let $c \in \mathbb{C}$.

Denote by $\mathbf{f}_c(z) = z^{q/p} + c$ the correspondence[3] on $\widehat{\mathbb{C}}$ given by

$$\mathbf{f}_c(z) \coloneqq \left\{ w \in \widehat{\mathbb{C}} : (w - c)^p = z^q \right\}$$

for all $z \in \widehat{\mathbb{C}}$.

A version of *Julia set* $J(\mathbf{f}_c)$ is defined as the closure of the union of all repelling periodic orbits of $\mathbf{f}_c$, see for example, [Siq15, Definition 6.31] or [SS17, Section 2.1]. Denote by $\mathbf{f}_c|_J$ a map given by $\mathbf{f}_c|_J(z) \coloneqq J(\mathbf{f}_c) \cap \mathbf{f}_c(z)$ for all $z \in J(\mathbf{f}_c)$.

Set $P_c \coloneqq \overline{\bigcup_{n\in\mathbb{N}} \mathbf{f}_c^n(0)}$ and

$$M_{q/p,0} \coloneqq \{ c \in \mathbb{C} : \exists (x_1, x_2, \dots) \in \mathcal{O}_\omega(\mathbf{f}_c) \text{ such that } x_1 = 0 \text{ and } \{x_n\}_{n\in\mathbb{N}} \text{ is bounded} \}.$$

[3] The superscript q/p in $z^{q/p} + c$ is merely a notation, not a fraction. The two correspondences $z^{2/1} + c$ and $z^{4/2} + c$ are different by this definition.

A number $c \in \mathbb{C}$ is called a *simple center* if $c \neq 0$ and there is only one bounded orbit $(x_1, x_2, \dots) \in \mathcal{O}_\omega(\mathbf{f}_c)$ with $x_1 = 0$ and such a bounded orbit is a cycle, see for example, [Siq22, Section 2.2].

Theorem F. There is an open set $H_{q/p}$ containing both $\mathbb{C} \setminus M_{q/p,0}$ and every simple center such that for every $c \in H_{q/p}$, the following statements holds:

(i) The set $\mathbb{C} \setminus P_c$ is a hyperbolic Riemann surface.
(ii) The statements (i), (ii), (iii), (a), and (b) in Theorem C hold for the correspondence $\mathbf{f}_c|_J$ on the compact metric space $(J(\mathbf{f}_c), d_c)$, where d_c refers to the hyperbolic metric on $\mathbb{C} \setminus P_c$.

Remark. C. Siqueira pointed out in [Siq15, Section 1.2] that one of the motivations to study the correspondences $z^{q/p} + c$ is the density of hyperbolicity. Specifically, Fatou conjectured in 1920 that hyperbolic maps are dense within the space of rational maps with fixed degrees. C. Siqueira introduced a version of hyperbolicity for this family of holomorphic correspondences in [Siq22, Definition 5.6] and proved that there is an open set $H_{q/p}$ containing both $\mathbb{C} \setminus M_{q/p,0}$ and every simple center with the property that $\mathbf{f}_c$ is hyperbolic for all $c \in H_{q/p}$, see [Siq22, Corollary 5.7.1]. But now we do not know further relations between the two different $H_{q/p}$ on which Theorem F holds and on which $\mathbf{f}_c$ is hyperbolic, respectively.

Note that 0 is not a simple center, so Theorem F does not work when c is close to 0. For c in a neighborhood of 0, we have the following result.

Theorem G. There is an open neighborhood $U_{q/p}$ of 0 with the property that for every $c \in U_{q/p}$, the statements (i), (ii), (iii), (a), and (b) in Theorem C hold for the correspondence $\mathbf{f}_c|_J$ on the compact space $J(\mathbf{f}_c)$ equipped with the Euclidian metric on $\mathbb{C}$.

Strategy of this work. In this subsection, we discuss our innovations, the difficulties, and the methods to overcome these difficulties.

Correspondences assign each point a set, which means they are multi-valued, while the continuous maps are single-valued. To emphasize this difference, we use *single-valued continuous maps* to refer to continuous maps. Despite this difference, J. Kelly and T. Tennant managed to define the topological entropy of correspondences in a natural way using (n, ϵ)-separated sets and (n, ϵ)-spanning sets [KT17, Definition 2.5]. We extend this idea to define the topological pressure for correspondences (see Section 4.2).

In contrast, measure-theoretic entropy is harder to define for correspondences, because, given a subset of a space, there are no canonical distributions on it. We overcome this difficulty by assigning a distribution on the image of each point under a correspondence, i.e., we consider a transition probability kernel. Roughly speaking, a *transition probability kernel* on a space X assigns each point a probability measure on X. We introduce the measure-theoretic entropy for a transition probability kernel using the entropy of partitions (See Section 5.3).

To establish our Variational Principle, we need a relation between correspondences and transition probability kernels. The most natural relation is *support*. Recall that a correspondence T on X assigns each point $x \in X$ a closed subset $T(x)$ of X, and a transition probability kernel on X assigns each point a probability measure on X. Recall that a transition probability kernel is *supported by* a correspondence T if, for each point $x \in X$, the probability measure is supported on the closed subset $T(x)$. We conjecture a version of the Variational Principle, see (1.1).

Now we sketch the main results and their proofs.

We first establish the characterizations of both the topological pressure of correspondences and the measure-theoretic entropy for transition probability kernels in terms of the shift map on the orbit space. We briefly explain the dynamics of such a shift map now.

Let T be a correspondence on a compact metric space X, $\mathcal{Q}$ be a transition probability kernel on X, and μ be a probability measure on X invariant under $\mathcal{Q}$, i.e., the pushforward of μ under $\mathcal{Q}$ is still μ (see Definition 5.10 for details). Let σ be the shift map on $X^\omega \coloneqq \{(x_1, x_2, \dots) : x_k \in X \text{ for all } k \in \mathbb{N}\}$, i.e., $\sigma(x_1, x_2, \dots) = (x_2, x_3, \dots)$. It turns out that $\mathcal{O}_\omega(T) \coloneqq \{(x_1, x_2, \dots) \in X^\omega : x_{k+1} \in T(x_k)\}$ is an invariant set of σ. Moreover, for the potential function $\phi \colon \mathcal{O}_2(T) \to \mathbb{R}$, we also lift it to the orbit space $\mathcal{O}_\omega(T)$: denote by $\widetilde{\phi}$ the function on $\mathcal{O}_\omega(T)$ given by $\widetilde{\phi}(x_1, x_2, \dots) \coloneqq \phi(x_1)$. Now we state the following characterizations:

First, the (classical) topological pressure of σ restricted on $\mathcal{O}_\omega(T)$ with respect to the potential $\widetilde{\phi}$ is equal to the topological pressure of T with respect to the potential ϕ (see Theorem 4.9).

Second, we consider the Markov process with the initial distribution μ and the transition probability kernel $\mathcal{Q}$. Denote by $\mu\mathcal{Q}^\omega$ the distribution of forward infinite orbits of the Markov process. It is a probability measure on X^ω. The assumption that μ is invariant under $\mathcal{Q}$ implies that $\mu\mathcal{Q}^\omega$ is invariant under σ (see Subsection 5.4). We prove that the (classical) measure-theoretic entropy of σ with respect to the invariant measure $\mu\mathcal{Q}^\omega$ is equal to the measure-theoretic entropy of $\mathcal{Q}$ with respect to the invariant measure μ (see Theorem 5.24).

Thereby, the following conjecture about σ is equivalent to our (conjectured) Variational Principle:

$$P\bigl(\sigma|_{\mathcal{O}_\omega(T)}, \widetilde{\phi}\bigr) = \sup_\nu \biggl\{ h_\nu(\sigma) + \int \widetilde{\phi} \, \mathrm{d}\nu \biggr\}, \tag{1.6}$$

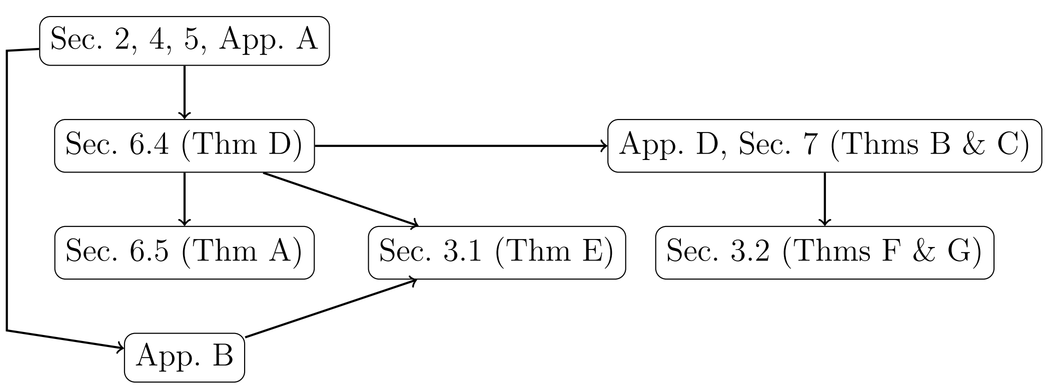
where ν ranges over all Borel probability measures on $\mathcal{O}_\omega(T)$ induced by Markov processes and invariant under σ (see Subsection 7.2 for details).

Note that if the supremum in (1.6) is taken by letting ν range over all Borel probability measures on $\mathcal{O}_\omega(T)$ invariant under σ, then (1.6) holds, ensured by the (classical) Variational Principle. From this perspective, we establish Theorem D. But some σ-invariant measures on $\mathcal{O}_\omega(T)$ are not induced by Markov processes, which is the difficulty of establishing the Variational Principle for correspondences. For the Variational Principle and the corresponding thermodynamic formalism, we use two kinds of methods in Sections 6 and 7, respectively.

In Section 6, we introduce the forward expansiveness for correspondences (see Definition 6.1) and establish the Variational Principle for forward expansive correspondences (Theorem A). We overcome the difficulty that a probability measure invariant under the shift map may not be induced by a Markov process in this section. Specifically, for an arbitrary Borel probability measure ν on the orbit space $\mathcal{O}_\omega(T)$ which is invariant under the shift map, the projection of ν onto the first two coordinates can induce a measure μ and a conditional transition probability kernel $\mathcal{Q}$. It turns out that $\mathcal{Q}$ is supported by a forward expansive correspondence T and that μ is $\mathcal{Q}$-invariant. Moreover, the measure-theoretic entropy of $\mu\mathcal{Q}^\omega$ is greater than or equal to the measure-theoretic entropy of ν. Therefore, to prove (1.6), it is enough to consider the case where ν is induced by a Markov process, and the measure-theoretic entropy of such ν corresponds to the measure-theoretic entropy of transition probability kernels.

In Section 7, we first introduce various properties for correspondences or potential functions, including specification property (Definition 7.1), Bowen summability (Definition 7.3), distance-expanding property (Definition 7.5), openness (Definition 7.11), and strong transitivity (Definition 7.12), and recall the topologically exact property (Definition 7.13) for correspondences. We prove that these properties imply some corresponding properties of the shift map on the orbit space. See Propositions D.3, D.4, D.6, D.7, and D.8 for precise statements. Then we give two versions of thermodynamic formalism for forward expansive correspondence T on a compact metric space X with a continuous potential function $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$ satisfying some of the properties above. The key tool for our approach is the Ruelle–Perron–Frobenius Theorem for the shift map. Since the Ruelle–Perron–Frobenius Theorem describes the equilibrium state for the shift map explicitly, we can verify that the equilibrium state is indeed induced by a Markov process. In this way, we get that equilibrium states for correspondences, and their various properties (uniqueness and equidistribution) come from the corresponding properties of the equilibrium states for shift maps.

Structures of this work. Let us highlight our results and structure of this work in more detail.



In Section 2, we fix some notations that will be used throughout this work.

In Section 4, we introduce the topological pressure $P(T,\phi)$ of a correspondence T with respect to a continuous potential function $\phi\colon \mathcal{O}_2(T)\to\mathbb{R}$ through the (n,ϵ)-separated set and (n,ϵ)-spanning set (Definition 4.6). Then we show that our topological pressure of a correspondence T with respect to a continuous potential function ϕ is equal to $P\bigl(\sigma,\widetilde{\phi}\bigr)$ (see Theorem 4.9), the topological pressure of the shift map σ on the orbit space $\mathcal{O}_\omega(T)$ with respect to the potential function $\widetilde{\phi}$ (given in (2.4)) induced by ϕ.

In Section 5, we discuss transition probability kernels and introduce the measure-theoretic entropy $h_\mu(\mathcal{Q})$ (Definition 5.22) for a transition probability kernel $\mathcal{Q}$ with respect to a $\mathcal{Q}$-invariant (Definition 5.10) probability measure μ through the entropy of partitions. Then we show that the measure-theoretic entropy of a transition probability kernel is equal to the measure-theoretic entropy of the shift map.

In Section 6, we introduce the forward expansiveness for correspondences (see Definition 6.1) and establish the Variational Principle for forward expansive correspondence (Theorem A). We also investigate general correspondences and establish Theorem D by utilizing what we called the "support" relation between transition probability kernels and correspondences (see Definition 5.3).

In Section 7, we first introduce various properties for correspondences or potential functions, including specification property (Definition 7.1), Bowen summability (Definition 7.3), distance-expanding property (Definition 7.5), openness (Definition 7.11), and strong transitivity (Definition 7.12), and recall the topologically exact property for correspondences (Definition 7.13). Then we give two versions of thermodynamic formalism for forward expansive correspondence T on a compact metric space X with a continuous potential function $\phi\colon \mathcal{O}_2(T)\to\mathbb{R}$ with some of the properties above.

Apart from that, in Section 3, we apply our theory to two examples: the Lee–Lyubich–Markorov–Mazor–Mukherjee anti-holomorphic correspondence and a family of hyperbolic holomorphic correspondence of the form $z^{q/p}+c$.

Acknowledgments. The authors want to thank Wenyuan Yang for interesting discussions on random walks. X. Li wants to thank Xianghui Shi for the useful comments.

2. Notation

We follow the convention $\mathbb{N} \coloneqq \{1, 2, 3, \dots\}$, $\mathbb{N}_0 \coloneqq \mathbb{N}\cup\{0\}$, and $\widehat{\mathbb{N}} \coloneqq \mathbb{N}\cup\{\omega\}$. Here ω is the least infinite ordinal. For each $n\in\mathbb{N}_0$, write $[n] \coloneqq \{0, 1, \dots, n\}$ and $(n] \coloneqq [n]\setminus\{0\}$.

Let $\mathbb{C}$ be the set of all complex numbers, $\widehat{\mathbb{C}} \coloneqq \mathbb{C}\cup\{\infty\}$, and $\mathbb{D} \coloneqq \{z\in\mathbb{C} : \lvert z\rvert<1\}$ be the unit disk on the complex plane.

For a set $\mathcal{D}$ consisting of other sets, denote by $\bigcup\mathcal{D}$ the union of all elements in $\mathcal{D}$.

When we use the notation $A\cap B\times C$ or $B\times C\cap A$ for sets A, B, and C, it should be interpreted as first performing the multiplication operation between sets B and C and subsequently finding the intersection of the result with A.

Let X be a set and $n \in \mathbb{N}$. Define $\iota_n \colon X^n \to X^n$ by

$$\iota_n(x_1, x_2, \dots, x_n) \coloneqq (x_n, x_{n-1}, \dots, x_1) \quad \text{for all } (x_1, \dots, x_n) \in X^n.$$

For a function $\phi \colon X \to \mathbb{R}$, write $\|\phi\|_\infty \coloneqq \sup\{|\phi(x)| : x \in X\}$. Let $\mathscr{M}(X)$ be a σ-algebra on X. Denote by $B(X, \mathbb{R})$ the set of real-valued bounded measurable functions on X and by $\mathcal{P}(X)$ the set of probability measures on X.

Let Y be a compact metric space. Denote by $\mathscr{B}(Y)$ the Borel σ-algebra on Y. Denote by $\mathcal{F}(Y)$ the set of all non-empty closed subsets of Y, and by $C(Y, \mathbb{R})$ the set of real-valued continuous functions on Y.

Let X be a compact metric space with the metric d and $T \colon X \to \mathcal{F}(X)$ be a map. For each subset $A \subseteq X$, set $T(A) \coloneqq \bigcup_{x \in A} T(x) \subseteq X$. For $n \in \mathbb{N}$, define $T^n(A) \subseteq X$ inductively on n with $T^1(A) \coloneqq T(A)$ and $T^{n+1}(A) \coloneqq T(T^n(A))$, for all $n \in \mathbb{N}$. Moreover, write $T^{-1}(A) \coloneqq \{x \in X : A \cap T(x) \neq \emptyset\} \subseteq X$. For $n \in \mathbb{N}$, define $T^{-n}(A) \subseteq X$ inductively on n with $T^{-(n+1)}(A) \coloneqq T^{-1}(T^{-n}(A))$, for all $n \in \mathbb{N}$. For each $n \in \mathbb{Z} \setminus \{0\}$ and each $x \in X$, write $T^n(x) \coloneqq T^n(\{x\})$.

For a subset $A \subseteq X$ and each $x \in X$, define $T|_A(x) \coloneqq T(x) \cap A$.

For each $n \in \mathbb{N}$, equip $X^n \coloneqq \{(x_1, \dots, x_n) : x_k \in X \text{ for all } k \in (n]\}$ with the metric d_n given by

$$d_n((x_1, \dots, x_n), (y_1, \dots, y_n)) \coloneqq \max\{d(x_i, y_i) : i \in (n]\} \tag{2.1}$$

for all $(x_1, \dots, x_n), (y_1, \dots, y_n) \in X^n$. Clearly $\iota_n \colon X^n \to X^n$ is an isometry. Similarly, equip $X^\omega \coloneqq \{(x_1, x_2, \dots) : x_k \in X \text{ for all } k \in \mathbb{N}\}$ with the metric d_ω given by

$$d_\omega((x_1, x_2, \dots), (y_1, y_2, \dots)) \coloneqq \sum_{k=1}^{+\infty} \frac{d(x_k, y_k)}{2^k(1 + d(x_k, y_k))} \tag{2.2}$$

for all $(x_1, x_2, \dots), (y_1, y_2, \dots) \in X^\omega$. With the metrics d_n for $n \in \widehat{\mathbb{N}}$, the topologies of X^n induced by these metrics are the product topologies.

For each $n \in \mathbb{N}$, write

$$\mathcal{O}_n(T) \coloneqq \{(x_1, \dots, x_n) \in X^n : x_{k+1} \in T(x_k) \text{ for each } k \in (n-1]\} \subseteq X^n.$$

Then the *orbit space* $\mathcal{O}_\omega(T)$ induced by T is given by

$$\mathcal{O}_\omega(T) \coloneqq \{(x_1, x_2, \dots) \in X^\omega : x_{k+1} \in T(x_k) \text{ for each } k \in \mathbb{N}\} \subseteq X^\omega. \tag{2.3}$$

For each $n \in \widehat{\mathbb{N}}$, We call an element in $\mathcal{O}_n(T)$ an *orbit*. A sequence of orbits $\bigl(x_1^{(j)}, x_2^{(j)}, \dots\bigr) \in \mathcal{O}_\omega(T)$, $j \in \mathbb{N}$, converges to an orbit $(x_1, x_2, \dots)$ if and only if $x_k^{(j)}$ converges to x_k as $j \to +\infty$ for each $k \in \mathbb{N}$.

Let $\varphi \colon X \to \mathbb{R}$ and $\phi \colon \mathcal{O}_2(T) \to \mathbb{R}$ be two continuous functions. Denote by $\widetilde{\varphi}$, $\widetilde{\phi} \colon \mathcal{O}_\omega(T) \to \mathbb{R}$ and $\widehat{\varphi} \colon \mathcal{O}_2(T) \to \mathbb{R}$ be functions given by

$$\widetilde{\varphi}(x_1, x_2, \dots) \coloneqq \varphi(x_1), \quad \widetilde{\phi}(x_1, x_2, \dots) \coloneqq \phi(x_1, x_2), \quad \text{and } \widehat{\varphi}(x_1, x_2) \coloneqq \varphi(x_1). \tag{2.4}$$

Clearly, they are all continuous functions.

Denote by $\widetilde{\pi}_1$, $\widetilde{\pi}_2 \colon \bigcup_{n \in \widehat{\mathbb{N}} \setminus \{1\}} X^n \to X$, and $\widetilde{\pi}_{12} \colon \bigcup_{n \in \widehat{\mathbb{N}} \setminus \{1\}} X^n \to X^2$ the projection maps given by

$$\widetilde{\pi}_{12} \colon (x_n)_n \mapsto (x_1, x_2), \quad \widetilde{\pi}_1 \colon (x_n)_n \mapsto x_1, \quad \widetilde{\pi}_2 \colon (x_n)_n \mapsto x_2. \tag{2.5}$$

If μ is a Borel probability measure on a subset A of X^n for some $n \in \widehat{\mathbb{N}} \setminus \{1\}$, then $\mu \circ \widetilde{\pi}_{12}^{-1}$ refers to a Borel probability measure on X^2 given by $\mu \circ \widetilde{\pi}_{12}^{-1}(B) := \mu(A \cap \widetilde{\pi}_{12}^{-1}(B))$ for all $B \in \mathscr{B}(X^2)$, and $\mu \circ \widetilde{\pi}_i^{-1}$ refers to a Borel probability measure on X given by $\mu \circ \widetilde{\pi}_i^{-1}(B) := \mu(A \cap \widetilde{\pi}_i^{-1}(B))$ for all $B \in \mathscr{B}(X)$, where $i = 1$ or $i = 2$.

3. Holomorphic and anti-holomorphic correspondences

In this section, we discuss our main motivation for the investigation in this work: two examples in complex dynamics, especially the first one. Since we will apply our theory developed throughout this work to them, the reader can skip this section in the first read.

3.1. **Lee–Lyubich–Markorov–Mazor–Mukherjee correspondences.** In this subsection, we aim to prove Theorem E.

Set $\mathbb{D}^* := \{z \in \widehat{\mathbb{C}} : |z| > 1\}$. Recall $d \in \mathbb{N}$ and that $f \colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a rational map of degree $d+1$ that is univalent on the open unit disk $\mathbb{D}$. Recall $\eta(z) = 1/\overline{z}$ and the correspondence $\mathfrak{C}^*(z) = \bigl\{w \in \widehat{\mathbb{C}} : \frac{f(w)-f(\eta(z))}{w-\eta(z)} = 0\bigr\}$ for all $z \in \widehat{\mathbb{C}}$.

Let $\Omega := f(\mathbb{D})$. An anti-holomorphic map $\tau \colon \Omega \to \widehat{\mathbb{C}}$ is given by

$$\tau := f \circ \eta \circ (f|_{\mathbb{D}})^{-1}. \tag{3.1}$$

Let $T(\tau) := \widehat{\mathbb{C}} \setminus \Omega$ and $S(\tau)$ be the singular set (consisting of all cusps and double points) of $\partial T(\tau) = \partial\Omega$. Set $T^0(\tau) := T(\tau) \setminus S(\tau)$ and $T^\infty(\tau) := \bigcup_{n \in \mathbb{N}_0} \tau^{-n}(T^0(\tau))$. Write $K(\tau) := \widehat{\mathbb{C}} \setminus T^\infty(\tau)$, which is called the *non-escaping set of* τ. It turns out that $K(\tau)$ is a closed subset of $\widehat{\mathbb{C}}$ by [LMM23, Proposition 2.2]. Write

$$\widetilde{K(\tau)} := f^{-1}(K(\tau)) \text{ and } \widetilde{T^\infty(\tau)} := f^{-1}(T^\infty(\tau)) = \widehat{\mathbb{C}} \setminus \widetilde{K(\tau)}. \tag{3.2}$$

The subset $\widetilde{K(\tau)}$ is closed in $\widehat{\mathbb{C}}$, and the subset $\widetilde{T^\infty(\tau)}$ is open.

About the dynamics of the correspondence, the following result is from [LMM23, Proposition 2.4].

Proposition 3.1. *For all $z, w \in \widehat{\mathbb{C}}$, if $w \in \mathfrak{C}^*(z)$, then $z \in \widetilde{K(\tau)}$ if and only if $w \in \widetilde{K(\tau)}$, and $z \in \widetilde{T^\infty(\tau)}$ if and only if $w \in \widetilde{T^\infty(\tau)}$.*

Set $V_0 := f^{-1}\bigl(\widehat{\mathbb{C}} \setminus \overline{\Omega}\bigr)$, $U_0 := f^{-1}(\partial\Omega \setminus S(\tau)) \cap \partial\mathbb{D}$, $V_n := f^{-1}\bigl(\tau^{-n}\bigl(\widehat{\mathbb{C}} \setminus \overline{\Omega}\bigr)\bigr) \cap \mathbb{D}^*$, $U_n := f^{-1}(\tau^{-(n-1)}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D}^*$, $V_{-n} := f^{-1}\bigl(\tau^{-n}\bigl(\widehat{\mathbb{C}} \setminus \overline{\Omega}\bigr)\bigr) \cap \mathbb{D}$, and $U_{-n} := f^{-1}(\tau^{-n}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D}$ for all $n \in \mathbb{N}$. We have the following lemma.

Lemma 3.2. *The collection $\{V_n\}_{n \in \mathbb{Z}} \cup \{U_n\}_{n \in \mathbb{Z}}$ is a partition of $\widetilde{T^\infty(\tau)}$.*

Proof. Since $\tau^{-1}(T^0(\tau)) \subseteq \Omega$ and $T^0(\tau) \subseteq \widehat{\mathbb{C}} \setminus \Omega$, we can see that $\{\tau^{-n}(T^0(\tau))\}_{n \in \mathbb{N}_0}$ a partition of $T^\infty(\tau)$. Note that $T^0(\tau) = \bigl(\widehat{\mathbb{C}} \setminus \Omega\bigr) \setminus S(\tau)$ is the disjoint union of $\widehat{\mathbb{C}} \setminus \overline{\Omega}$ and $\partial\Omega \setminus S(\tau)$, so $\bigl\{f^{-1}\bigl(\tau^{-n}\bigl(\widehat{\mathbb{C}} \setminus \overline{\Omega}\bigr)\bigr)\bigr\}_{n \in \mathbb{N}_0} \cup \{f^{-1}(\tau^{-n}(\partial\Omega \setminus S(\tau)))\}_{n \in \mathbb{N}_0}$ is a partition of $\widetilde{T^\infty(\tau)} = f^{-1}(T^\infty(\tau))$.

Recall that Ω is the univalent image of $\mathbb{D}$ under f. For each $n \in \mathbb{N}$, since $\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big) \subseteq \Omega$ and $(\tau^{-n}(\partial\Omega \setminus S(\tau))) \subseteq \Omega$, we have $f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \subseteq \widehat{\mathbb{C}} \setminus \partial\mathbb{D} = \mathbb{D} \cup \mathbb{D}^*$ and $f^{-1}(\tau^{-n}(\partial\Omega \setminus S(\tau))) \subseteq \mathbb{D} \cup \mathbb{D}^*$. Moreover, since $f(\mathbb{D}) = \Omega$, whose intersection with $\partial\Omega \setminus S(\tau)$ is empty, we have $f^{-1}(\partial\Omega \setminus S(\tau)) \subseteq \partial\mathbb{D} \cup \mathbb{D}^*$. From the arguments above, we conclude that the collection $\{V_n\}_{n\in\mathbb{Z}} \cup \{U_n\}_{n\in\mathbb{Z}}$ is a partition of $\widetilde{T^\infty}(\tau)$. $\square$

By (3.1), we have $\tau \circ f(z) = f \circ \eta \circ (f|_{\mathbb{D}})^{-1} \circ f(z) = f \circ \eta(z)$ for all $z \in \mathbb{D}$, and $\tau \circ f \circ \eta(z) = f(z)$ for all $z \in \mathbb{D}^*$. This implies for each $A \subset \widehat{\mathbb{C}}$,

$$f^{-1}(\tau^{-1}(A)) \cap \mathbb{D} = \eta(f^{-1}(A)) \cap \mathbb{D} \text{ and } \eta(f^{-1}(\tau^{-1}(A))) \cap \mathbb{D}^* = f^{-1}(A) \cap \mathbb{D}^*. \tag{3.3}$$

The following proposition is to describe the dynamics of $(\mathfrak{C}^*)^{-1}$ on $\{V_n\}_{n\in\mathbb{Z}} \cup \{U_n\}_{n\in\mathbb{Z}}$.

Proposition 3.3. *Let $n \in \mathbb{N}_0$ be arbitrary, then*

(i) $(\mathfrak{C}^*)^{-1}(V_n) \subseteq V_{n-1} \cup V_{-n-1}$,
(ii) $(\mathfrak{C}^*)^{-1}(V_{-n}) \subseteq V_{-n-1}$,
(iii) $(\mathfrak{C}^*)^{-1}(U_n) \subseteq U_{n-1} \cup U_{-n}$, *and*
(iv) $(\mathfrak{C}^*)^{-1}(U_{-n}) \subseteq U_{-n-1}$.

Proof. Recall $\mathfrak{C}^*(z) = \big\{w \in \widehat{\mathbb{C}} : \frac{f(w)-f(\eta(z))}{w-\eta(z)} = 0\big\}$ for all $z, w \in \widehat{\mathbb{C}}$. It follows that $(\mathfrak{C}^*)^{-1}(w) \subseteq \eta(f^{-1}(f(w)))$. If $z \in (\mathfrak{C}^*)^{-1}(w)$, then $f'(w) = 0$ or $\eta(z) \neq w$. Recall Ω is the univalent image of $\mathbb{D}$ under f, and thus $f\big(\overline{\mathbb{D}}\big) = \overline{\Omega}$ and $f(\partial\mathbb{D}) = \partial\Omega$. If $w \in \mathbb{D}$ and $z \in (\mathfrak{C}^*)^{-1})(w)$, since f is injective on $\mathbb{D}$ and thus $f'(w) \neq 0$, we have $\eta(z) \notin \mathbb{D}$. It follows that $\eta(z) \in \mathbb{D}^*$, i.e., $z \in \mathbb{D}$ because $f(\partial\mathbb{D}) = \partial\Omega$ and $f(\mathbb{D}) = \Omega$. As a result, $(\mathfrak{C}^*)^{-1}(\mathbb{D}) \subseteq \mathbb{D}$.

Firstly, $(\mathfrak{C}^*)^{-1}(V_0) \subseteq \eta\big(f^{-1}\big(f\big(f^{-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big)\big)\big) = \eta\big(f^{-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big)$. Recall $f(\overline{\mathbb{D}}) = \overline{\Omega}$, so $f^{-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big) \subseteq \mathbb{D}^*$. By (3.3),

$$\eta\big(f^{-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) = \eta\big(f^{-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \cap \mathbb{D} = f^{-1}\big(\tau^{-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \cap \mathbb{D} = V_{-1}.$$

Hence, $(\mathfrak{C}^*)^{-1}(V_0) \subseteq V_{-1}$.

For each $w \in U_0 = f^{-1}(\partial\Omega \setminus S(\tau)) \cap \partial\mathbb{D}$, if $z \in (\mathfrak{C}^*)^{-1}(w)$, i.e., $\frac{f(w)-f(\eta(z))}{(w-\eta(z))} = 0$, then $f(\eta(z)) = f(w) \in \partial\Omega \setminus S(\tau)$, and thus $\eta(z) \in \widehat{\mathbb{C}} \setminus \mathbb{D} = \mathbb{D}^* \cup \partial\mathbb{D}$, i.e., $z \in \mathbb{D} \cup \partial\mathbb{D}$. Moreover, we have $\eta(z) \neq w$ or $f'(w) = 0$. We argue by contradiction and assume $z \in \partial\mathbb{D}$, then $\eta(z) \neq w$ indicates that $f(w) = f(\eta(z))$ is a double point on $\partial\Omega$, and $f'(w) = 0$ indicates that $f(w)$ is a cusp on $\partial\Omega$. This contradicts $f(w) \in \partial\Omega \setminus S(\tau)$ and we conclude $z \in \mathbb{D}$, so $(\mathfrak{C}^*)^{-1}(U_0) \subseteq \mathbb{D}$. Consequently, by (3.3),

$$(\mathfrak{C}^*)^{-1}(U_0) \subseteq \eta(f^{-1}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D} = f^{-1}(\tau^{-1}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D} = U_{-1}.$$

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

Recall $(\mathfrak{C}^*)^{-1}(\mathbb{D}) \subseteq \mathbb{D}$ and $V_{-n} = f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \cap \mathbb{D}$. By (3.3),

$$(\mathfrak{C}^*)^{-1}(V_{-n}) \subseteq \eta\big(f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big)\big) \cap \mathbb{D} = f^{-1}\big(\tau^{-(n+1)}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \cap \mathbb{D} = V_{-n-1}.$$

Recall $f(\partial\mathbb{D}) = \partial\Omega$, so $f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \subseteq f^{-1}(\Omega) \subseteq \mathbb{D} \cup \mathbb{D}^*$. By (3.3) we have

$$\begin{aligned}(\mathfrak{C}^*)^{-1}(V_n) &\subseteq \eta\big(f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big)\big)\\ &= \big(\eta\big(f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big)\big) \cap \mathbb{D}\big) \cup \big(\eta\big(f^{-1}\big(\tau^{-n}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big)\big) \cap \mathbb{D}^*\big)\\ &= \big(f^{-1}\big(\tau^{-n-1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \cap \mathbb{D}\big) \cup \big(f^{-1}\big(\tau^{-n+1}\big(\widehat{\mathbb{C}} \setminus \overline{\Omega}\big)\big) \cap \mathbb{D}^*\big)\\ &= V_{-n-1} \cup V_{n-1}.\end{aligned}$$

Recall $(\mathfrak{C}^*)^{-1}(\mathbb{D}) \subseteq \mathbb{D}$ and $U_{-n} = f^{-1}(\tau^{-n}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D}$. By (3.3),

$$(\mathfrak{C}^*)^{-1}(U_{-n}) \subseteq \eta(f^{-1}(\tau^{-n}(\partial\Omega \setminus S(\tau)))) \cap \mathbb{D} = f^{-1}(\tau^{-(n+1)}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D} = U_{-n-1}.$$

Now we show $(\mathfrak{C}^*)^{-1}(U_n) \subseteq U_{n-1} \cup U_{-n}$. Recall $U_n = f^{-1}(\tau^{-(n-1)}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D}^*$. If $n \geqslant 2$, then $f^{-1}(\tau^{-n+1}(\partial\Omega \setminus S(\tau))) \subseteq f^{-1}(\Omega) \subseteq \mathbb{D} \cup \mathbb{D}^*$. By (3.3) we have

$$\begin{aligned}(\mathfrak{C}^*)^{-1}(U_n) &\subseteq \eta(f^{-1}(\tau^{-(n-1)}(\partial\Omega \setminus S(\tau))))\\ &= (\eta(f^{-1}(\tau^{-(n-1)}(\partial\Omega \setminus S(\tau)))) \cap \mathbb{D}) \cup \eta(f^{-1}(\tau^{-(n-1)}(\partial\Omega \setminus S(\tau)))) \cap \mathbb{D}^*)\\ &= (f^{-1}(\tau^{-n}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D}) \cup (f^{-1}(\tau^{-(n-2)}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D}^*)\\ &= U_{-n} \cup U_{n-1}.\end{aligned}$$

Assume $n = 1$. if $w \in U_1$ and $z \in (\mathfrak{C}^*)^{-1}(w)$, then $f(\eta(z)) = f(w) \in \partial\Omega \setminus S(\tau)$ implies $z \in \mathbb{D} \cup \partial\mathbb{D}$. If $z \in \partial\mathbb{D}$, then $f(z) = f(\eta(z)) \in \partial\Omega \setminus S(\tau)$, and thus $z \in f^{-1}(\partial\Omega \setminus S(\tau)) \cap \partial\mathbb{D} = U_0$. If $z \in \mathbb{D}$, then $\tau(f(z)) = f(\eta(z)) \in \partial\Omega \setminus S(\tau)$, and thereby, $z \in \tau^{-1}(f^{-1}(\partial\Omega \setminus S(\tau))) \cap \mathbb{D} = U_{-1}$. Therefore we have $(\mathfrak{C}^*)^{-1}(U_1) \subseteq U_0 \cup U_{-1}$. $\square$

Proposition 3.4. *If $\mathcal{Q}$ is a transition probability kernel on $\widehat{\mathbb{C}}$ supported by $\mathfrak{C}^*$ and μ is a $\mathcal{Q}$-invariant probability measure on $\widehat{\mathbb{C}}$, then $\mu\big(\widetilde{K(\tau)}\big) = 1$.*

Proof. By [MA99, Theorem 3.1], if A_1, $A_2 \in \mathscr{B}\big(\widehat{\mathbb{C}}\big)$ satisfy $(\mathfrak{C}^*)^{-1}(A_1) \subseteq A_2$, then $\mu(A_2) \geqslant \mu(A_1)$.

As a result, the statements in Proposition 3.3 indicate $\mu(V_{-n-1}) \geqslant \mu(V_{-n})$ and $\mu(U_{-n-1}) \geqslant \mu(U_{-n})$ for all $n \in \mathbb{N}_0$. For each $n \in \mathbb{N}_0$ and $k \in \mathbb{N}$, we have

$$k\mu(V_{-n}) \leqslant \sum_{j=0}^{k-1} \mu(V_{-n-j}) \leqslant \mu\big(\widetilde{T^\infty(\tau)}\big),$$

where the last inequality is ensured by Lemma 3.2. Hence we have $k\mu(V_{-n}) \leqslant \frac{1}{k}$ for all $k \in \mathbb{N}$, and thus $\mu(V_{-n}) = 0$ for each $n \in \mathbb{N}_0$. Similarly, $\mu(U_{-n}) = 0$ for each $n \in \mathbb{N}_0$.

Again, by Proposition 3.3, for each $n \in \mathbb{N}$ we have $\mu(V_n) \leqslant \mu(V_{n-1}) + \mu(V_{-n-1}) = \mu(V_{n-1})$ and $\mu(U_n) \leqslant \mu(U_{n-1}) + \mu(U_{-n}) = \mu(U_{n-1})$. This implies $\mu(V_n) \leqslant \mu(V_{n-1}) \leqslant \cdots \leqslant \mu(V_0) = 0$, and thus $\mu(V_n) = 0$ for every $n \in \mathbb{N}$. Similarly, $\mu(U_n) = 0$ for every $n \in \mathbb{N}$.

By Lemma 3.2, we conclude $\mu\big(\widetilde{T^\infty(\tau)}\big) = 0$, and therefore, $\mu\big(\widetilde{K(\tau)}\big) = 1$. $\square$

Propositions 3.1, 3.4, and 6.17 imply the following corollary which leads us to shift our focus to $\mathfrak{C}^*|_{\widetilde{K(\tau)}}$.

Corollary 3.5. *Let $\phi \in C(\mathcal{O}_2(\mathfrak{C}^*), \mathbb{R})$. Set $\phi_K := \phi|_{\mathcal{O}_2(\mathfrak{C}^*|_{\widetilde{K(\tau)}})}$. Then $\mathfrak{C}^*|_{\widetilde{K(\tau)}}$ is a correspondence on $\widetilde{K(\tau)}$ and $P(\mathfrak{C}^*, \phi) = P\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}}, \phi_K\big)$*

The following proposition is to describe the dynamics of $\mathfrak{C}^*|_{\widetilde{K(\tau)}}$.

Proposition 3.6 ([LMM23, Proposition 2.5])**.** *The following statements hold:*

(i) *For all $z \in \widetilde{K(\tau)}$, we have $\#\mathfrak{C}^*|_{\widetilde{K(\tau)}}(z) = \#\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}}\big)^{-1}(z) = d$.*

(ii) *If $w \in \mathfrak{C}^*|_{\widetilde{K(\tau)}}(z)$, then $z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$ implies that $w \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$, moreover, $w \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}}$ implies that $z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}}$.*

(iii) *The correspondence $\mathfrak{C}^*|_{\widetilde{K(\tau)}}$ has one forward branch carrying $\widetilde{K(\tau)} \cap \overline{\mathbb{D}}$ onto itself with degree d, which is topologically conjugate to $\tau\colon K(\tau) \to K(\tau)$, and the remaining forward branches carry $\widetilde{K(\tau)} \cap \overline{\mathbb{D}}$ onto $\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$.*

(iv) *The correspondence $\mathfrak{C}^*|_{\widetilde{K(\tau)}}$ has a backward branch carrying $\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$ onto itself with degree d, which is topologically conjugate to $\tau\colon K(\tau) \to K(\tau)$, and the remaining backward branches carry $\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$ onto $\widetilde{K(\tau)} \cap \overline{\mathbb{D}}$.*

Write $G := \mathfrak{C}^*|_{\widetilde{K(\tau)}}$. Proposition 3.6 (iii) indicates that $\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}}$ is induced by a single-valued continuous map, so we suppose it is induced by $g\colon \widetilde{K(\tau)} \cap \overline{\mathbb{D}} \to \widetilde{K(\tau)} \cap \overline{\mathbb{D}}$, i.e., $\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}} = \mathcal{C}_g$ (for the definition of $\mathcal{C}_g$, the correspondence induced by g, see Appendix B.2 for details). Similarly, Proposition 3.6 (iv) indicates that $\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}^*}}\big)^{-1} = \mathcal{C}_{g^*}$, where $g^*\colon \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*} \to \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$ is a single-valued continuous map.

Let $\phi \in C(\mathcal{O}_2(\mathfrak{C}^*), \mathbb{R})$. Recall $\phi_K = \phi|_{\mathcal{O}_2(\mathfrak{C}^*|_{\widetilde{K(\tau)}})} \in C(\mathcal{O}_2(G), \mathbb{R})$. Note that each $(x, y) \in \mathcal{O}_2(\mathcal{C}_g)$ is of the form $(x, g(x))$, so functions on $\mathcal{O}_2\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}}\big) = \mathcal{O}_2(\mathcal{C}_g)$ actually only depend on the first coordinate. As a result, we can choose $\varphi \in C\big(\widetilde{K(\tau)} \cap \overline{\mathbb{D}}, \mathbb{R}\big)$ such that $\varphi(x) = \phi(x, y)$ for all $(x, y) \in \mathcal{O}_2\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}}\big)$. Similarly, we can choose $\varphi^* \in C\big(\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}, \mathbb{R}\big)$ such that $\varphi^*(y) = \phi(x, y)$ for all $(x, y) \in \mathcal{O}_2\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}^*}}\big)$.

With the dynamics of $G = \mathfrak{C}^*|_{\widetilde{K(\tau)}}$ given by Proposition 3.6, we estimate the topological pressure $P(G, \phi_K)$ (see Subsection 4.2).

Proposition 3.7. *Let $\phi \in C(\mathcal{O}_2(\mathfrak{C}^*), \mathbb{R})$ and G, g, g^*, ϕ_K, φ, and φ^* be given above. We have $P(G, \phi_K) = \max\{P(g, \varphi),\ P(g^*, \varphi^*)\}$.*

Proof. We briefly recall the definition of topological pressure for correspondences from Definition 4.6: Let T be a correspondence on a compact metric space (X, d) and $\psi \in C(\mathcal{O}_2(T), \mathbb{R})$. The topological pressure is

$$P(T, \psi) = \lim_{\epsilon \to 0^+} \limsup_{n \to +\infty} \frac{1}{n} \log\Big(\sup_{E_n(\epsilon)} \sum_{\underline{x} \in E_n(\epsilon)} \exp(S_n\psi(\underline{x}))\Big),$$

where $E_n(\epsilon)$ ranges over all ϵ-separated subsets of $(\mathcal{O}_{n+1}(T), d_{n+1})$, and $S_n\psi(x_1, \dots, x_{n+1}) = \sum_{k=1}^{n} \psi(x_k, x_{k+1})$ for all $(x_1, \dots, x_{n+1}) \in \mathcal{O}_{n+1}(T)$.

Now we return to the estimation on the topological pressure of $G = \mathfrak{C}^*|_{\widetilde{K(\tau)}}$. Let d be the spherical metric on $\widehat{\mathbb{C}}$ and d_n be the metrics given by (2.1) and (2.2).

Fix arbitrary $n \in \mathbb{N}$ and $\epsilon > 0$, write $S_n\phi(x_1, \dots, x_{n+1}) \coloneqq \sum_{k=1}^{n} \phi(x_k, x_{k+1})$ for all $(x_1, \dots, x_{n+1}) \in \mathcal{O}_{n+1}(G)$,

$$\alpha(n, \epsilon) \coloneqq \sup_{E_n(\epsilon)} \sum_{\underline{x} \in E_n(\epsilon)} \exp(S_n\phi(\underline{x})), \tag{3.4}$$

where $E_n(\epsilon)$ ranges over all ϵ-separated subsets of $(\mathcal{O}_{n+1}(\mathcal{C}_g), d_{n+1})$, and

$$\beta(n, \epsilon) \coloneqq \sup_{F_n(\epsilon)} \sum_{\underline{x} \in F_n(\epsilon)} \exp(S_n\phi(\underline{x})), \tag{3.5}$$

where $F_n(\epsilon)$ ranges over all ϵ-separated subsets of $(\mathcal{O}_{n+1}(\mathcal{C}_{g^*}^{-1}), d_{n+1})$.

Recall $\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}} = \mathcal{C}_g$, $\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}^*}} = \mathcal{C}_{g^*}^{-1}$, $\varphi(x) = \phi(x, y)$ for all $(x, y) \in \mathcal{O}_2(\mathcal{C}_g)$, and $\varphi^*(y) = \phi(x, y)$ for all $(x, y) \in \mathcal{O}_2(\mathcal{C}_{g^*})$. By Propositions B.3 and 4.8, we have $P(g, \varphi) = P\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}}, \phi|_{\mathcal{O}_2(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}}})}\big)$ and $P(g^*, \varphi^*) = P\big(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}^*}}, \phi|_{\mathcal{O}_2(\mathfrak{C}^*|_{\widetilde{K(\tau)}\cap\overline{\mathbb{D}^*}})}\big)$. By (3.4), (3.5), and the definition of the topological pressure for correspondences, we have

$$\begin{aligned} P(g, \varphi) &= \lim_{\epsilon \to 0^+} \limsup_{n \to +\infty} \frac{1}{n} \log(\alpha(n, \epsilon)), \\ P(g^*, \varphi^*) &= \lim_{\epsilon \to 0^+} \limsup_{n \to +\infty} \frac{1}{n} \log(\beta(n, \epsilon)). \end{aligned} \tag{3.6}$$

Fix arbitrary $\epsilon > 0$, $n \in \mathbb{N}$, and ϵ-separated subset $W_n(\epsilon)$ of $\mathcal{O}_{n+1}(G)$. By Proposition 3.6 (ii), we can write $W_n(\epsilon) = \bigcup_{k=-1}^{n} W_{n,k}(\epsilon)$, where

$$W_{n,k}(\epsilon) \coloneqq \{(x_0, \dots, x_n) \in W_n(\epsilon) : x_i \in \mathbb{D} \text{ for } i \leqslant k \text{ and } x_i \in \mathbb{D}^* \text{ for } i > k\}.$$

For each $k \in [n]$ and each $(x_0, \dots, x_k) \in E_k(\epsilon/2)$, set

$$W_{n,k,x_0,\dots,x_k}(\epsilon) \coloneqq \{(y_0, \dots, y_n) \in W_{n,k}(\epsilon) : d(x_i, y_i) < \epsilon/2 \text{ for all } i \in [k]\},$$

and fix an arbitrary maximal $\frac{\epsilon}{2}$-separated subset $E_k(\epsilon/2)$ of $\mathcal{O}_{k+1}(\mathcal{C}_g)$. For each $(y_0, \dots, y_n) \in W_{n,k}(\epsilon)$, the maximality ensures that there must be some $(x_0, \dots, x_k) \in E_k(\epsilon/2)$ such that $d(x_i, y_i) < \epsilon/2$ for all $i \in [k]$, so

$$\bigcup_{(x_0,\dots,x_k) \in E_k(\frac{\epsilon}{2})} W_{n,k,x_0,\dots,x_k}(\epsilon) = W_{n,k}(\epsilon). \tag{3.7}$$

Fix arbitrary $k \in [n-1]$ and $(x_0, \dots, x_k) \in E_k(\epsilon/2)$. For each $\underline{y} = (y_0, \dots, y_n) \in W_{n,k,x_0,\dots,x_k}(\epsilon)$, $d(x_i, y_i) < \epsilon/2$ for all $i \in [k]$ implies that

$$\sum_{j=0}^{k-1} \phi(y_j, y_{j+1}) \leqslant \sum_{j=0}^{k-1} \phi(x_j, x_{j+1}) + k\Delta\big(\phi, \frac{\epsilon}{2}\big),$$

where $\Delta(\phi, \delta) \coloneqq \sup\{|\phi(x_1, x_2) - \phi(y_1, y_2)| : d(x_1, y_1) < \delta \text{ and } d(x_2, y_2) < \delta\}$ for all $\delta > 0$. As a result, we have

$$S_n\phi(\underline{y}) \leqslant \sum_{j=0}^{k-1} \phi(x_j, x_{j+1}) + k\Delta\Big(\phi, \frac{\epsilon}{2}\Big) + \|\phi\|_\infty + \sum_{j=k+1}^{n-1} \phi(y_j, y_{j+1}). \tag{3.8}$$

Because $W_{n,k,x_0,\dots,x_k}(\epsilon)$ is contained in $W_n(\epsilon)$, an ϵ-separated subset of $\mathcal{O}_{n+1}(G)$, for each pair of distinct orbits $(y_0, \dots, y_n)$, $(z_0, \dots, z_n) \in W_{n,k,x_0,\dots,x_k}(\epsilon)$, there exists $l \in [n]$ such that $d(y_l, z_l) \geqslant \epsilon$. Such an integer l must be greater than k, because for each $j \in [k]$, we have $d(y_j, z_j) \leqslant d(x_j, y_j) + d(x_j, z_j) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. As a result, the set $\{(y_{k+1}, \dots, y_n) : (y_1, \dots, y_n) \in W_{n,k,x_0,\dots,x_k}(\epsilon)\}$ is an ϵ-separated subset of $\mathcal{O}_{n-k}(\mathcal{C}_{g^*}^{-1})$. Thus by (3.8) and (3.5) we have

$$\sum_{\underline{y} \in W_{n,k,x_0,\dots,x_k}(\epsilon)} e^{S_n\phi(\underline{y})} \leqslant \beta(n-k-1, \epsilon) \exp\bigg(\sum_{j=0}^{k-1} \phi(x_j, x_{j+1}) + k\Delta\Big(\phi, \frac{\epsilon}{2}\Big) + \|\phi\|_\infty\bigg).$$

Consequently, by (3.7) and (3.4) we have

$$\sum_{\underline{y} \in W_{n,k}(\epsilon)} e^{S_n\phi(\underline{y})} \leqslant \alpha\Big(k, \frac{\epsilon}{2}\Big) \beta(n-k-1, \epsilon) \exp\Big(n\Delta\Big(\phi, \frac{\epsilon}{2}\Big) + \|\phi\|_\infty\Big). \tag{3.9}$$

Note that (3.9) holds for $k \in [n-1]$, so we need to consider $k = -1$ and $k = n$ independently. Specifically, $W_{n,-1}(\epsilon) = \{(x_0, \dots, x_n) \in W_n(\epsilon) : x_i \in \mathbb{D}^* \text{ for all } i \in [n]\}$ is ϵ-separated in $\mathcal{O}_{n+1}(\mathcal{C}_{g^*}^{-1})$ and $W_{n,n}(\epsilon) = \{(x_0, \dots, x_n) \in W_n(\epsilon) : x_i \in \mathbb{D} \text{ for all } i \in [n]\}$ is ϵ-separated in $\mathcal{O}_{n+1}(\mathcal{C}_g)$. By (3.4) and (3.5) we have $\sum_{\underline{y} \in W_{n,-1}(\epsilon)} \exp(S_n\phi(\underline{y})) \leqslant \beta(n, \epsilon)$ and $\sum_{\underline{y} \in W_{n,n}(\epsilon)} \exp(S_n\phi(\underline{y})) \leqslant \alpha(n, \epsilon)$. By (3.9) and since $W_n(\epsilon) = \bigcup_{k=-1}^{n} W_{n,k}(\epsilon)$, we have

$$\sum_{\underline{y} \in W_n(\epsilon)} e^{S_n\phi(\underline{y})} \leqslant \alpha(n, \epsilon) + \beta(n, \epsilon) + e^{n\Delta(\phi, \frac{\epsilon}{2}) + \|\phi\|_\infty} \sum_{k=1}^{n-1} \alpha\Big(k, \frac{\epsilon}{2}\Big) \beta(n-k-1, \epsilon).$$

Since $\lim_{\epsilon \to 0^+} \Delta\big(\phi, \frac{\epsilon}{2}\big) = 0$ due to the uniform continuity of ϕ, applying (3.6) we conclude

$$\begin{aligned} P(G, \phi_K) &= \lim_{\epsilon \to 0^+} \limsup_{n \to +\infty} \frac{1}{n} \log\Big(\sup_{W_n(\epsilon)} \sum_{\underline{y} \in W_n(\epsilon)} e^{S_n\phi(\underline{y})}\Big) \\ &\leqslant \max\{P(g, \varphi),\ P(g^*, \varphi^*)\}. \end{aligned} \tag{3.10}$$

Additionally, for arbitrary $n \in \mathbb{N}$ and $\epsilon > 0$, every ϵ-separated subset of $\mathcal{O}_{n+1}(\mathcal{C}_g)$ is also an ϵ-separated subset of $\mathcal{O}_{n+1}(G)$. Thus by Definition 4.6 we have $P(G, \phi_K) \geqslant P\big(\mathcal{C}_g, \phi|_{\mathcal{O}_2(\mathcal{C}_g)}\big)$. Similarly, we have $P(G, \phi_K) \geqslant P\big(\mathcal{C}_{g^*}^{-1}, \phi|_{\mathcal{O}_2(\mathcal{C}_{g^*}^{-1})}\big)$. Recall $P\big(\mathcal{C}_g, \phi|_{\mathcal{O}_2(\mathcal{C}_g)}\big) = P(g, \varphi)$ and $P\big(\mathcal{C}_{g^*}^{-1}, \phi|_{\mathcal{O}_2(\mathcal{C}_{g^*}^{-1})}\big) = P(g^*, \varphi^*)$. Hence $P(G, \phi_K) \geqslant \max\{P(g, \varphi),\ P(g^*, \varphi^*)\}$. Therefore, by (3.10) we conclude $P(G, \phi_K) = \max\{P(g, \varphi),\ P(g^*, \varphi^*)\}$. $\square$

Now, we prove Theorem E.

Proof of Theorem E. In this proof, if μ is a Borel probability measure on some Borel subset K of $\widehat{\mathbb{C}}$, then $\widehat{\mu}$ will refer to the Borel probability measure on $\widehat{\mathbb{C}}$ given by $\widehat{\mu}(A) := \mu(K \cap A)$ for all $A \in \mathscr{B}\big(\widehat{\mathbb{C}}\big)$. Corollary 3.5 and Proposition 3.7 imply that $P(\mathfrak{C}^*, \phi) = \max\{P(g, \varphi),\ P(g^*, \varphi^*)\}$. We establish Theorem E by discussing the following two cases:

Case 1. $P(\mathfrak{C}^*, \phi) = P(g, \varphi)$.

By the classical Variational Principle (B.7), we have

$$P(g, \varphi) = \sup_{\mu} \biggl\{ h_\mu(g) + \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}}} \varphi \, \mathrm{d}\mu \biggr\}, \tag{3.11}$$

where μ ranges over all g-invariant probability measures on $\widetilde{K(\tau)} \cap \overline{\mathbb{D}}$.

Fix an arbitrary g-invariant Borel probability measure μ on $\widetilde{K(\tau)} \cap \overline{\mathbb{D}}$. By Lemma B.2 and (B.5), μ is $\widehat{g}$-invariant and $h_\mu(g) = h_\mu(\widehat{g})$, where $\widehat{g}$ is the transition probability kernel on $\widetilde{K(\tau)} \cap \overline{\mathbb{D}}$ induced by g given in Definition B.1. We choose a Borel measurable branch a_0 of $\mathfrak{C}^*$, where the existence of a_0 is ensured by [MA99, Lemma 1.1]. Let $\mathcal{S}$ be a transition probability kernel on $\widehat{\mathbb{C}}$ given by

$$\mathcal{S}(z, A) := \begin{cases} \mathbb{1}_A(g(z)), & z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}}, \\ \mathbb{1}_A(a_0(z)), & z \notin \widetilde{K(\tau)} \cap \overline{\mathbb{D}} \end{cases}$$

for all $z \in \widehat{\mathbb{C}}$ and $A \in \mathscr{B}\big(\widehat{\mathbb{C}}\big)$. It follows that $\mathcal{S}_z = \delta_{g(z)}$ for all $z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}}$ and that $\mathcal{S}$ is supported by $\mathfrak{C}^*$. By Lemma 5.28, $\widehat{\mu}$ is $\mathcal{S}$-invariant and $h_{\widehat{\mu}}(\mathcal{S}) = h_\mu(\widehat{g}) = h_\mu(g)$. Recall $\varphi(z) = \phi(z, g(z))$ for all $z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}}$. We have

$$\begin{aligned} \int_{\widehat{\mathbb{C}}} \int_{\mathfrak{C}^*(z)} \phi(z, w) \, \mathrm{d}\mathcal{S}_z(w) \, \mathrm{d}\widehat{\mu}(z) &= \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}}} \int_{\mathfrak{C}^*(z)} \phi(z, w) \, \mathrm{d}\delta_{g(z)}(w) \, \mathrm{d}\mu(z) \\ &= \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}}} \phi(z, g(z)) \, \mathrm{d}\mu(z) \\ &= \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}}} \varphi \, \mathrm{d}\mu. \end{aligned} \tag{3.12}$$

Recall $\mathcal{S}$ is supported on $\mathfrak{C}^*$. By (3.11), (3.12) $h_{\widehat{\mu}}(\mathcal{S}) = h_\mu(g)$, and $P(\mathfrak{C}^*, \phi) = P(g, \varphi)$, we have

$$P(\mathfrak{C}^*, \phi) \leqslant \sup_{\mathcal{Q}, \mu} \biggl\{ h_\mu(\mathcal{Q}) + \int_{\widehat{\mathbb{C}}} \int_{\mathfrak{C}^*(x_1)} \phi(x_1, x_2) \, \mathrm{d}\mathcal{Q}_{x_1}(x_2) \, \mathrm{d}\mu(x_1) \biggr\}, \tag{3.13}$$

where $\mathcal{Q}$ ranges over all transition probability kernels on $\widehat{\mathbb{C}}$ supported by $\mathfrak{C}^*$ and μ ranges over all $\mathcal{Q}$-invariant probability measures on $\widehat{\mathbb{C}}$. Therefore, (1.5) follows by Theorem D.

Case 2. $P(\mathfrak{C}^*, \phi) = P(g^*, \varphi^*)$.

By the classical Variational Principle (B.7), we have

$$P(g^*, \varphi^*) = \sup_{\mu} \left\{ h_\mu(g^*) + \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}} \varphi^* \, \mathrm{d}\mu \right\}, \tag{3.14}$$

where μ ranges over all g^*-invariant probability measures on $\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$.

We proceed with the proof in a manner similar to the previous case. Fix an arbitrary g^*-invariant Borel probability measure μ on $\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$. We choose a Borel measurable branch a_1 of $(\mathfrak{C}^*)^{-1}$. Let $\mathcal{S}$ be a transition probability kernel on $\widehat{\mathbb{C}}$ given by

$$\mathcal{S}(z, A) := \begin{cases} \mathbb{1}_A(g^*(z)), & z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}, \\ \mathbb{1}_A(a_1(z)), & z \notin \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*} \end{cases}$$

for all $z \in \widehat{\mathbb{C}}$ and $A \in \mathscr{B}(\widehat{\mathbb{C}})$. It is supported by $(\mathfrak{C}^*)^{-1}$, and thus the measure $\widehat{\mu}\mathcal{S}^{[1]}$ is supported on $\mathcal{O}_2((\mathfrak{C}^*)^{-1})$. By (B.5), Lemmas B.2 and 5.28, $\widehat{\mu}$ is $\mathcal{S}$-invariant and $h_{\widehat{\mu}}(\mathcal{S}) = h_\mu(g^*)$. Recall $\varphi^*(z) = \phi(g^*(z), z)$ for all $z \in \widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}$. By (A.11) in Lemma A.9, we have

$$\int_{\widehat{\mathbb{C}}^2} \phi(w, z) \, \mathrm{d}\big(\widehat{\mu}\mathcal{S}^{[1]}\big)(z, w) = \int_{\widehat{\mathbb{C}}} \int_{\widehat{\mathbb{C}}} \phi(w, z) \, \mathrm{d}\mathcal{S}_z(w) \, \mathrm{d}\widehat{\mu}(z) = \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}} \varphi^* \, \mathrm{d}\mu, \tag{3.15}$$

which corresponds to (3.12) in the previous case. By (A.10), we have $\big(\widehat{\mu}\mathcal{S}^{[1]}\big) \circ \widetilde{\pi}_2^{-1} = \widehat{\mu}\mathcal{S} = \widehat{\mu}$. Choose a backward conditional transition probability kernel $\mathcal{R}$ of $\widehat{\mu}\mathcal{S}^{[1]}$ from $\widehat{\mathbb{C}}$ to $\widehat{\mathbb{C}}$ supported on $\mathcal{O}_2((\mathfrak{C}^*)^{-1})$. Definition A.14 (a) and the fact that $\mu\mathcal{S}^{[1]}$ is supported by $(\mathfrak{C}^*)^{-1}$ indicate that $\mathcal{R}$ is supported on $\mathfrak{C}^*$. By Remark A.15, Definition A.14 (b) leads to $\big(\widehat{\mu}\mathcal{S}^{[1]}\big) \circ \iota_2^{-1} = \widehat{\mu}\mathcal{R}^{[1]}$, where $\iota_2(z, w) = (w, z)$ for all $z, w \in \widehat{\mathbb{C}}$. By Proposition 5.23, $\widehat{\mu}$ is $\mathcal{R}$-invariant and $h_{\widehat{\mu}}(\mathcal{R}) = h_{\widehat{\mu}}(\mathcal{S}) = h_\mu(g^*)$. By (A.11), (3.15), and $\big(\widehat{\mu}\mathcal{S}^{[1]}\big) \circ \iota_2^{-1} = \widehat{\mu}\mathcal{R}^{[1]}$, we have

$$\begin{aligned} \int_{\widehat{\mathbb{C}}} \int_{\mathfrak{C}^*(z)} \phi(z, w) \, \mathrm{d}\mathcal{R}_z(w) \, \mathrm{d}\widehat{\mu}(z) &= \int_{\widehat{\mathbb{C}}^2} \phi \, \mathrm{d}\big(\widehat{\mu}\mathcal{R}^{[1]}\big) \\ &= \int_{\widehat{\mathbb{C}}^2} \phi \circ \iota_2 \, \mathrm{d}\big(\widehat{\mu}\mathcal{S}^{[1]}\big) \\ &= \int_{\widetilde{K(\tau)} \cap \overline{\mathbb{D}^*}} \varphi^* \, \mathrm{d}\mu. \end{aligned} \tag{3.16}$$

Recall $\mathcal{R}$ is supported on $\mathfrak{C}^*$. By (3.14), (3.16), $h_{\widehat{\mu}}(\mathcal{R}) = h_\mu(g^*)$, and $P(\mathfrak{C}^*, \phi) = P(g^*, \varphi^*)$, we get that (3.13) holds, and therefore, (1.5) follows by Theorem D. $\square$

3.2. **A family of hyperbolic holomorphic correspondences.** In this subsection, we aim to establish Theorems F and G. To begin with, we explain in detail the definition of Julia set of $\mathbf{f}_c = z^{q/p} + c$ mentioned in Subsection 1, see for example, [Siq15, Definition 6.31] or [SS17, Section 2.1].

A *periodic orbit* is a sequence $\{z_k\}_0^n$, where $n \in \mathbb{N}$, $z_k \in \widehat{\mathbb{C}}$ for each $k \in [n]$, satisfying $z_k \in \mathbf{f}_c(z_{k-1})$ for each $k \in (n]$ and $z_n = z_0$. For each $k \in (n]$, if z_{k-1} does not belong to $\{0, \infty\}$, we can choose a branch of the holomorphic function

$\phi_k \colon z \mapsto \exp\bigl(\frac{1}{q}\log(z^p)\bigr) + c$ in a neighborhood of z_{k-1} which maps z_{k-1} to z_k. We say that a periodic orbit $\{z_k\}_0^n$ is *repelling* if none of elements in that orbit belong to $\{0, \infty\}$ and $|(\phi_n \circ \cdots \circ \phi_2 \circ \phi_1)'(z_0)| > 1$. The *Julia set* $J(\mathbf{f}_c)$ is defined as the closure of the union of all repelling periodic orbits of $\mathbf{f}_c$.

Recall $P_c := \overline{\bigcup_{n\in\mathbb{N}} \mathbf{f}_c^n(0)}$. The following proposition is formulated from [Siq22, Theorems 4.4, 5.1, and 5.8].

Proposition 3.8. *There is an open set $H_{q/p}$ containing both $\mathbb{C} \setminus M_{q/p,0}$ and every simple center with the property that for each $c \in H_{q/p}$, the following statements hold:*

(i) $\boldsymbol{f}_c^{-1}(J(\boldsymbol{f}_c)) = J(\boldsymbol{f}_c) \neq \emptyset$.
(ii) *The set $\mathbb{C} \setminus P_c$ is a hyperbolic Riemann surface and $J(\boldsymbol{f}_c) \subseteq \mathbb{C} \setminus P_c$.*
(iii) *If we denote by d_c the hyperbolic metric on $\mathbb{C} \setminus P_c$, then there exist constants $\lambda > 1$ and $\delta > 0$ depending on p, q, and c with the following property:*

 If a pair of distinct points z_1, $z_2 \in J(\boldsymbol{f}_c)$ satisfy $d_c(z_1, z_2) < \delta$, then for each $w_1 \in J(\boldsymbol{f}_c) \cap \boldsymbol{f}_c(z_1)$ and each $w_2 \in J(\boldsymbol{f}_c) \cap \boldsymbol{f}_c(z_2)$, we have $d_c(w_1, w_2) > \lambda d_c(z_1, z_2)$.
(iv) *For every open set V in $\mathbb{C}$ that intersects $J(\boldsymbol{f}_c)$, there exists $n \in \mathbb{N}$ such that $\boldsymbol{f}_c^n(V \cap J(\boldsymbol{f}_c)) = J(\boldsymbol{f}_c)$.*

The following proposition about $\mathbf{f}_c$ when c is closed to 0 is formulated from [Siq15, Corollaries 4.6,4.8, Theorems 3.5, and 2.7].

Proposition 3.9. *There is an open neighborhood $U_{q/p}$ of 0 such that for every $c \in U_{q/p}$, the following statements hold:*

(i) $\boldsymbol{f}_c^{-1}(J(\boldsymbol{f}_c)) = J(\boldsymbol{f}_c) \neq \emptyset$.
(ii) *There exist constants $\lambda > 1$ and $\delta > 0$ depending on p, q, and c with the following property:*

 If a pair of distinct points z_1, $z_2 \in J(\boldsymbol{f}_c)$ satisfy $|z_1 - z_2| < \delta$, then for each $w_1 \in J(\boldsymbol{f}_c) \cap \boldsymbol{f}_c(z_1)$ and each $w_2 \in J(\boldsymbol{f}_c) \cap \boldsymbol{f}_c(z_2)$, we have $|w_1 - w_2| > \lambda |z_1 - z_2|$.
(iii) *For every open set V in $\mathbb{C}$ that intersects $J(\boldsymbol{f}_c)$, there exists $n \in \mathbb{N}$ such that $\boldsymbol{f}_c^n(V \cap J(\boldsymbol{f}_c)) = J(\boldsymbol{f}_c)$.*

Recall $\mathbf{f}_c|_J(z) = J(\mathbf{f}_c) \cap \mathbf{f}_c(z)$ for all $z \in J(\mathbf{f}_c)$. for all $z \in \widehat{\mathbb{C}}$. Now we are prepared to establish Theorems F and G.

Proof of Theorems F and G. Note that $0 \in M_{q/p}$ is not a simple center. We choose the open sets $H_{q/p}$ and $U_{q/p}$ as in Propositions 3.8 and 3.9, respectively, such that $H_{q/p} \cap U_{q/p} = \emptyset$. For every $c \in H_{q/p}$, we denote by d_c the hyperbolic metric on the hyperbolic Riemann surface $\mathbb{C} \setminus P_c$, where the hyperbolicity of $\mathbb{C} \setminus P_c$ is ensured by Proposition 3.8 (ii). For every $c \in U_{q/p}$, we denote by d_c the Euclidian metric on $\mathbb{C}$. By Theorem C, it suffices to show that $\mathbf{f}_c|_J$ is an open, distance-expanding, topologically exact correspondence on the compact metric space $(J(\mathbf{f}_c), d_c)$ for all $c \in H_{q/p} \cup U_{q/p}$.

Fix an arbitrary $c \in H_{q/p} \cup U_{q/p}$.

First, we show that $\mathbf{f}_c|_J$ is a correspondence on $J(\mathbf{f}_c)$. Indeed, for every $z \in J(\mathbf{f}_c)$, by Propositions 3.8 (i) and 3.9 (i), there is a point $w \in J(\mathbf{f}_c)$ with $z \in \mathbf{f}_c^{-1}(w)$, i.e., $w \in \mathbf{f}_c(z)$. Consequently, $\mathbf{f}_c|_J(z) = \mathbf{f}_c(z) \cap J(\mathbf{f}_c)$ is non-empty and closed for all $z \in J(\mathbf{f}_c)$. Moreover, the set $\mathcal{O}_2(\mathbf{f}_c|_J) = \mathcal{O}_2(\mathbf{f}_c) \cap J(\mathbf{f}_c)^2$ is closed in $J(\mathbf{f}_c)^2$. Hence it follows that $\mathbf{f}_c|_J$ is a correspondence on $J(\mathbf{f}_c)$.

Second, the openness of $\mathbf{f}_c|_J$ follows from $\mathbf{f}_c^{-1}(J(\mathbf{f}_c)) = J(\mathbf{f}_c)$, i.e., Propositions 3.8 (i) and 3.9 (i). Specifically, we fix arbitrary $z \in J(\mathbf{f}_c)$, an open neighborhood V of z in $J(\mathbf{f}_c)$, and $w \in \mathbf{f}_c|_J(z)$. For every point $w' \in J(\mathbf{f}_c)$ which is sufficiently close to w, a branch of $\mathbf{f}_c^{-1}$ gives a point $z' \in V$ such that $w' \in \mathbf{f}_c(z')$. This implies $z' \in \mathbf{f}_c^{-1}(J(\mathbf{f}_c)) = J(\mathbf{f}_c)$, so $w' \in \mathbf{f}_c|_J(z') \subseteq \mathbf{f}_c|_J(V)$. The argument above shows that $\mathbf{f}_c|_J(V)$ contains a neighborhood of w in $J(\mathbf{f}_c)$. Hence we conclude that $\mathbf{f}_c|_J$ is open.

Third, Propositions 3.8 (iii) and 3.9 (ii) indicate that $\mathbf{f}_c|_J$, as correspondence on the compact metric space $(J(\mathbf{f}_c), d_c)$, is distance-expanding.

Fourth, $\mathbf{f}_c^{-1}(J(\mathbf{f}_c)) = J(\mathbf{f}_c)$ implies that for arbitrary $W \subseteq \widehat{\mathbb{C}}$ and $n \in \mathbb{N}$, we have $(\mathbf{f}_c|_J)^n(W \cap J(\mathbf{f}_c)) = \mathbf{f}_c^n(W) \cap J(\mathbf{f}_c)$. Thus Propositions 3.8 (vi) and 3.9 (iii) imply that $\mathbf{f}_c|_J$ is topologically exact.

Hence, for all $c \in H_{q/p} \cup U_{q/p}$, the correspondence $\mathbf{f}_c|_J$ satisfies all the hypothesis in Theorem C, and therefore Theorem C directly yields Theorems F and G. $\square$

4. Topological pressure of correspondences

In this section, we introduce and discuss the topological pressure of correspondences. First, we recall the definition of correspondence in Subsection 4.1. Then in Subsection 4.2, we introduce the topological pressure of a correspondence with respect to a continuous potential function. Finally, in Subsection 4.3, we define a shift map for a correspondence and relate the topological pressure of this shift map to that of the correspondence, see Theorem 4.9, the main theorem of this subsection.

4.1. **Definition of correspondences.**

In this subsection, we state our definition of correspondences on compact metric spaces.

Recall from Section 2 that for a compact metric space X, the set $\mathcal{F}(X)$ consists of all non-empty closed subsets of X. The following lemma is established in [IM06, Theorems 1, 2, and 3].

Lemma 4.1. *Let (X, d) be a compact metric space. For a map $T\colon X \to \mathcal{F}(X)$, the following statements are equivalent:*

(i) (Upper-semicontinuity)[4] *For every $x \in X$ and an arbitrary open neighborhood $\mathcal{U}$ of $T(x)$, there exists an open neighborhood $\mathcal{V}$ of x such that $T(y) \subseteq \mathcal{U}$ for each $y \in \mathcal{V}$.*

(ii) $\mathcal{O}_2(T) = \{(x_1, x_2) \in X^2 : x_2 \in T(x_1)\}$ *is closed in X^2.*

(iii) $\mathcal{O}_n(T)$ *is closed in X^n for each $n \in \widehat{\mathbb{N}}$.*

[4]The notion of upper-semicontinuity in this setting is also discussed in [AF09, Definition 1.4.1].

Definition 4.2 (Correspondence)**.** Let (X,d) be a compact metric space. We say that a map $T\colon X\to\mathcal{F}(X)$ is a *correspondence on* X if it satisfies one of the equivalent conditions (i), (ii), and (iii) in Lemma 4.1.

J. Aubin and H. Frankowska discussed upper-semicontinuity (see Lemma 4.1 (i)), lower-semicontinuity, and continuity for what they called "set-valued maps" in [AF09, Chapter 1]. Let us recall the notion of continuity for correspondences on compact metric spaces from [AF09, Section 9.4.1, footnote 6].

Definition 4.3 (Continuity)**.** Let (X,d) be a compact metric space and $T\colon X\to\mathcal{F}(X)$ be a correspondence on X. if T is continuous with respect to the metric d on X and the Hausdorff distance on $\mathcal{F}(X)$, then we say that T is a *continuous* correspondence.

Recall $T^{-1}(x)=\{y\in X : x\in T(y)\}$ for all $x\in X$ from Section 2.

Lemma 4.4. *If T is a correspondence on a compact metric space X satisfying $T(X)=X$, then so is T^{-1}.*

Proof. The property $T(X)=X$ implies $T^{-1}(x)\neq\emptyset$ for all $x\in X$. It follows that $T^{-1}(X)=X$. Since $\mathcal{O}_2(T)$ is compact, we have $T^{-1}(x)=\{y\in X : x\in T(y)\}=\widetilde{\pi}_1(X\times\{x\}\cap\mathcal{O}_2(T))$ is compact for all $x\in X$, and thus is closed in X, where $\widetilde{\pi}_1$ is the projection map given in (2.5). Consequently, $T^{-1}(x)\in\mathcal{F}(X)$ for all $x\in X$. Moreover,

$$\mathcal{O}_2\big(T^{-1}\big)=\big\{(x,y)\in X^2 : y\in T^{-1}(x)\big\}=\big\{(x,y)\in X^2 : x\in T(y)\big\}=\iota_2(\mathcal{O}_2(T)),$$

where $\iota_2\colon X^2\to X^2$ is the isometry given by $\iota_2(x,y):=(y,x)$ in Section 2. Thus $\mathcal{O}_2(T^{-1})$ is closed in X_2 and therefore T^{-1} is a correspondence on X. $\square$

4.2. **Definition of topological pressure for correspondences.** In this subsection, we introduce a new version of topological pressure of a correspondence with respect to a continuous potential function ϕ through the (n,ϵ)-separated sets and (n,ϵ)-spanning sets. This definition naturally generalizes the definition of the topological pressure of a single-valued continuous map. When ϕ vanishes, our notion of topological pressure coincides with the notion of topological entropy for correspondences in [KT17, Definition 2.5].

For $\epsilon>0$ and a metric space (Y,ρ), we say that $E\subseteq Y$ is *ϵ-separated* if for each pair of distinct points $x,\,y\in E$, we have $\rho(x,y)\geqslant\epsilon$. We say that $F\subseteq Y$ is *ϵ-spanning* if for each $y\in Y$ there exists $x\in F$ such that $\rho(x,y)<\epsilon$. For each $\delta>0$ and each continuous function $g\colon Y\to\mathbb{R}$, set $\Delta(g,\delta):=\sup\{|g(x)-g(y)| : x,\,y\in Y \text{ and } \rho(x,y)\leqslant\delta\}$.

Let T be a correspondence on compact metric space (X,d) and $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ be a continuous function. For each $n\in\mathbb{N}$, the function $S_n\phi\colon\mathcal{O}_{n+1}(T)\to\mathbb{R}$ is given by:

$$S_n\phi(x_1,\dots,x_{n+1}):=\sum_{i=1}^{n}\phi(x_i,x_{i+1}). \tag{4.1}$$

For each $n \in \mathbb{N}$ and each $\epsilon > 0$, write

$$\begin{aligned}
s_n(T, \phi, \epsilon) &:= \sup\Big\{\sum_{\underline{x} \in E} \exp(S_n\phi(\underline{x})) : E \text{ is an } \epsilon\text{-separated subset of } \mathcal{O}_{n+1}(T)\Big\}, \\
r_n(T, \phi, \epsilon) &:= \inf\Big\{\sum_{\underline{x} \in F} \exp(S_n\phi(\underline{x})) : F \text{ is an } \epsilon\text{-spanning subset of } \mathcal{O}_{n+1}(T)\Big\}, \\
s(T, \phi, \epsilon) &:= \limsup_{n \to +\infty} \frac{1}{n} \log(s_n(T, \phi, \epsilon)), \text{ and} \\
r(T, \phi, \epsilon) &:= \limsup_{n \to +\infty} \frac{1}{n} \log(r_n(T, \phi, \epsilon)).
\end{aligned}$$

We establish some estimates for these quantities.

By choosing an orbit $\underline{x}_0 \in \mathcal{O}_{n+1}(T)$ and focusing on the ϵ-separated subset $\{\underline{x}_0\}$ of $\mathcal{O}_{n+1}(T)$, we have $s_n(T, \phi, \epsilon) = \exp(S_n\phi(\underline{x}_0)) \geqslant \exp(-n\|\phi\|_\infty)$ and

$$s(T, \phi, \epsilon) \geqslant \limsup_{n \to +\infty} \frac{1}{n} \log(\exp(-n\|\phi\|_\infty)) = -\|\phi\|_\infty.$$

For an arbitrary ϵ-spanning set $F \subseteq \mathcal{O}_{n+1}(T)$, we can choose an orbit $\underline{x}_0 \in F$, and thus we have $\sum_{\underline{x} \in F} \exp(S_n\phi(\underline{x})) \geqslant \exp(S_n\phi(\underline{x}_0)) \geqslant \exp(-n\|\phi\|_\infty)$ and

(4.2)
$$r(T, \phi, \epsilon) = \limsup_{n \to +\infty} \frac{1}{n} \log(r_n(T, \phi, \epsilon)) \geqslant \limsup_{n \to +\infty} \frac{1}{n} \log(\exp(-n\|\phi\|_\infty)) = -\|\phi\|_\infty.$$

On the other hand, both $r(T, \phi, \epsilon)$ and $s(T, \phi, \epsilon)$ may be $+\infty$.

Since for arbitrary $\epsilon_2 > \epsilon_1 > 0$, an ϵ_2-separated set is also ϵ_1-separated, we have $s_n(T, \phi, \epsilon_2) \leqslant s_n(T, \phi, \epsilon_1)$, i.e., $s_n(T, \phi, \epsilon)$ is decreasing in ϵ, and thus $s(T, \phi, \epsilon)$ is decreasing in ϵ. Similarly, an ϵ_1-spanning set is also ϵ_2-spanning, so we have $r_n(T, \phi, \epsilon_2) \leqslant r_n(T, \phi, \epsilon_1)$, i.e., $r_n(T, \phi, \epsilon)$ is decreasing in ϵ, and thus $r(T, \phi, \epsilon)$ is decreasing in ϵ. As a result, the following limits exist:

$$P_s(T, \phi) := \lim_{\epsilon \to 0^+} s(T, \phi, \epsilon), \qquad P_r(T, \phi) := \lim_{\epsilon \to 0^+} r(T, \phi, \epsilon).$$

Proposition 4.5. *Let T be a correspondence on a compact metric space (X, d) and $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$ a continuous function. Then $P_s(T, \phi) = P_r(T, \phi)$.*

Proof. For each $n \in \mathbb{N}$ and each $\epsilon > 0$, choose a maximal ϵ-separated subset $E \subseteq \mathcal{O}_{n+1}(T)$. For each $\underline{y} \in \mathcal{O}_{n+1}(T)$, since $E \cup \{\underline{y}\}$ is not ϵ-separated, there exists $\underline{x} \in E$ such that $d_{n+1}(\underline{x}, \underline{y}) < \epsilon$. Thus E is ϵ-spanning in $\mathcal{O}_{n+1}(T)$. Thereby, we have $s_n(T, \phi, \epsilon) \geqslant \sum_{\underline{x} \in E} \exp(S_n\phi(\underline{x})) \geqslant r_n(T, \phi, \epsilon)$. This implies $s(T, \phi, \epsilon) \geqslant r(T, \phi, \epsilon)$, and hence we get $P_s(T, \phi) \geqslant P_r(T, \phi)$.

For each $n \in \mathbb{N}$ and each $\epsilon > 0$, choose an arbitrary ϵ-separated set $E \subseteq \mathcal{O}_{n+1}(T)$ and an arbitrary $\frac{\epsilon}{2}$-spanning set $F \subseteq \mathcal{O}_{n+1}(T)$. For each orbit $\underline{x} \in E$, since F is $\frac{\epsilon}{2}$-spanning, there exists $\gamma(\underline{x}) \in F$ such that $d_{n+1}(\underline{x}, \gamma(\underline{x})) < \epsilon/2$. For each pair of distinct orbits $\underline{x}$, $\underline{y} \in E$, since $d_{n+1}(\underline{x}, \underline{y}) \geqslant \epsilon$, $d_{n+1}(\underline{x}, \gamma(\underline{x})) < \epsilon/2$, and $d_{n+1}(\underline{y}, \gamma(\underline{y})) <$

$\epsilon/2$, we have $\gamma(\underline{x}) \neq \gamma(\underline{y})$. Thereby, $\gamma\colon E \to F$ is injective, and thus

$$\begin{aligned}\sum_{\underline{y}\in F}\exp(S_n\phi(\underline{y})) &\geqslant \sum_{\underline{x}\in E}\exp(S_n\phi(\gamma(\underline{x}))) \\ &\geqslant \sum_{\underline{x}\in E}\exp(S_n\phi(\underline{x}) - \Delta(S_n\phi,\epsilon/2)) \\ &= \exp(-\Delta(S_n\phi,\epsilon/2))\sum_{\underline{x}\in E}\exp(S_n\phi(\underline{x})),\end{aligned}$$

where

$$\Delta(S_n\phi,\epsilon/2) \coloneqq \sup\bigl\{\bigl|S_n\phi\bigl(\underline{y}_1\bigr) - S_n\phi\bigl(\underline{y}_2\bigr)\bigr| : \underline{y}_1,\, \underline{y}_2 \in \mathcal{O}_{n+1}(T),\, d_{n+1}\bigl(\underline{y}_1,\underline{y}_2\bigr) \leqslant \epsilon/2\bigr\}.$$

Recall $d_{n+1}((x_1,\dots,x_{n+1}),(x'_1,\dots,x'_{n+1})) = \max\{d(x_i,x'_i) : 1 \leqslant i \leqslant n+1\}$. If

$$(x_1,\dots,x_{n+1}),(x'_1,\dots,x'_{n+1}) \in \mathcal{O}_n(T)$$

and

$$d_{n+1}((x_1,\dots,x_{n+1}),(x'_1,\dots,x'_{n+1})) \leqslant \epsilon/2,$$

then $d(x_i,x'_i) \leqslant \epsilon/2$ for all $i \in (n+1]$, and thereby, we have

$$\begin{aligned}|S_n\phi(x_1,\dots,x_{n+1}) - S_n\phi(x'_1,\dots,x'_{n+1})| &= \left|\sum_{i=1}^{n}(\phi(x_i,x_{i+1}) - \phi(x'_i,x'_{i+1}))\right| \\ &\leqslant \sum_{i=1}^{n}|\phi(x_i,x_{i+1}) - \phi(x'_i,x'_{i+1})| \\ &\leqslant n\Delta(\phi,\epsilon/2).\end{aligned}$$

This implies that $\Delta(S_n\phi,\epsilon/2) \leqslant n\Delta(\phi,\epsilon/2)$. As a result, we get

$$\sum_{\underline{y}\in F}\exp(S_n\phi(\underline{y})) \geqslant \exp(-n\Delta(\phi,\epsilon/2))\sum_{\underline{x}\in E}\exp(S_n\phi(\underline{x})).$$

Since E and F are chosen arbitrarily, we have

$$r_n(T,\phi,\epsilon/2) \geqslant \exp(-n\Delta(\phi,\epsilon/2))s_n(T,\phi,\epsilon).$$

Thus,

$$r(T,\phi,\epsilon/2) \geqslant s(T,\phi,\epsilon) - \Delta(\phi,\epsilon/2). \tag{4.3}$$

Since X is compact and ϕ is continuous, ϕ is uniformly continuous, i.e., for an arbitrary $\delta > 0$, there exists $\lambda > 0$ such that $\Delta(\phi,\lambda) < \delta$. Thus we have $\lim_{\epsilon\to 0^+}\Delta(\phi,\epsilon/2) = 0$. Consequently, by taking $\epsilon \to 0^+$ in (4.3), we get $P_r(T,\phi) \geqslant P_s(T,\phi)$.

Therefore, we conclude that $P_r(T,\phi) = P_s(T,\phi)$. □

Definition 4.6 (Topological pressure)**.** Let T be a correspondence on a compact metric space (X,d) and $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$ be a continuous function. The *topological pressure* $P(T,\phi)$ is defined as

$$P(T,\phi) \coloneqq P_s(T,\phi) = P_r(T,\phi).$$

In other words,

$$
\begin{aligned}
P(T,\phi) &= \lim_{\epsilon\to0^+}\limsup_{n\to+\infty}\frac{1}{n}\log\biggl(\sup_{E_n(\epsilon)}\sum_{\underline{x}\in E_n(\epsilon)}\exp(S_n\phi(\underline{x}))\biggr)\\
&= \lim_{\epsilon\to0^+}\limsup_{n\to+\infty}\frac{1}{n}\log\biggl(\inf_{F_n(\epsilon)}\sum_{\underline{x}\in F_n(\epsilon)}\exp(S_n\phi(\underline{x}))\biggr),
\end{aligned}
\tag{4.4}
$$

where $E_n(\epsilon)$ ranges over all ϵ-separated subsets of $(\mathcal{O}_{n+1}(T),d_{n+1})$ and $F_n(\epsilon)$ ranges over all ϵ-spanning subsets of $(\mathcal{O}_{n+1}(T),d_{n+1})$.

In particular, if $\phi\equiv0$, we call $P(T,\mathbf{0})$ the *topological entropy* of T and denote it by $h(T)$.

Remark 4.7. Recall from (4.2) that $r(T,\phi,\epsilon)\geqslant-\|\phi\|_\infty$. This implies

$$
P(T,\phi)=P_r(T,\phi)=\lim_{\epsilon\to0^+}r(T,\phi,\epsilon)\geqslant-\|\phi\|_\infty>-\infty.
\tag{4.5}
$$

Let $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ be a continuous function. Denote by $\overline{\phi}\colon\mathcal{O}_2(T^{-1})$ the conjugate function given by $\overline{\phi}(x,y):=\phi(y,x)$ for all $(x,y)\in\mathcal{O}_2(T^{-1})$. The definition of topological pressure for correspondences can lead to the following proposition.

Proposition 4.8. *Let T be a correspondence on a compact metric space X satisfying $T(X)=X$ and $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ be a continuous function. Then*

$$
P(T,\phi)=P\bigl(T^{-1},\overline{\phi}\bigr).
$$

Proof. For each $n\in\mathbb{N}$ and $\underline{x}=(x_1,\dots,x_n)\in X^n$, $\underline{x}=(x_1,x_2,\dots,x_n)\in\mathcal{O}_n(T)$ if and only if $\iota_n(\underline{x})=(x_n,x_{n-1},\dots,x_1)\in\mathcal{O}_n(T^{-1})$. Consequently, the isometry ι_n sends $\mathcal{O}_n(T)$ onto $\mathcal{O}_n(T^{-1})$. From (4.1) we can see that $S_n\phi(\underline{x})=S_n\overline{\phi}(\iota_{n+1}(\underline{x}))$ holds for all $\underline{x}\in\mathcal{O}_{n+1}(T)$. Since ι_{n+1} is an isometry, for each $\epsilon>0$, $E_n(\epsilon)\subseteq\mathcal{O}_{n+1}(T)$ is ϵ-separated if and only if $\iota_{n+1}(E_n(\epsilon))\subseteq\mathcal{O}_{n+1}(T^{-1})$, so by (4.4) we conclude $P(T,\phi)=P\bigl(T^{-1},\overline{\phi}\bigr)$. □

4.3. **A characterization of the topological pressure.** We will prove in this subsection that our topological pressure of a correspondence T with respect to a continuous potential function ϕ is equal to $P\bigl(\sigma,\widetilde{\phi}\bigr)$, the topological pressure of the shift map σ on the orbit space $\mathcal{O}_\omega(T)$ with respect to the potential function $\widetilde{\phi}$ (given in (2.4)) induced by ϕ (see Theorem 4.9 for the precise statement).

Let T be a correspondence on a compact metric space (X,d). We consider a dynamical system $(\mathcal{O}_\omega(T),\sigma)$, where $\mathcal{O}_\omega(T)$ is equipped with the metric d_ω and $\sigma\colon\mathcal{O}_\omega(T)\to\mathcal{O}_\omega(T)$ is the shift map given by

$$
\sigma(x_1,x_2,x_3,\dots):=(x_2,x_3,\dots)
\tag{4.6}
$$

for all $(x_1,x_2,\dots)\in X^\omega$.

Theorem 4.9. *Let T be a correspondence on a compact metric space (X,d), $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ be a continuous function, and σ be the shift map on $\mathcal{O}_\omega(T)$ given above. Then we have*

$$
P(T,\phi)=P\bigl(\sigma,\widetilde{\phi}\bigr),
$$

where $P(\sigma, \widetilde{\phi})$ refers to the topological pressure of the dynamical system $(\mathcal{O}_\omega(T), \sigma)$ with the potential function $\widetilde{\phi}$ given in (2.4).

Proof. We divide this proof into two steps. Let $\epsilon > 0$ be arbitrary and denote $\tilde{\epsilon} := \epsilon/(1+\epsilon)$.

Step 1. We show $P(T, \phi) \leqslant P(\sigma, \widetilde{\phi})$.

Let $n \in \mathbb{N}$. For every $(x_1, \dots, x_{n+1}) \in \mathcal{O}_{n+1}(T)$, we choose $x_{n+2}, x_{n+3}, \dots \in X$ such that $(x_1, \dots, x_{n+1}, x_{n+2}, \dots) \in \mathcal{O}_\omega(T)$. Denote by $\tau \colon E_n(\epsilon) \to \mathcal{O}_\omega(T)$ the map that extends each $(x_1, \dots, x_{n+1}) \in E_n(\epsilon)$ to the orbit $(x_1, \dots, x_{n+1}, x_{n+2}, \dots) \in \mathcal{O}_\omega(T)$. The map τ is injective.

Fix arbitrary $n \in \mathbb{N}$ and an ϵ-separated subset $E_n(\epsilon)$ of $(\mathcal{O}_{n+1}(T), d_{n+1})$.

For an arbitrary pair of distinct orbits $(x_1, \dots, x_{n+1}), (y_1, \dots, y_{n+1}) \in E_n(\epsilon)$, we have

$$\epsilon \leqslant d_{n+1}((x_1, \dots, x_{n+1}), (y_1, \dots, y_{n+1})) = \max\{d(x_i, y_i) : 1 \leqslant i \leqslant n+1\}.$$

Choose $k \in (n+1]$ such that $d(x_k, y_k) \geqslant \epsilon$, then

$$\begin{aligned} &d_\omega\bigl(\sigma^{k-1}(\tau(x_1, \dots, x_{n+1})), \sigma^{k-1}(\tau(y_1, \dots, y_{n+1}))\bigr) \\ &\qquad = d_\omega((x_k, \dots, x_{n+1}, \dots), (y_k, \dots, y_{n+1}, \dots)) \geqslant \frac{1}{2} \cdot \frac{d(x_k, y_k)}{1 + d(x_k, y_k)} \geqslant \frac{\tilde{\epsilon}}{2}. \end{aligned}$$

This implies that $\tau(E_n(\epsilon))$ is $(n+1, \tilde{\epsilon}/2)$-separated in the dynamical system $(\mathcal{O}_\omega(T), \sigma)$.

Since

$$\sum_{\underline{x} \in E_n(\epsilon)} e^{S_n\phi(\underline{x})} = \sum_{(x_1, \dots, x_{n+1}, \dots) \in \tau(E_n(\epsilon))} e^{\sum_{j=1}^{n} \phi(x_j, x_{j+1})} = \sum_{\underline{x} \in \tau(E_n(\epsilon))} e^{\sum_{j=1}^{n} \widetilde{\phi}(\sigma^{j-1}(\underline{x}))},$$

and the ϵ-separated set $E_n(\epsilon)$ is chosen arbitrarily, we have

$$\begin{aligned} &\sup\Bigl\{ \sum_{\underline{x} \in E_n(\epsilon)} e^{S_n\phi(\underline{x})} : E_n(\epsilon) \text{ is } \epsilon\text{-separated in } \mathcal{O}_{n+1}(T) \Bigr\} \\ &\qquad \leqslant \sup\Biggl\{ \sum_{\underline{x} \in \widetilde{E}_n(\frac{\tilde{\epsilon}}{2})} e^{\sum_{j=0}^{n-1} \widetilde{\phi}(\sigma^j(\underline{x}))} : \widetilde{E}_n(\tilde{\epsilon}/2) \text{ is } (n+1, \tilde{\epsilon}/2)\text{-separated in } (\mathcal{O}_\omega(T), \sigma) \Biggr\}. \end{aligned}$$

This implies that

$$\begin{aligned} P(T, \phi) &= \lim_{\epsilon \to 0^+} \limsup_{n \to +\infty} \frac{1}{n} \log\Biggl(\sup_{E_n(\epsilon)} \sum_{\underline{x} \in E_n(\epsilon)} \exp(S_n\phi(\underline{x})) \Biggr) \\ &\leqslant \lim_{\epsilon \to 0^+} \limsup_{n \to +\infty} \frac{1}{n} \log\Biggl(\sup_{\widetilde{E}_n(\tilde{\epsilon}/2)} \sum_{\underline{x} \in \widetilde{E}_n(\tilde{\epsilon}/2)} \exp\Biggl(\sum_{j=0}^{n-1} \widetilde{\phi}(\sigma^j(\underline{x})) \Biggr) \Biggr) \\ &= P(\sigma, \widetilde{\phi}), \end{aligned}$$

where $E_n(\epsilon)$ ranges over all ϵ-separated subsets of $\mathcal{O}_{n+1}(T)$, $\widetilde{E}_n(\tilde{\epsilon}/2)$ ranges over all $(n+1, \tilde{\epsilon}/2)$-separated sets of the dynamical system $(\mathcal{O}_\omega(T), \sigma)$, and the last equality holds because $\lim_{\epsilon \to 0^+} \tilde{\epsilon}/2 = 0$. Hence we conclude $P(T, \phi) \leqslant P(\sigma, \widetilde{\phi})$.

Step 2. We show $P(T,\phi) \geqslant P\big(\sigma, \widetilde{\phi}\big)$.

Fix arbitrary $n \in \mathbb{N}$ and an ϵ-spanning subset $F_n(\epsilon)$ of $(\mathcal{O}_{n+1}(T), d_{n+1})$.

For every $(x_1', \dots, x_{n+1}', x_{n+2}', \dots) \in \mathcal{O}_\omega(T)$, since $(x_1', \dots, x_{n+1}')$ is in $\mathcal{O}_{n+1}(T)$, we can choose an orbit $(x_1, \dots, x_{n+1}) \in F_n(\epsilon)$ such that $d_{n+1}((x_1', \dots, x_{n+1}'), (x_1, \dots, x_{n+1})) < \epsilon$, i.e., $d(x_k, x_k') < \epsilon$ for all $k \in (n+1]$. For each $k \in [n - \lfloor \sqrt{n} \rfloor - 1]\}$, we have

$$\begin{aligned}
&d_\omega\big(\sigma^k(x_1', \dots, x_{n+1}', x_{n+2}', \dots), \sigma^k(\tau(x_1, \dots, x_{n+1}))\big) \\
&\quad = d_\omega((x_{k+1}', \dots, x_{n+1}', x_{n+2}', \dots), (x_{k+1}, \dots, x_{n+1}, x_{n+2}, \dots)) \\
&\quad = \sum_{j=1}^{+\infty} \frac{1}{2^j} \frac{d(x_{k+j}', x_{k+j})}{1 + d(x_{k+j}', x_{k+j})} \\
&\quad < \sum_{j=1}^{n-k} \frac{1}{2^j} \tilde{\epsilon} + \sum_{j=n-k+1}^{+\infty} \frac{1}{2^j} \\
&\quad < \tilde{\epsilon} + 2^{-(n-k)} \\
&\quad \leqslant \tilde{\epsilon} + 2^{-\lfloor \sqrt{n} \rfloor - 1} \\
&\quad \leqslant \tilde{\epsilon} + 2^{-\sqrt{n}}.
\end{aligned}$$

Hence $\tau(F_n(\epsilon))$ is an $\big(n - \lfloor \sqrt{n} \rfloor, \tilde{\epsilon} + 2^{-\sqrt{n}}\big)$-spanning set in the dynamical system $(\mathcal{O}_\omega(T), \sigma)$.

Recall $\|\phi\|_\infty = \sup\{|\phi(x_1, x_2)| : (x_1, x_2) \in \mathcal{O}_2(T)\}$. We have

$$\begin{aligned}
\sum_{\underline{x} \in F_n(\epsilon)} e^{S_n \phi(\underline{x})} &= \sum_{(x_1, \dots, x_{n+1}) \in F_n(\epsilon)} e^{\sum_{j=1}^{n} \phi(x_j, x_{j+1})} \\
&= \sum_{\underline{x} \in \tau(F_n(\epsilon))} e^{\sum_{j=0}^{n-1} \widetilde{\phi}(\sigma^j(\underline{x}))} \\
&\geqslant \sum_{\underline{x} \in \tau(F_n(\epsilon))} e^{\sum_{j=0}^{n - \lfloor \sqrt{n} \rfloor - 1} \widetilde{\phi}(\sigma^j(\underline{x})) - \lfloor \sqrt{n} \rfloor \|\phi\|_\infty} \\
&\geqslant e^{-\sqrt{n} \|\phi\|_\infty} \sum_{\underline{x} \in \tau(F_n(\epsilon))} e^{\sum_{j=0}^{n - \lfloor \sqrt{n} \rfloor - 1} \widetilde{\phi}(\sigma^j(\underline{x}))}.
\end{aligned} \tag{4.7}$$

For each $\delta > 0$ and each $m \in \mathbb{N}$, write

$$\alpha(m, \delta) := \inf \left\{ \sum_{\underline{x} \in \widetilde{F}_m(\delta)} e^{\sum_{j=0}^{m-1} \widetilde{\phi}(\sigma^j(\underline{x}))} : \widetilde{F}_m(\delta) \text{ is } (m, \delta)\text{-spanning in } (\mathcal{O}_\omega(T), \sigma) \right\},$$

then we have $P\big(\sigma, \widetilde{\phi}\big) = \lim_{\delta \to 0^+} \limsup_{m \to +\infty} \frac{1}{m} \log(\alpha(m, \delta))$.

Since for $\delta_2 > \delta_1 > 0$, an (m, δ_1)-spanning set is also (m, δ_2)-spanning, we can see that $\alpha(m, \delta)$ is decreasing in δ for each $m \in \mathbb{N}$.

Since $\tau(F_n(\epsilon))$ is an $\bigl(n-\lfloor\sqrt{n}\rfloor, \tilde{\epsilon}+2^{-\sqrt{n}}\bigr)$-spanning set in $(\mathcal{O}_\omega(T),\sigma)$, (4.7) implies

$$\sum_{\underline{x}\in F_n(\epsilon)} e^{S_n\phi(\underline{x})} \geqslant e^{-\sqrt{n}\|\phi\|_\infty}\alpha\bigl(n-\lfloor\sqrt{n}\rfloor, \tilde{\epsilon}+2^{-\sqrt{n}}\bigr).$$

Let $F_n(\epsilon)$ range over all ϵ-spanning subsets of $(\mathcal{O}_{n+1}(T), d_{n+1})$ and take an infimum in the inequality above, we get

$$r_n(T,\phi,\epsilon) = \inf_{F_n(\epsilon)} \sum_{\underline{x}\in F_n(\epsilon)} e^{S_n\phi(\underline{x})} \geqslant e^{-\sqrt{n}\|\phi\|_\infty}\alpha\bigl(n-\lfloor\sqrt{n}\rfloor, \tilde{\epsilon}+2^{-\sqrt{n}}\bigr).$$

This implies

$$\begin{aligned}
P(T,\phi) &= \lim_{\epsilon\to 0^+}\limsup_{n\to+\infty}\frac{1}{n}\log(r_n(T,\phi,\epsilon))\\
&\geqslant \lim_{\epsilon\to 0^+}\limsup_{n\to+\infty}\bigl(n^{-1}\log\bigl(\alpha\bigl(n-\lfloor\sqrt{n}\rfloor, \tilde{\epsilon}+2^{-\sqrt{n}}\bigr)\bigr)-\|\phi\|_\infty n^{-1/2}\bigr)\\
&\geqslant \lim_{\epsilon\to 0^+}\limsup_{n\to+\infty}(n-\lfloor\sqrt{n}\rfloor)^{-1}\log\bigl(\alpha\bigl(n-\lfloor\sqrt{n}\rfloor, \tilde{\epsilon}+\epsilon\bigr)\bigr)\\
&= P\bigl(\sigma,\widetilde{\phi}\bigr),
\end{aligned}$$

where the last equality holds because $n-\lfloor\sqrt{n}\rfloor$ ranges over all positive integers as n ranges over all positive integers and $\lim_{\epsilon\to 0^+}\tilde{\epsilon}+\epsilon=0$. □

5. Measure-theoretic entropy of transition probability kernels

This section is devoted to introduce and discuss the measure-theoretic entropy of transition probability kernels. We discuss some basic notions and properties about transition probability kernels in Subsections 5.1 and 5.2 and introduce the measure-theoretic entropy of transition probability kernels in Subsection 5.3. Finally, in Subsection 5.4, we define another shift map for a correspondence and relate the measure-theoretic entropy of this shift map to that of the correspondence, see Theorem 5.24, the main theorem in this subsection.

5.1. **Basic properties of transition probability kernels.** In this subsection, we recall the definition of transition probability kernels (see for example, [MT12, Section 3.4.1]), which are also called Markovian transition kernels (see for example, [Ga16, Section 6.1]). Intuitively, a transition probability kernel on a measurable space X assigns each $x\in X$ a probability measure on X. We fix some notations in probability to prepare for the next subsection. Moreover, we recall how a transition probability kernel pushes a function forward (Definition 5.4) and pulls a measure back (Definition 5.6), how to compose transition probability kernels (Definition 5.11), and we show some properties of these actions and compositions such as the association law (see for example, [Re84, Section 1.1]).

Definition 5.1 (Transition probability kernels)**.** Let $(X,\mathscr{M}(X))$ and $(Y,\mathscr{M}(Y))$ be measurable spaces, where X and Y are sets and $\mathscr{M}(X)$ and $\mathscr{M}(Y)$ are σ-algebras on X and Y, respectively. A *transition probability kernel from Y to X* is a map $\mathcal{Q}\colon Y\times\mathscr{M}(X)\to[0,1]$ satisfying the following two properties:

(i) For every $y \in Y$, the map $\mathscr{M}(X) \ni A \mapsto \mathcal{Q}(y, A)$ is a probability measure on the measurable space $(X, \mathscr{M}(X))$.

(ii) For every $A \in \mathscr{M}(X)$, the map $Y \ni y \mapsto \mathcal{Q}(y, A)$ is $\mathscr{M}(Y)$-measurable.

With the notations above, for every $y \in Y$, denote by $\mathcal{Q}_y$ the probability measure that assigns a measurable set $A \in \mathscr{M}(X)$ the value $\mathcal{Q}(y, A)$. In other words,

$$\mathcal{Q}_y(A) \coloneqq \mathcal{Q}(y, A).$$

Moreover, if $Y = X$, we also call a transition probability kernel from Y to X a *transition probability kernel on* X.

Remark 5.2. As an interpretation of Definition 5.1, we can think of a transition probability kernel $\mathcal{Q}$ from Y to X as a stochastic process that transmits each point $y \in Y$ "randomly" to a point $x \in X$ with distribution $\mathcal{Q}_y$.

Definition 5.3. Let T be a correspondence on a compact metric space X and $\mathcal{Q}$ be a transition probability kernel on X. We say that $\mathcal{Q}$ is *supported by* T if the measure $\mathcal{Q}_x$ is supported on the closed set $T(x)$ (i.e., $\operatorname{supp} \mathcal{Q}_x \subseteq T(x)$) for every $x \in X$.

Given a measurable map $F \colon Y \to X$, we can pull back a function $f \colon X \to \mathbb{R}$ (with the resulting pullback $F^* \colon f \mapsto f \circ F$) or push forward a probability measure μ on Y (with the resulting pushforward $F_* \colon \mu \mapsto \mu \circ F^{-1}$). For a Markov chain with the state space $X = (d]$ and a transition matrix P, a distribution $p = (p_1, \dots p_d)$ on X becomes pP after one step of the Markov process. Definitions B.1 and C.1 show that transition probability kernels generalize measurable maps and transition matrices. Their actions on functions and measures are standard, see for example, [Ga16, Section 6.1] and [MT12, Section 3.4.2]. We recall them below.

Definition 5.4. Let $(X, \mathscr{M}(X))$ and $(Y, \mathscr{M}(Y))$ be measurable spaces, $f \in B(X, \mathbb{R})$ be a bounded measurable function, and $\mathcal{Q}$ be a transition probability kernel from Y to X. The *pullback* function $\mathcal{Q}f \colon Y \to \mathbb{R}$ of f by $\mathcal{Q}$ is given by

$$\mathcal{Q}f(y) \coloneqq \int_X f(x) \, \mathrm{d}\mathcal{Q}_y(x).$$

As an operator acting on bounded measurable functions, $\mathcal{Q}$ is linear, i.e.,

$$\mathcal{Q}(af + g) = a\mathcal{Q}f + \mathcal{Q}g \tag{5.1}$$

for all $f, g \in B(X, \mathbb{R})$, and $a \in \mathbb{R}$. Moreover, $\mathcal{Q}$ as an operator is continuous in the sense of the following lemma.

Lemma 5.5. *Let $(X, \mathscr{M}(X))$ and $(Y, \mathscr{M}(Y))$ be measurable spaces, $\mathcal{Q}$ be a transition probability kernel from Y to X, and $f \in B(X, \mathbb{R})$. If a sequence of uniformly bounded measurable functions $f_n \in B(X, \mathbb{R})$, $n \in \mathbb{N}$, converges pointwise to $f \colon X \to \mathbb{R}$ as $n \to +\infty$, then $\mathcal{Q}f_n$ converges pointwise to $\mathcal{Q}f$ as $n \to +\infty$. Moreover, $\mathcal{Q}f$ is measurable and $\|\mathcal{Q}f\|_\infty \leqslant \|f\|_\infty$.*

We include a proof of Lemma 5.5 in Appendix A.1.

Definition 5.6. Let $(X, \mathscr{M}(X))$ and $(Y, \mathscr{M}(Y))$ be measurable spaces, $\mu \in \mathcal{P}(Y)$ be a probability measure on Y, and $\mathcal{Q}$ be a transition probability kernel from Y to X. The *pushforward* probability measure $\mu\mathcal{Q}$ on X of μ by $\mathcal{Q}$ is given by

$$(\mu\mathcal{Q})(A) := \int_Y \mathcal{Q}(y, A) \,\mathrm{d}\mu(y)$$

for all $A \in \mathscr{M}(X)$.

We shall check that $\mu\mathcal{Q}$ is a probability measure on X:

(i) Since $\mathcal{Q}(y, A) \in [0, 1]$, we have $(\mu\mathcal{Q})(A) = \int_Y \mathcal{Q}(y, A) \,\mathrm{d}\mu(y) \geq 0$ for all $A \in \mathscr{M}(X)$.

(ii) $(\mu\mathcal{Q})(X) = \int_Y \mathcal{Q}(y, X) \,\mathrm{d}\mu(y) = \int_Y 1 \,\mathrm{d}\mu(y) = 1$.

(iii) If $A_1, A_2, \dots \in \mathscr{M}(X)$ are mutually disjoint, then

$$\begin{aligned}(\mu\mathcal{Q})\biggl(\bigcup_{n=1}^{+\infty} A_n\biggr) &= \int_Y \mathcal{Q}\biggl(y, \bigcup_{n=1}^{+\infty} A_n\biggr) \,\mathrm{d}\mu(y) \\ &= \int_Y \sum_{n=1}^{+\infty} \mathcal{Q}(y, A_n) \,\mathrm{d}\mu(y) \\ &= \sum_{n=1}^{+\infty} \int_Y \mathcal{Q}(y, A_n) \,\mathrm{d}\mu(y) \\ &= \sum_{n=1}^{+\infty} (\mu\mathcal{Q})(A_n).\end{aligned}$$

Hence we conclude that $\mu\mathcal{Q}$ given in Definition 5.6 is indeed a probability measure on X.

Remark 5.7. Recall from Remark 5.2 that $\mathcal{Q}$ can be considered as a stochastic process. Under this perspective, for an initial distribution μ on Y, if we set a random point $y \in Y$ with the distribution μ and transmit it to a point $x \in X$ through the stochastic process $\mathcal{Q}$, then the resulting distribution of $x \in X$ is $\mu\mathcal{Q}$.

Remark 5.8. Notice that $\delta_y\mathcal{Q} = \mathcal{Q}_y$ holds for all $y \in Y$, where δ_y refers to the Dirac measure at the point y. This is because for each $A \in \mathscr{M}(X)$ and each $y \in Y$, we have

$$(\delta_y\mathcal{Q})(A) = \int_Y \mathcal{Q}(z, A) \,\mathrm{d}\delta_y(z) = \mathcal{Q}(y, A) = \mathcal{Q}_y(A).$$

Proposition 5.9. *Let $(X, \mathscr{M}(X))$ and $(Y, \mathscr{M}(Y))$ be measurable spaces, $\mu \in \mathcal{P}(Y)$, $\mathcal{Q}$ be a transition probability kernel from Y to X, and $f \in B(X, \mathbb{R})$. We have*

$$\int_Y \mathcal{Q}f \,\mathrm{d}\mu = \int_X f \,\mathrm{d}(\mu\mathcal{Q}). \tag{5.2}$$

Proposition 5.9 is standard for any specialist. But we include a proof in Appendix A.1 since we cannot locate a precise reference.

Definition 5.10. Let $(X, \mathscr{M}(X))$ be a measurable space and $\mathcal{Q}$ be a transition probability kernel on X. We say that a probability measure μ on X is *$\mathcal{Q}$-invariant* if $\mu\mathcal{Q} = \mu$. Denote by $\mathcal{M}(X, \mathcal{Q})$ the set of all $\mathcal{Q}$-invariant probability measures on X.

The following notion comes from [Ga16, Definition 6.1].

Definition 5.11. Let $(X, \mathscr{M}(X))$, $(Y, \mathscr{M}(Y))$, and $(Z, \mathscr{M}(Z))$ be measurable spaces, $\mathcal{Q}$ be a transition probability kernel from Y to X, and $\mathcal{Q}'$ be a transition probability kernel from Z to Y. The transition probability kernel $\mathcal{Q}'\mathcal{Q}$ from Z to X is given by

$$(\mathcal{Q}'\mathcal{Q})(z, A) := (\mathcal{Q}'_z\mathcal{Q})(A)$$

for all $z \in Z$ and $A \in \mathscr{M}(X)$, where $\mathcal{Q}'_z\mathcal{Q}$ is a probability measure on X defined in Definition 5.6.

We check in Appendix A.1 that $\mathcal{Q}'\mathcal{Q}$ is indeed a transition probability kernel from Z to X, see Lemma A.1.

Remark 5.12. Recall from Remark 5.2 that $\mathcal{Q}$ and $\mathcal{Q}'$ can be considered as stochastic processes. Under this perspective, $\mathcal{Q}'\mathcal{Q}$ refers to the composition process of the two processes, i.e., the stochastic process that transmits each point $z \in Z$ randomly to an intermediate point $y \in Y$ through the process $\mathcal{Q}'$, and next transmits y randomly to a point $z \in Z$ through the process $\mathcal{Q}$.

Lemma 5.13. *Let $(X, \mathscr{M}(X))$, $(Y, \mathscr{M}(Y))$, and $(Z, \mathscr{M}(Z))$ be measurable spaces, $\mathcal{Q}$ be a transition probability kernel from Y to X, $\mathcal{Q}'$ be a transition probability kernel from Z to Y, and $\mu \in \mathcal{P}(Z)$. We have the law of association:*

$$\mu(\mathcal{Q}'\mathcal{Q}) = (\mu\mathcal{Q}')\mathcal{Q}.$$

We include a proof of Lemma 5.13 in Appendix A.1.

For measurable spaces X_1, X_2, and X_3, transition probability kernels $\mathcal{Q}$ from X_3 to X_2 and $\mathcal{Q}'$ from X_2 to X_1, and $\mu \in \mathcal{P}(X_3)$, Lemma 5.13 ensures that we can write $\mu\mathcal{Q}\mathcal{Q}'$ without parentheses.

5.2. **Transition probability kernels $\mathcal{Q}^{[n]}$ and $\mathcal{Q}^{\omega}$.** In this subsection, we give the definition of transition probability kernels $\mathcal{Q}^{[n]}$ ($n \in \mathbb{N}_0$) and $\mathcal{Q}^{\omega}$, which are repeatedly used in the sequel. First, we discuss some intuition.

Recall from Remark 5.2 that a transition probability kernel $\mathcal{Q}$ on a measurable space X can be considered as a stochastic process. For $n \in \mathbb{N}$, if we iterate this stochastic process n times, then from an initial point $x_0 \in X$, we can get an orbit $(x_0, x_1, \dots, x_n) \in X^{n+1}$. The stochastic process that transmits each $x_0 \in X$ randomly to an orbit $(x_0, x_1, \dots, x_n)$ with the distribution that comes from the n-fold iterate of $\mathcal{Q}$ is formed by the transition probability kernel $\mathcal{Q}^{[n]}$. From this intuition, we notice that $\mathcal{Q}^{[n]}$ is a transition probability kernel from X to X^{n+1}.

If we iterate this stochastic process infinitely many times, then from an initial point $x_0 \in X$, we can get an infinite forward orbit $(x_0, x_1, \dots) \in X^{\omega}$. The stochastic process that transmits each $x_0 \in X$ randomly to an orbit $(x_0, x_1, \dots)$ with the distribution that comes from the infinite iterates of $\mathcal{Q}$ is formed by the transition probability $\mathcal{Q}^{\omega}$.

We notice from this intuition that $\mathcal{Q}^\omega$ is a transition probability kernel from X to X^ω.

Another important thing in this subsection is the measure $\mu\mathcal{Q}^\omega$, where μ is a probability measure on X and $\mathcal{Q}$ is a transition probability kernel on X. The probability measure μ as an initial distribution and the transition probability kernel $\mathcal{Q}$ as a stochastic process can generate a Markov process. The distribution of the forward infinite orbit in this Markov process is denoted by $\mu\mathcal{Q}^\omega$.

We discuss the above notations more carefully below.

We always use $(X, \mathscr{M}(X))$ to denote a measurable space where X is a set and $\mathscr{M}(X)$ is a σ-algebra on X.

Consider $m \in \mathbb{N} \setminus \{1\}$, $n_1, n_2, \dots, n_{m-1} \in \mathbb{N}$, $n_m \in \widehat{\mathbb{N}}$, and a subset $B_i \subseteq X^{n_i}$ for each $i \in (m]$. Set $N_0 \coloneqq 0$ and $N_i \coloneqq \sum_{j=1}^{i} n_j \in \widehat{\mathbb{N}}$ for each $i \in (m]$. The notation $B_1 \times \cdots \times B_m$ refers to a subset of X^{N_m} given below:

$$\begin{aligned} B_1 \times \cdots \times B_m \coloneqq \bigl\{ &(x_1, x_2, \dots) \in X^{N_m} : \\ &(x_{N_{k-1}+1}, x_{N_{k-1}+2}, \dots, x_{N_{k-1}+n_k}) \in B_k \text{ for each } k \in (m] \bigr\}, \end{aligned}$$

where $(x_{N_{m-1}+1}, x_{N_{m-1}+2}, \dots, x_{N_{m-1}+n_m})$ means $(x_{N_{m-1}+1}, x_{N_{m-1}+2}, \dots)$ if $n_m = \omega$.

For each $n \in \mathbb{N}$, denote by $\mathscr{M}(X^n)$ the σ-algebra on X^n generated by $\bigcup_{i=0}^{n-1} \{X^i \times A \times X^{n-1-i} : A \in \mathscr{M}(X)\}$. Denote by $\mathscr{M}(X^\omega)$ the σ-algebra on $X^\omega \coloneqq \{(x_1, x_2, \dots) : x_1, x_2, \dots \in X\}$ generated by $\bigcup_{i=0}^{+\infty} \{X^i \times A \times X^\omega : A \in \mathscr{M}(X)\}$.

For each $A_{n+1} \subseteq X^{n+1}$ and each $(x_1, \dots, x_n) \in X^n$, write

$$\pi_{n+1}(x_1, \dots, x_n; A_{n+1}) \coloneqq \{x_{n+1} \in X : (x_1, \dots, x_n, x_{n+1}) \in A_{n+1}\}. \tag{5.3}$$

Definition 5.14. Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, where X is a set and $\mathscr{M}(X)$ is a σ-algebra on X. Define the transition probability kernel $\mathcal{Q}^{[n]}$ from X to X^{n+1} inductively on $n \in \mathbb{N}_0$ as follows:

First, $\mathcal{Q}^{[0]} \coloneqq \widehat{\mathrm{id}_X}$, where $\widehat{\mathrm{id}_X}$ is a transition probability kernel given by $\widehat{\mathrm{id}_X}(x, A) \coloneqq \mathbb{1}_A(x)$ for all $x \in X$ and $A \in \mathscr{M}(X)$. This means $\mathcal{Q}_x^{[0]} = \delta_x$, the Dirac measure at $x \in X$, for all $x \in X$. If $\mathcal{Q}^{[n-1]}$ has been defined for some $n \in \mathbb{N}$, we define $\mathcal{Q}^{[n]}$ as:

$$\mathcal{Q}^{[n]}(x, A_{n+1}) \coloneqq \int_{(x_1, \dots, x_n) \in X^n} \mathcal{Q}(x_n, \pi_{n+1}(x_1, \dots, x_n; A_{n+1})) \,\mathrm{d}\mathcal{Q}_x^{[n-1]}(x_1, \dots, x_n) \tag{5.4}$$

for all $x \in X$ and $A_{n+1} \in \mathscr{M}\bigl(X^{n+1}\bigr)$.

Remark 5.15. Recall from Section 2 that $[n] = \{0, 1, \dots, n\}$. Our notation $\mathcal{Q}^{[n]}$ is compatible with the notation in the proof of [Ga16, Corollary 6.4], in which its author uses a finite subset $U = \{t_1, \dots, t_k\}$ of $\mathbb{R}_+$ as a superscript to indicate the distribution of following orbits of a continuous-time Markov process: $(x_{t_1}, \dots, x_{t_k})$, where x_t denotes the random variable at the time t in the Markov process. In addition, see [MT12, Section 3.4.1] for a related notion P_x^n.

We check in Appendix A.1 that $\mathcal{Q}^{[n]}$ defined above is indeed a transition probability kernel for each $n \in \mathbb{N}$, see Lemma A.2.

Remark 5.16. Recall from Remark 5.2 that $\mathcal{Q}$ can be considered as a stochastic process and we can iterate it n times. If we start at a point $x \in X$, then $\mathcal{Q}_x^{[n]}$ is the distribution of orbits of the n-step stochastic process, containing $n+1$ entries from the 0-th step to the n-th step.

Lemma 5.17. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, m, $n \in \mathbb{N}_0$, and $A \in \mathscr{M}(X^{n+1})$. For each $x \in X$, we have*

$$\mathcal{Q}^{[n]}(x, A) = \mathcal{Q}^{[n+m]}(x, A \times X^m). \tag{5.5}$$

We include a proof of Lemma 5.17 in Appendix A.1.

Applying the Kolmogorov extension theorem (see Lemma A.3 for details), we get the following definition.

Definition 5.18. Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$. Define the transition probability kernel $\mathcal{Q}^\omega$ from X to X^ω as the unique transition probability kernel from X to X^ω with the property that for each $x \in X$, each $n \in \mathbb{N}_0$, and each measurable set $A \in \mathscr{M}(X^{n+1})$, the following equality holds:

$$\mathcal{Q}^\omega(x, A \times X^\omega) = \mathcal{Q}^{[n]}(x, A). \tag{5.6}$$

Remark 5.19. For each $\mu \in \mathcal{P}(X)$, Definition 5.6 and (5.6) imply that

$$(\mu\mathcal{Q}^\omega)(A \times X^\omega) = \big(\mu\mathcal{Q}^{[n]}\big)(A). \tag{5.7}$$

5.3. **Definition of measure-theoretic entropy for transition probability kernels.** In this subsection, we introduce the measure-theoretic entropy for transition probability kernels (Definition 5.22). As we can see in Definition B.1, transition probability kernels generalize measurable maps. Our definition of the measure-theoretic entropy of transition probability kernels uses the entropy of partitions and generalizes naturally the definition of the measure-theoretic entropy of measurable maps.

A *finite measurable partition* $\mathcal{A}$ of a measurable space $(Y, \mathscr{M}(Y))$ is a finite collection of mutually disjoint measurable subsets $\{A_1, \dots, A_n\}$ satisfying $\bigcup_{i=1}^n A_i = Y$, where $n \in \mathbb{N}$.

For arbitrary finite measurable partitions $\mathcal{A}_1, \dots, \mathcal{A}_m$ of measurable spaces $(X^{n_1}, \mathscr{M}(X^{n_1}))$, $\dots$, $(X_{n_m}, \mathscr{M}(X^{n_m}))$, respectively, their product is given by

$$\mathcal{A}_1 \times \cdots \times \mathcal{A}_m := \{A_1 \times \cdots \times A_m : A_i \in \mathcal{A}_i \text{ for every } i \in (m]\} \subseteq \mathscr{M}(X^{n_1 + \cdots + n_m}).$$

It is a finite measurable partition of $(X^{n_1+\cdots+n_m}, \mathscr{M}(X^{n_1+\cdots+n_m}))$.

For a finite measurable partition $\mathcal{A}$ and $n \in \mathbb{N}$, write

$$\mathcal{A}^n := \underbrace{\mathcal{A} \times \mathcal{A} \times \cdots \times \mathcal{A}}_{n \text{ copies of } \mathcal{A}}.$$

Recall that for a finite measurable partition $\mathcal{A}$ and a probability measure μ on X, the entropy of $\mathcal{A}$ is given by

$$H_\mu(\mathcal{A}) := -\sum_{A \in \mathcal{A}} \mu(A) \log(\mu(A)). \tag{5.8}$$

We refer the reader to [PU10, Chapter 2] for basic properties of the entropy $H_\nu(\mathcal{A}_1)$ of a finite measure partition $\mathcal{A}_1$ and the *conditional entropy* $H_\nu(\mathcal{A}_1|\mathcal{A}_2)$ of a finite measurable partition $\mathcal{A}_1$ given another finite measurable partition $\mathcal{A}_2$.

Proposition 5.20. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $\mu \in \mathcal{M}(X, \mathcal{Q})$, and $\mathcal{A}$ be a finite measurable partition of X. Then we have*

$$H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n) = \sum_{k=0}^{n-1} H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A} \times \{X^\omega\} \bigm| \{X\} \times \mathcal{A}^k \times \{X^\omega\}\bigr),$$

where $\{X^\omega\}$ (resp., $\{X\}$) is the partition of X^ω (resp., X) into only one subset.

Moreover, the limit $\lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n)$ exists.

Proof. By (5.7) and (5.8), we have $H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n) = H_{\mu\mathcal{Q}^\omega}(\mathcal{A}^n \times \{X^\omega\})$. By (A.9) in Corollary A.7 and (5.8), we have $H_{\mu\mathcal{Q}^\omega}\bigl(\{X\} \times \mathcal{A}^k \times \{X^\omega\}\bigr) = H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A}^k \times \{X^\omega\}\bigr)$. Thus

$$\begin{aligned} H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n) &= H_{\mu\mathcal{Q}^\omega}(\mathcal{A}^n \times \{X^\omega\}) \\ &= \sum_{k=0}^{n-1} \bigl(H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A}^{k+1} \times \{X^\omega\}\bigr) - H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A}^k \times \{X^\omega\}\bigr) \bigr) \\ &= \sum_{k=0}^{n-1} \bigl(H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A}^{k+1} \times \{X^\omega\}\bigr) - H_{\mu\mathcal{Q}^\omega}\bigl(\{X\} \times \mathcal{A}^k \times \{X^\omega\}\bigr) \bigr) \\ &= \sum_{k=0}^{n-1} H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A} \times \{X^\omega\} \bigm| \{X\} \times \mathcal{A}^k \times \{X^\omega\}\bigr). \end{aligned}$$

Here $\mathcal{A}^0 \times \{X^\omega\}$ is $\{X^\omega\}$, whose entropy is 0.

Since $H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A} \times \{X^\omega\} \bigm| \{X\} \times \mathcal{A}^k \times \{X^\omega\}\bigr)$ is non-negative and it decreases as k increases, we get

$$\lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n) = \lim_{k\to+\infty} H_{\mu\mathcal{Q}^\omega}\bigl(\mathcal{A} \times \{X^\omega\} \bigm| \{X\} \times \mathcal{A}^k \times \{X^\omega\}\bigr).$$

Therefore, the limit $\lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n)$ exists. □

This proposition guarantees that $h_\mu(\mathcal{Q}, \mathcal{A})$ in the next definition is well-defined.

Definition 5.21. Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, μ be a $\mathcal{Q}$-invariant probability measure on X, and $\mathcal{A}$ be a finite measurable partition of X. Then $h_\mu(\mathcal{Q}, \mathcal{A})$, the *measure-theoretic entropy of $\mathcal{Q}$ with respect to the partition $\mathcal{A}$*, is defined as

$$h_\mu(\mathcal{Q}, \mathcal{A}) \coloneqq \lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n).$$

We can now formulate our definition of the measure-theoretic entropy of a transition probability kernel.

Definition 5.22. Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$ and μ be a $\mathcal{Q}$-invariant probability measure on X. The *measure-theoretic entropy* $h_\mu(\mathcal{Q})$ (of $\mathcal{Q}$ for μ) is given by

$$h_\mu(\mathcal{Q}) \coloneqq \sup_{\mathcal{A}} h_\mu(\mathcal{Q}, \mathcal{A}),$$

where $\mathcal{A}$ ranges over all finite measurable partitions of X.

We can establish a useful result through this definition.

Proposition 5.23. *Let $\mathcal{Q}$, $\mathcal{R}$ be transition probability kernels on a measurable space $(X, \mathscr{M}(X))$ and $\mu \in \mathcal{P}(X)$. If $\mu\mathcal{Q}^{[1]} = \left(\mu\mathcal{R}^{[1]}\right) \circ \iota_2^{-1}$, then $\mu \in \mathcal{M}(X, \mathcal{Q}) \cap \mathcal{M}(X, \mathcal{R})$ and $h_\mu(\mathcal{Q}) = h_\mu(\mathcal{R})$.*

Proof. Suppose $\mu\mathcal{Q}^{[1]} = \left(\mu\mathcal{R}^{[1]}\right) \circ \iota_2^{-1}$. We have shown in Lemma A.10 that $\mu \in \mathcal{M}(X, \mathcal{Q}) \cap \mathcal{M}(X, \mathcal{R})$.

By Lemma A.10 and (5.8), for every finite measurable partition $\mathcal{A}$ of X and every $n \in \mathbb{N}$, we have $h_{\mu\mathcal{Q}^{[n]}}(\mathcal{A}^{n+1}) = h_{\mu\mathcal{R}^{[n]}}(\mathcal{A}^{n+1})$. Consequently, by Definition 5.21, we have $h_\mu(\mathcal{Q}, \mathcal{A}) = h_\mu(\mathcal{R}, \mathcal{A})$. Therefore, by Definition 5.22, $h_\mu(\mathcal{Q}) = h_\mu(\mathcal{R})$. $\square$

5.4. **A characterization of the measure-theoretic entropy.** This subsection aims to establish Theorem 5.24.

Let $(X, \mathscr{M}(X))$ be a measurable space. Denote by σ the shift map on X^ω given by

$$\sigma(x_1, x_2, \dots) \coloneqq (x_2, x_3, \dots) \tag{5.9}$$

for all $(x_1, x_2, \dots) \in X^\omega$.

For two arbitrary finite measurable partitions $\mathcal{A}_1$ and $\mathcal{A}_2$ of X, the finite measurable partition $\mathcal{A}_1 \vee \mathcal{A}_2$ is given by $\mathcal{A}_1 \vee \mathcal{A}_2 \coloneqq \{A_1 \cap A_2 : A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2\}$.

Let $\mathcal{Q}$ be a transition probability kernel on X and μ be a $\mathcal{Q}$-invariant probability measure on X. For arbitrary $n \in \mathbb{N}$ and measurable set $A \in \mathscr{M}(X^n)$, by (A.9) in Corollary A.7 we have

$$(\mu\mathcal{Q}^\omega)(\sigma^{-1}(A \times X^\omega)) = (\mu\mathcal{Q}^\omega)(X \times A \times X^\omega) = (\mu\mathcal{Q}^\omega)(A \times X^\omega).$$

By the Dynkin's π-λ Theorem, we get $(\mu\mathcal{Q}^\omega) \circ \sigma^{-1} = \mu\mathcal{Q}^\omega$, which means that $\mu\mathcal{Q}^\omega$ is σ-invariant. As a result, the invariant measure $\mu\mathcal{Q}^\omega$ and the *measure-theoretic entropy $h_{\mu\mathcal{Q}^\omega}(\sigma)$ of σ for $\mu\mathcal{Q}^\omega$* are well-defined in the sense of classical ergodic theory for single-valued maps (see for example, [PU10, Chapter 2]).

Theorem 5.24. *Let $\mathcal{Q}$ be a transition probability kernel on the measurable space $(X, \mathscr{M}(X))$, μ be a $\mathcal{Q}$-invariant probability measure on X, and σ be the shift map on X^ω. Then we have*

$$h_\mu(\mathcal{Q}) = h_{\mu\mathcal{Q}^\omega}(\sigma).$$

Before establishing Theorem 5.24, we give three lemmas.

Lemma 5.25. *Let $\mathcal{Q}$ be a transition probability kernel on the measurable space $(X, \mathscr{M}(X))$, $\mu \in \mathcal{M}(X, \mathcal{Q})$, σ be the shift map on X^ω, and $\mathcal{A}$ be a finite measurable partition of X. Then*

$$h_\mu(\mathcal{Q}, \mathcal{A}) = h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A} \times \{X^\omega\}).$$

Here the definition of the *measure-theoretic entropy* $h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A} \times \{X^\omega\})$ *of the single-valued map* σ *with respect to the partition* $\mathcal{A} \times \{X^\omega\}$ can be found in [PU10, Lemma 2.4.2].

Proof. First, for each $n \in \mathbb{N}$, we have

$$\bigvee_{j=0}^{n-1} \sigma^{-j}(\mathcal{A} \times \{X^\omega\}) = \bigvee_{j=0}^{n-1} \{X^j\} \times \mathcal{A} \times \{X^\omega\} = \mathcal{A}^n \times \{X^\omega\}.$$

Recall $H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n) = H_{\mu\mathcal{Q}^\omega}(\mathcal{A}^n \times \{X^\omega\})$ from the proof of Proposition 5.20. Therefore,

$$\begin{aligned} h_\mu(\mathcal{Q}, \mathcal{A}) &= \lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^{[n-1]}}(\mathcal{A}^n) \\ &= \lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^\omega}(\mathcal{A}^n \times \{X^\omega\}) \\ &= \lim_{n\to+\infty} \frac{1}{n} H_{\mu\mathcal{Q}^\omega}\left(\bigvee_{j=0}^{n-1} \sigma^{-j}(\mathcal{A} \times \{X^\omega\})\right) \\ &= h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A} \times \{X^\omega\}), \end{aligned}$$

where the last equality follows from [PU10, Lemma 2.4.2]. $\square$

The next two lemmas come from [Wa82]:

Lemma 5.26 ([Wa82, Theorem 4.12 (vi)]). *Let* $(Y, \mathscr{M}(Y))$ *be a measurable space,* $F\colon Y \to Y$ *be a measurable map,* μ *be an* F*-invariant probability measure on* Y*,* $\mathcal{A}$ *be a finite measurable partition of* Y*, and* $k \in \mathbb{N}$*. Then we have*

$$h_\mu(F, \mathcal{A}) = h_\mu\big(F, \mathcal{A} \vee F^{-1}(\mathcal{A}) \vee \cdots \vee F^{-k}(\mathcal{A})\big).$$

Lemma 5.27 ([Wa82, Theorem 4.21]). *Let* $(Y, \mathscr{M}(Y))$ *be a measurable space,* $F\colon Y \to Y$ *be a measurable map,* μ *be an* F*-invariant probability measure on* Y*, and* $\mathcal{S}$ *be a sub-algebra of* $\mathscr{M}(Y)$ *satisfying that* $\mathscr{M}(Y)$ *is the* σ*-algebra generated by* $\mathcal{S}$*. Then we have*

$$h_\mu(F) = \sup_{\mathcal{A}} h_\mu(F, \mathcal{A}),$$

where $\mathcal{A}$ *ranges over all finite partitions of* Y *satisfying* $\mathcal{A} \subseteq \mathcal{S}$*.*

Proof of Theorem 5.24. First, we set some notations:

$$\begin{aligned} \mathfrak{A}_0 &\coloneqq \{\mathcal{A}^n \times \{X^\omega\} : \mathcal{A} \text{ is a finite measurable partition of } X \text{ and } n \in \mathbb{N}\}, \\ \mathcal{G} &\coloneqq \{B_1 \times \cdots \times B_n \times X^\omega : n \in \mathbb{N} \text{ and } B_1, \dots, B_n \in \mathscr{M}(X)\}, \\ \mathcal{S} &\coloneqq \{E : E \text{ is a finite union of sets in } \mathcal{G}\}, \\ \mathfrak{A} &\coloneqq \{\mathcal{D} : \mathcal{D} \text{ is a finite measurable partition of } X^\omega \text{ and } \mathcal{D} \subseteq \mathcal{S}\}. \end{aligned}$$

We can verify that $\mathcal{S}$ is a sub-algebra of $\mathscr{M}(X^\omega)$ and that $\mathscr{M}(X^\omega)$ is the σ-algebra generated by $\mathcal{S}$. So by Lemma 5.27, we get

$$h_{\mu\mathcal{Q}^\omega}(\sigma) = \sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{D}) : \mathcal{D} \in \mathfrak{A}\}. \tag{5.10}$$

Fix an arbitrary $\mathcal{D} \in \mathfrak{A}$. Suppose $\mathcal{D} = \{D_1, \dots, D_p\}$, $D_i = G_{i1} \cup \dots \cup G_{iq_i} \in \mathcal{S}$ for all $i \in (p]$, and $G_{ij} = B_{ij1} \times \dots \times B_{ijr_{ij}} \times X^\omega \in \mathcal{G}$ for all $i \in (p]$ and $j \in (q_i]$, where $B_{ijk} \in \mathscr{M}(X)$ for all $i \in (p]$, $j \in (q_i]$, and $k \in (r_{ij}]$. Set $\mathcal{A} := \bigvee_{i=1}^{p} \bigvee_{j=1}^{q_i} \bigvee_{k=1}^{r_{ij}} \{B_{ijk}, B_{ijk}^c\}$, a finite measurable partition of X, then $\mathcal{A}^n \times \{X^\omega\} \in \mathfrak{A}_0$ is a finer partition of X^ω than $\mathcal{D}$, where $n = \max\{r_{ij} : i \in (p], j \in (q_i]\}$, and a partition is finer than another means that each element in this partition contains in an element in another partition.

From the discussion above we conclude that for each partition in $\mathfrak{A}$, we can find a finer partition in $\mathfrak{A}_0$.

We can conclude by (B.3) that if a finite measurable partition $\mathcal{A}'$ of X^ω is finer than another finite measurable partition $\mathcal{A}''$, then $h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A}') \geqslant h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A}'')$. Thus $\sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{D}) : \mathcal{D} \in \mathfrak{A}\} \leqslant \sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A}_0) : \mathcal{A}_0 \in \mathfrak{A}_0\}$. Moreover, we can check that $\mathfrak{A}_0 \subseteq \mathfrak{A}$, and thus

$$\sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{D}) : \mathcal{D} \in \mathfrak{A}\} = \sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A}_0) : \mathcal{A}_0 \in \mathfrak{A}_0\}. \tag{5.11}$$

For each $\mathcal{A}^n \times \{X^\omega\} \in \mathfrak{A}_0$ where $\mathcal{A}$ is a measurable partition of X and $n \in \mathbb{N}$, recall $\bigvee_{j=0}^{n-1} \sigma^{-j}(\mathcal{A} \times \{X^\omega\}) = \mathcal{A}^n \times \{X^\omega\}$ from the proof of Lemma 5.25. By (5.10), (5.11), and Lemma 5.26, we have

$$\begin{aligned} h_{\mu\mathcal{Q}^\omega}(\sigma) &= \sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{D}) : \mathcal{D} \in \mathfrak{A}\} \\ &= \sup\{h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A}_0) : \mathcal{A}_0 \in \mathfrak{A}_0\} \\ &= \sup_{\mathcal{A}} h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A}^n \times \{X^\omega\}) \\ &= \sup_{\mathcal{A}} h_{\mu\mathcal{Q}^\omega}\left(\sigma, \bigvee_{i=0}^{n-1} \sigma^{-i}(\mathcal{A} \times \{X^\omega\})\right) \\ &= \sup_{\mathcal{A}} h_{\mu\mathcal{Q}^\omega}(\sigma, \mathcal{A} \times \{X^\omega\}), \end{aligned}$$

where $\mathcal{A}$ ranges over all finite measurable partitions of X.

Therefore, by Definition 5.22 and Lemma 5.25 we get $h_{\mu\mathcal{Q}^\omega}(\sigma) = \sup_{\mathcal{A}} h_\mu(\mathcal{Q}, \mathcal{A}) = h_\mu(\mathcal{Q})$, where $\mathcal{A}$ ranges over all finite measurable partitions of X. $\square$

A lemma which is useful in Subsection 3.1 follows from Theorem 5.24:

Lemma 5.28. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $Y \in \mathscr{M}(X)$, and $\mathcal{Q}'$ be a transition probability kernel on the measurable space $(Y, \mathscr{M}(Y))$, where $\mathscr{M}(Y)$ refers to the σ-algebra induced by $\mathscr{M}(X)$. Suppose $\mu \in \mathcal{M}(Y, \mathcal{Q}')$ and $\widehat{\mu} \in \mathcal{P}(X)$ satisfy $\widehat{\mu}(A) = \mu(A)$ for all $A \in \mathscr{M}(Y)$. If for each $A \in \mathscr{M}(Y)$, the equality $\mathcal{Q}(y, A) = \mathcal{Q}'(y, A)$ holds for μ-almost every $y \in Y$, then we have $\widehat{\mu} \in \mathcal{M}(X, \mathcal{Q})$ and $h_{\widehat{\mu}}(\mathcal{Q}) = h_\mu(\mathcal{Q}')$.*

Proof. Denote by σ_X the shift map on X^ω given in (5.9), and $\sigma_Y := \sigma_X|_{Y^\omega}$ be the shift map on Y^ω. The conditions above indicate that $\widehat{\mu}\mathcal{Q} = \widehat{\mu}$ by (5.6), and that the measure $\widehat{\mu}\mathcal{Q}^\omega$ is the extension of the measure $\mu\mathcal{Q}'^\omega$ from Y^ω to X^ω, i.e., $(\widehat{\mu}\mathcal{Q}^\omega)(B) = (\mu\mathcal{Q}'^\omega)(B \cap Y^\omega)$ holds for all $B \in \mathscr{M}(X^\omega)$. As a result, the inclusion map from Y^ω to X^ω is an isomorphism between measure-preserving systems $(Y^\omega, \mathscr{M}(Y^\omega), \sigma_Y, \mu\mathcal{Q}'^\omega)$

and $(X^\omega, \mathscr{M}(X^\omega), \sigma_X, \widehat{\mu}\mathcal{Q}^\omega)$, so $h_{\mu\mathcal{Q}'^\omega}(\sigma_Y) = h_{\widehat{\mu}\mathcal{Q}^\omega}(\sigma_X)$. Then $h_{\widehat{\mu}}(\mathcal{Q}) = h_\mu(\mathcal{Q}')$ follows from Theorem 5.24. $\square$

6. Variational Principle for forward expansive correspondences

In this section, we establish the Variational Principle when the correspondence T on a compact metric space (X, d) has a property called *forward expansive.* More precisely, we will prove Theorem A, the first main result of this work. The proof of this theorem is the most technical part of this work. In Subsection 6.1, we introduce forward expansiveness for correspondences. Then in Subsection 6.2, we prove a lemma for Subsection 6.3. Subsection 6.3 is devoted to establishing the Rokhlin formula for measure-theoretic entropy of transition probability kernels and of the corresponding shift maps. Theorem D is established in Subsection 6.4, and is used in the proof of Theorem A. Finally, in Subsection 6.5, we establish an inequality about measure-theoretic entropy (Proposition 6.18) by the Rokhlin formulas and prove Theorem A.

6.1. **Forward expansiveness.** R. K. Williams has defined a type of expansiveness for correspondences or set-valued functions in [Wi70, Definition 3], which is called *RW-expansiveness* by M. J. Pacifico and J. Vieitez in [PV17, Definition 3.2]. In Definition 6.1 below, what we call *forward expansiveness* is inspired but different from what M. J. Pacifico and J. Vieitez called *RW-expansiveness*.

Definition 6.1 (Forward expansiveness)**.** Let T be a correspondence on a compact metric space (X, d). We say that T is *forward expansive*, if there exists a number $\epsilon > 0$ such that for each pair of distinct orbits $(x_1, x_2, \dots), (y_1, y_2, \dots) \in \mathcal{O}_\omega(T)$, we have $d(x_n, y_n) > \epsilon$ for some $n \in \mathbb{N}$. Such a positive number ϵ is called an *expansive constant* of T.

Remark 6.2. Let T be a forward expansive correspondence on a compact metric space (X, d) with an expansive constant $\epsilon > 0$. Fix an arbitrary point $x_1 \in X$ and choose an orbit $(x_1, x_2, \dots) \in \mathcal{O}_\omega(T)$. For a pair of distinct points x_0, $x_0' \in X$ satisfying $x_1 \in T(x_0) \cap T(x_0')$, we have $(x_0, x_1, \dots) \in \mathcal{O}_\omega(T)$ and $(x_0', x_1, \dots) \in \mathcal{O}_\omega(T)$. By the forward expansiveness of T, we must have $d(x_0, x_0') > \epsilon$. Since X is compact, we know that $T^{-1}(x) = \{y \in X : x \in T(y)\}$ is finite for all $x \in X$. Moreover, for each $x \in X$, we have $\#T^{-1}(x) \leqslant M_\epsilon$, where M_ϵ refers to the largest cardinality of an ϵ-separated subset of X.

If the correspondence T on X degenerates to a singe-valued continuous map, then the forward expansiveness for T is equivalent to the forward expansiveness for the corresponding single-valued map, see Appendix B.3 (i) for the precise statement. Moreover, Propositions D.1 and D.2 show that the forward expansiveness of $(\mathcal{O}_\omega(T), \sigma)$ and the forward expansiveness of T are equivalent.

Let X be a compact metric space and T be a correspondence on X. In Subsections 5.4 and 4.3, we give two different shift maps, one on X^ω and one on $\mathcal{O}_\omega(T)$. To distinguish them, we denote by $\sigma' \colon X^\omega \to X^\omega$ the shift map in Subsection 5.4 given by $\sigma'(x_1, x_2, \dots) \coloneqq (x_2, x_3, \dots)$, and by $\sigma \colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ the shift map in Subsection 4.3 given by (4.6).

Lemma 6.3. *If a correspondence T on a compact metric space X is forward expansive, then for every $A \in \mathscr{B}(X)$, the set $T(A)$ is Borel measurable.*

Proof. Recall from Remark 6.2 that if two different points x, $y \in X$ satisfy $T(x) \cap T(y) \neq \emptyset$, then $d(x, y) > \epsilon$.

Fix an arbitrary $x \in X$. The arguments above implies that $T(y_1) \cap T(y_2) = \emptyset$ for all y_1, $y_2 \in \overline{B_{\epsilon/2}(x)} := \{y \in X : d(x, y) \leqslant \epsilon/2\}$ with $y_1 \neq y_2$, so for each $z \in T\bigl(\overline{B_{\epsilon/2}(x)}\bigr)$, there is exactly one $y \in \overline{B_{\epsilon/2}(x)}$ satisfying $z \in T(y)$. Suppose $g_x \colon T\bigl(\overline{B_{\epsilon/2}(x)}\bigr) \to \overline{B_{\epsilon/2}(x)}$ is the map with the property that $g_x(z)$ is the unique point $y \in \overline{B_{\epsilon/2}(x)}$ with $z \in T(y)$, i.e., $T = g_x^{-1}$ on $\overline{B_{\epsilon/2}(x)}$. Since $\mathcal{O}_2(T)$ is compact by the definition of correspondences, $T\bigl(\overline{B_{\epsilon/2}(x)}\bigr)$, the projection of $\mathcal{O}_2(T) \cap \overline{B_{\epsilon/2}(x)} \times X$ on the second coordinate, is compact. Since $\overline{B_{\epsilon/2}(x)}$ is compact and $\bigl\{(y, z) : z \in T\bigl(\overline{B_{\epsilon/2}(x)}\bigr), y = g_x(z)\bigr\} = \bigl\{(y, z) : y \in \overline{B_{\epsilon/2}(x)}, z \in T(y)\bigr\} = \mathcal{O}_2(T) \cap \overline{B_{\epsilon/2}(x)} \times X$ is compact, we get that $g_x \colon T\bigl(\overline{B_{\epsilon/2}(x)}\bigr) \to \overline{B_{\epsilon/2}(x)}$ is continuous. Thereby, for each Borel set $A \subseteq \overline{B_{\epsilon/2}(x)}$, the set $T(A) = g_x^{-1}(A)$ is Borel measurable in $T\bigl(\overline{B_{\epsilon/2}(x)}\bigr)$, and thus is Borel measurable in X due to the fact that $T\bigl(\overline{B_{\epsilon/2}(x)}\bigr)$ is a closed subset of X.

Since X is compact, we can choose a finite collection of points $\{x_1, \dots, x_n\} \subseteq X$ such that $\bigl\{\overline{B_{\epsilon/2}(x_i)}\bigr\}_{i=1}^n$ covers X. Let $A \in \mathscr{B}(X)$ be arbitrary. For each $i \in (n]$, the Borel set $A \cap \overline{B_{\epsilon/2}(x_i)}$ is contained in $\overline{B_{\epsilon/2}(x_i)}$, and thus $T\bigl(A \cap \overline{B_{\epsilon/2}(x_i)}\bigr)$ is Borel measurable. As a result, therefore,

$$T(A) = \bigcup_{i=1}^n T\bigl(A \cap \overline{B_{\epsilon/2}(x_i)}\bigr)$$

is Borel measurable. □

Corollary 6.4. *Let (X, d) be a compact metric space. If a continuous map $f \colon X \to X$ is forward expansive, then for each Borel measurable set $A \in \mathscr{B}(X)$, the set $f(A)$ is Borel measurable.*

Proof. Let $\mathcal{C}_f$ be the correspondence on X induced by f, see Appendix B.2 for details. By Appendix B.3 (i), $\mathcal{C}_f$ is forward expansive. Therefore, by Lemma 6.3, $f(A) = \mathcal{C}_f(A)$ is Borel measurable for all $A \in \mathscr{B}(X)$. □

Lemma 6.3 and Corollary 6.4 are important in Sections 6 and 7, because Theorems A, B, and C all assume that the correspondence T is forward expansive. When the correspondence T is forward expansive and at the same time $\sigma \colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ is forward expansive, we can write $T(A)$ as an Borel set of X for every $A \in \mathscr{B}(X)$ and $\sigma(B)$ as an Borel set of $\mathcal{O}_\omega(T)$ for every $B \in \mathscr{B}(\mathcal{O}_\omega(T))$.

Let T be a correspondence on a compact metric space (X, d) and $\mathcal{A}$ be a finite Borel measurable partition on X. Set

$$\operatorname{mesh} \mathcal{A} := \sup\{\operatorname{diam} B : B \in \mathcal{A}\}.$$

For each $n \in \mathbb{N}$, the pair $(T, \mathcal{A})$ induces a finite Borel measurable partition $\widetilde{\mathcal{A}}_T^n$ of the orbit space $\mathcal{O}_\omega(T)$ given by

$$\widetilde{\mathcal{A}}_T^n := \{B_1 \times \cdots \times B_n \times X^\omega \cap \mathcal{O}_\omega(T) : B_1, \ldots, B_n \in \mathcal{A}\}. \tag{6.1}$$

Note that $\widetilde{\mathcal{A}}_T^n = \bigvee_{k=0}^{n-1} \sigma^{-k}\big(\widetilde{\mathcal{A}}_T^1\big)$ for all $n \in \mathbb{N}$. For each $x \in \mathcal{O}_\omega(T)$ and each $n \in \mathbb{N}$, denote by $\widetilde{\mathcal{A}}_T^n(x)$ the element in $\widetilde{\mathcal{A}}_T^n$ containing x.

Lemma 6.5. *Let T be a forward expansive correspondence on a compact metric space (X, d) with expansive constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$. There exists $L \in \mathbb{N}$ with the following property:*

For each $n \in \mathbb{N}$ greater than L, if two orbits $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n) \in \mathcal{O}_n(T)$ satisfy $d(x_k, y_k) < \epsilon_2$ for all $k \in (n]$, then $d(x_k, y_k) < \epsilon_1$ holds for all $k \in (n - L]$.

Proof. We argue by contradiction and assume that for every $l \in \mathbb{N}$, we can choose two orbits $\big(x_1^l, \ldots, x_{n_l}^l\big)$, $\big(y_1^l, \ldots, y_{n_l}^l\big) \in \mathcal{O}_{n_l}(T)$, $n_l \in \mathbb{N}$, $n_l > l$ satisfying that $d\big(x_k^l, y_k^l\big) < \epsilon_2$ for every $k \in (n_l]$, and that there exists $j \in (n_l - l]$ such that $d\big(x_j^l, y_j^l\big) \geqslant \epsilon_1$. Assume $j = 1$, otherwise substitute $\big(x_j^l, \ldots, x_{n_l}^l\big)$ and $\big(y_j^l, \ldots, y_{n_l}^l\big)$ for $\big(x_1^l, \ldots, x_{n_l}^l\big)$ and $\big(y_1^l, \ldots, y_{n_l}^l\big)$, respectively.

Extend each pair of orbits $\big(x_1^l, \ldots, x_{n_l}^l\big)$, $\big(y_1^l, \ldots, y_{n_l}^l\big) \in \mathcal{O}_{n_l}(T)$ to $\big(x_1^l, x_2^l, \ldots\big)$, $\big(y_1^l, y_2^l, \ldots\big) \in \mathcal{O}_\omega(T)$. Since $\mathcal{O}_\omega(T) \times \mathcal{O}_\omega(T)$ is compact, we can choose an increasing sequence of positive integers $l_r \in \mathbb{N}$, $r \in \mathbb{N}$, such that $\big(x_1^{l_r}, x_2^{l_r}, \ldots\big)$ and $\big(y_1^{l_r}, y_2^{l_r}, \ldots\big)$ converge to $(x_1^0, x_2^0, \ldots) \in \mathcal{O}_\omega(T)$ and $(y_1^0, y_2^0, \ldots) \in \mathcal{O}_\omega(T)$ as $r \to +\infty$, respectively, i.e., for each $k \in \mathbb{N}$, $x_k^{l_r}$ and $y_k^{l_r}$ converge to x_k^0 and y_k^0 as $r \to +\infty$, respectively.

Fix an arbitrary $k \in \mathbb{N}$, since for each $r \in \mathbb{N}$ with $n_{l_r} > l_r \geqslant k$, we have $d\big(x_k^{l_r}, y_k^{l_r}\big) < \epsilon_2$, and since l_r tends to $+\infty$ as $r \to +\infty$, we get $d\big(x_k^0, y_k^0\big) = \lim_{r \to +\infty} d\big(x_k^{l_r}, y_k^{l_r}\big) \leqslant \epsilon_2$. This implies $x_k^0 = y_k^0$ for each $k \in \mathbb{N}$ because ϵ_2 is an expansive constant of T.

Recall $d\big(x_1^l, y_1^l\big) \geqslant \epsilon_1$ for all $l \in \mathbb{N}$. Thus we have

$$0 = d(x_1^0, y_1^0) = \lim_{r \to +\infty} d\big(x_1^{l_r}, y_1^{l_r}\big) \geqslant \epsilon_1 > 0,$$

which is impossible. $\square$

This lemma leads to the following corollaries.

Corollary 6.6. *Let T be a forward expansive correspondence on a compact metric space (X, d) with an expansive constant $\epsilon > 0$ and $\mathcal{A}$ be a finite Borel measurable partition of X with* $\operatorname{mesh} \mathcal{A} < \epsilon$. *Then* $\lim_{n \to +\infty} \operatorname{mesh} \widetilde{\mathcal{A}}_T^n = 0$.

Proof. Fix an arbitrary $\delta \in (0, \epsilon)$. Since ϵ is an expansive constant for T, δ is also an expansive constant for T. Choose $N \in \mathbb{N}$ such that $\frac{1}{2^N} < \frac{\delta}{2}$. By Lemma 6.5, we can choose $L \in \mathbb{N}$ greater than N with the following property:

For each $n \in \mathbb{N}$ greater than L, if two orbits $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n) \in \mathcal{O}_n(T)$ satisfy $d(x_k, y_k) < \epsilon$ for all $k \in (n]$, then $d(x_k, y_k) < \frac{\delta}{2}$ holds for all $k \in (n - L]$.

Fix an arbitrary $n \in \mathbb{N}$ greater than $N + L$. If two orbits $(x_1, x_2, \ldots), (y_1, y_2, \ldots) \in \mathcal{O}_\omega(T)$ belong to the same element in the partition $\widetilde{\mathcal{A}}_T^n$, then we have $d(x_k, y_k) < \epsilon$ for all $k \in (n]$ by $\operatorname{mesh} \mathcal{A} < \epsilon$. Then it follows from the property of L that $d(x_k, y_k) < \frac{\delta}{2}$ for all $k \in (n-L]$, so

$$d_\omega((x_1, x_2, \dots), (y_1, y_2, \dots)) \leqslant \sum_{k=1}^{n-L} \frac{1}{2^k} \frac{\delta}{1+\delta} + \sum_{k=n-L+1}^{+\infty} \frac{1}{2^k} < \frac{\delta}{2} + \frac{1}{2^N} \leqslant \delta.$$

Consequently, we have $\operatorname{mesh} \widetilde{\mathcal{A}}_T^n \leqslant \delta$. Since δ is chosen arbitrarily, we conclude that $\lim_{n\to+\infty} \operatorname{mesh} \widetilde{\mathcal{A}}_T^n = 0$. □

Corollary 6.7. *Let T be a forward expansive correspondence on a compact metric space X with an expansive constant $\epsilon > 0$, ν be a σ-invariant Borel probability measure on $\mathcal{O}_\omega(T)$, and $\mathcal{A}$ be a finite Borel measurable partition of X with* $\operatorname{mesh} \mathcal{A} < \epsilon$. *Then the partition $\widetilde{\mathcal{A}}_T^1$ of $\mathcal{O}_\omega(T)$ is a finite one-sided generator for ν, i.e., if $\underline{x}$, $\underline{y} \in \mathcal{O}_\omega(T)$ satisfy $\widetilde{\mathcal{A}}_T^1(\sigma^n(\underline{x})) = \widetilde{\mathcal{A}}_T^1(\sigma^n(\underline{y}))$ for all $n \in \mathbb{N}$, then $\underline{x} = \underline{y}$. Moreover, we have*

$$h_\nu(\sigma) = h_\nu\bigl(\sigma, \widetilde{\mathcal{A}}_T^1\bigr). \tag{6.2}$$

Proof. By Corollary 6.6 and Lemma D.1, we can choose $n \in \mathbb{N}$ such that $\operatorname{mesh} \widetilde{\mathcal{A}}_T^n$ is less than some expansive constant for σ. By [PU10, Lemma 3.5.5], we get that the partition $\widetilde{\mathcal{A}}_T^n$ is a finite one-sided generator for ν, i.e., if $\underline{x}$, $\underline{y} \in \mathcal{O}_\omega(T)$ satisfy $\widetilde{\mathcal{A}}_T^n(\sigma^m(\underline{x})) = \widetilde{\mathcal{A}}_T^n(\sigma^m(\underline{y}))$ for all $m \in \mathbb{N}$, then $\underline{x} = \underline{y}$. Recall $\widetilde{\mathcal{A}}_T^n = \bigvee_{k=0}^{n-1} \sigma^{-k}\bigl(\widetilde{\mathcal{A}}_T^1\bigr)$, so $\widetilde{\mathcal{A}}_T^n(\sigma^m(\underline{x})) = \widetilde{\mathcal{A}}_T^n(\sigma^m(\underline{y}))$ is equivalent to the statement that $\widetilde{\mathcal{A}}_T^1\bigl(\sigma^{m+k}(\underline{x})\bigr) = \widetilde{\mathcal{A}}_T^1\bigl(\sigma^{m+k}(\underline{y})\bigr)$ holds for all $k \in [n-1]$. Hence, the fact that $\widetilde{\mathcal{A}}_T^n$ is a finite one-sided generator for ν implies that $\widetilde{\mathcal{A}}_T^1$ is a finite one-sided generator for ν. Finally, (6.2) follows by [PU10, Theorem 2.8.7 (b)]. □

6.2. **Properties of the induced partition $\widetilde{\mathcal{A}}_T^n$ of the orbit space.** We show the following lemma in this subsection.

Proposition 6.8. *Let T be a forward expansive correspondence on a compact metric space (X, d) with an expansive constant $\epsilon > 0$, $\mathcal{A}$ be a finite measurable partition of X with* $\operatorname{mesh} \mathcal{A} < \epsilon$, *$\nu \in \mathcal{P}(\mathcal{O}_\omega(T), \mathbb{R})$, and $f \in B(\mathcal{O}_\omega(T), \mathbb{R})$. Then for ν-almost every $\underline{x} \in \mathcal{O}_\omega(T)$, the following properties hold:*

(i) *For all $n \in \mathbb{N}$, we have $\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr) > 0$.*
(ii) *The following limit exists and the equality holds:*

$$\lim_{n\to+\infty} \frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} f \,\mathrm{d}\nu = f(\underline{x}).$$

Proof. Set $S := \bigl\{\underline{x} \in \mathcal{O}_\omega(T) : \nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr) = 0 \text{ for some } n \in \mathbb{N}\bigr\}$.

For each $N \in \mathbb{N}$, set

$$J_N := \biggl\{\underline{x} \in \mathcal{O}_\omega(T) : \nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr) > 0 \text{ for all } n \in \mathbb{N} \text{ and } \limsup_{m\to+\infty} \frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^m(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^m(\underline{x})} \lvert f(\overline{x}) - f(\underline{x})\rvert \,\mathrm{d}\nu(\overline{x}) > \frac{1}{N}\biggr\}.$$

Note that all orbits in $\mathcal{O}_\omega(T) \setminus \bigl(S \cup \bigl(\bigcup_{N=1}^{+\infty} J_N\bigr)\bigr)$ satisfy properties (i) and (ii), so we aim to show that this set is of full measure in terms of ν.

To this end, we first show $S \in \mathscr{B}(\mathcal{O}_\omega(T))$ and $\nu(S) = 0$.

Write $\mathcal{D} := \bigl\{A \in \bigcup_{n=1}^{+\infty} \widetilde{\mathcal{A}}_T^n : \nu(A) = 0\bigr\}$. Note $S = \bigcup \mathcal{D} \in \mathscr{B}(\mathcal{O}_\omega(T))$. Since $\bigcup_{n=1}^{+\infty} \widetilde{\mathcal{A}}_T^n$, a countable union of finite sets, is countable, its subset $\mathcal{D}$ is also countable. Consequently, we get $\nu(S) = \nu(\bigcup \mathcal{D}) = 0$.

Next, we fix an arbitrary $N \in \mathbb{N}$ and turn our attention to the structure of J_N.

Fix an arbitrary $n \in \mathbb{N}$. For each $A \in \widetilde{\mathcal{A}}_T^n$, since $f : \mathcal{O}_\omega(T) \to \mathbb{R}$ is Borel measurable, the integral $\int_A |f(\overline{x}) - f(\underline{x})| \,\mathrm{d}\nu(\overline{x})$, as a function of $\underline{x}$, is Borel measurable on A. If $A \setminus S \neq \emptyset$, then $\nu(A) > 0$. In this case, the function

$$\frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |f(\overline{x}) - f(\underline{x})| \,\mathrm{d}\nu(\overline{x}) = \frac{1}{\nu(A)} \int_A |f(\overline{x}) - f(\underline{x})| \,\mathrm{d}\nu(\overline{x}),$$

in the variable $\underline{x}$, is Borel measurable on $A \setminus S$. Thus

$$\frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |f(\overline{x}) - f(\underline{x})| \,\mathrm{d}\nu(\overline{x}),$$

as a function of $\underline{x}$, is Borel measurable on $\mathcal{O}_\omega(T) \setminus S$. Since $n \in \mathbb{N}$ is arbitrary, we conclude that J_N is a Borel set.

We choose $M > 0$ such that $|f(\underline{x})| \leqslant M$ for all $\underline{x} \in \mathcal{O}_\omega(T)$. Fix an arbitrary $\delta > 0$. By Lusin's theorem (see for example, [Fol99, Theorem 7.10]), we can choose a continuous function $g : \mathcal{O}_\omega(T) \to \mathbb{R}$ with the following properties:

(a) For each $\underline{x} \in \mathcal{O}_\omega(T)$, we have $|g(\underline{x})| \leqslant M$.

(b) The compact set $K := \{\underline{x} \in \mathcal{O}_\omega(T) : f(\underline{x}) = g(\underline{x})\}$ satisfies $\nu(K) > 1 - \delta$.

Fix an arbitrary $\underline{x} \in J_N$. We have

$$\begin{aligned}\frac{1}{N} &< \limsup_{n\to+\infty} \frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |f(\overline{x}) - f(\underline{x})| \,\mathrm{d}\nu(\overline{x}) \\ &\leqslant |f(\underline{x}) - g(\underline{x})| + h(\underline{x}) + \limsup_{n\to+\infty} \frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |g(\overline{x}) - g(\underline{x})| \,\mathrm{d}\nu(\overline{x}),\end{aligned} \tag{6.3}$$

where

$$h(\underline{x}) := \limsup_{n\to+\infty} \frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |f(\overline{x}) - g(\overline{x})| \,\mathrm{d}\nu(\overline{x}).$$

By Corollary 6.6 and the fact that g is continuous, we have

$$\limsup_{n\to+\infty} \frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |g(\overline{x}) - g(\underline{x})| \,\mathrm{d}\nu(\overline{x}) = 0.$$

As a result, (6.3) implies that either $f(\underline{x}) \neq g(\underline{x})$, in which case $\underline{x} \in \mathcal{O}_\omega(T) \setminus K$, or $h(\underline{x}) > 1/N$.

Set

$$I_N := \bigl\{\underline{x} \in \mathcal{O}_\omega(T) : \nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr) > 0 \text{ for all } n \in \mathbb{N} \text{ and } h(\underline{x}) > 1/N\bigr\}.$$

Then $J_N \subseteq I_N \cup (\mathcal{O}_\omega(T) \setminus K)$. Write

$$\mathcal{C} := \left\{ A \in \bigcup_{n=1}^{+\infty} \widetilde{\mathcal{A}}_T^n : \int_A |f(\overline{x}) - g(\overline{x})| \,\mathrm{d}\nu(\overline{x}) > \frac{\nu(A)}{N} \right\}.$$

For each $\underline{x} \in I_N$, choose $n \in \mathbb{N}$ such that

$$\frac{1}{\nu\bigl(\widetilde{\mathcal{A}}_T^n(\underline{x})\bigr)} \int_{\widetilde{\mathcal{A}}_T^n(\underline{x})} |f(\overline{x}) - g(\overline{x})| \,\mathrm{d}\nu(\overline{x}) > \frac{1}{N}.$$

This inequality implies that $\underline{x} \in \widetilde{\mathcal{A}}_T^n(\underline{x}) \in \mathcal{C}$, and thus $\underline{x} \in \bigcup \mathcal{C}$. Hence, $I_N \subseteq \bigcup \mathcal{C}$.

For each $n \in \mathbb{N}$, define $\mathcal{C}_n \subseteq \mathcal{C}$ inductively as follows:

(1) First, set $\mathcal{C}_1 := \widetilde{\mathcal{A}}_T^1 \cap \mathcal{C}$.

(2) For each $n \in \mathbb{N}$, set $\mathcal{C}_{n+1} := \mathcal{C}_n \cup \bigl\{ A \in \widetilde{\mathcal{A}}_T^{n+1} \cap \mathcal{C} : A \cap \bigcup \mathcal{C}_n = \emptyset \bigr\}$.

From the inductive definition, we can see $\mathcal{C}_n \subseteq \bigcup_{k=1}^n \widetilde{\mathcal{A}}_T^k$ for all $n \in \mathbb{N}$, and thus $\mathcal{C}_n$ is a finite set for all $n \in \mathbb{N}$.

Claim 1. For each $n \in \mathbb{N}$, elements in $\mathcal{C}_n$ are mutually disjoint.

Indeed, we establish the claim by induction on n. First, $\mathcal{C}_1 \subseteq \widetilde{\mathcal{A}}_T^1$ implies that elements in $\mathcal{C}_1$ are mutually disjoint. Moreover, for each $n \in \mathbb{N}$, $\mathcal{C}_{n+1} \setminus \mathcal{C}_n \subseteq \widetilde{\mathcal{A}}_T^{n+1}$ implies that elements in $\mathcal{C}_{n+1} \setminus \mathcal{C}_n$ are mutually disjoint, and the requirement $A \cap \bigcup \mathcal{C}_n = \emptyset$ in the definition of $\mathcal{C}_{n+1}$ implies that each element in $\mathcal{C}_{n+1} \setminus \mathcal{C}_n$ does not intersect with any element in $\mathcal{C}_n$. The above arguments show that for each $n \in \mathbb{N}$, if elements in $\mathcal{C}_n$ are mutually disjoint, then the elements in $\mathcal{C}_{n+1}$ are mutually disjoint. Claim 1 now follows by induction on n.

Let $\mathcal{C}' := \bigcup_{n=1}^{+\infty} \mathcal{C}_n \subseteq \mathcal{C}$. Since $\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq \cdots$, elements in $\mathcal{C}'$ are also mutually disjoint. The fact that $\mathcal{C}_n$ is finite for each $n \in \mathbb{N}$ implies that $\mathcal{C}'$ is countable.

Claim 2. $\bigcup \mathcal{C}' = \bigcup \mathcal{C}$.

We already have $\bigcup \mathcal{C}' \subseteq \bigcup \mathcal{C}$ since $\mathcal{C}' \subseteq \mathcal{C}$. We argue by contradiction and suppose $\bigcup \mathcal{C}' \subsetneqq \bigcup \mathcal{C}$. Choose $\underline{x} \in (\bigcup \mathcal{C}) \setminus (\bigcup \mathcal{C}')$. Then we choose $n \in \mathbb{N}$ and $A \in \mathcal{C}$ with the following properties:

(1) $\underline{x} \in A \in \widetilde{\mathcal{A}}_T^n$;

(2) For each $k \in (n-1]$, there is no Borel set $A' \in \mathcal{C}$ with $\underline{x} \in A' \in \widetilde{\mathcal{A}}_T^k$.

We have $A \notin \mathcal{C}'$ since $\underline{x} \in A$ and $\underline{x} \notin \bigcup \mathcal{C}'$. If $n = 1$, then $A \in \widetilde{\mathcal{A}}_T^1 \cap \mathcal{C} = \mathcal{C}_1 \subseteq \mathcal{C}'$. This contradicts $A \notin \mathcal{C}'$. If $n > 1$, then $\mathcal{C}_n \subseteq \mathcal{C}'$ and $A \notin \mathcal{C}'$ indicate that

$$A \notin \mathcal{C}_n = \mathcal{C}_{n-1} \cup \Bigl\{ B \in \widetilde{\mathcal{A}}_T^n \cap \mathcal{C} : B \cap \bigcup \mathcal{C}_{n-1} = \emptyset \Bigr\}.$$

Since $A \in \widetilde{\mathcal{A}}_T^n \cap \mathcal{C}$, we get $A \cap \bigcup \mathcal{C}_{n-1}$ must be non-empty. Choose $B \in \mathcal{C}_{n-1} \subseteq \bigcup_{k=1}^{n-1} \widetilde{\mathcal{A}}_T^k$ intersecting with A and then choose $k \in (n-1]$ such that $B \in \widetilde{\mathcal{A}}_T^k$. Since the partition $\widetilde{\mathcal{A}}_T^n$ is finer than the partition $\widetilde{\mathcal{A}}_T^k$, $A \cap B \neq \emptyset$ implies $A \subseteq B$. As a result, we have $\underline{x} \in A \subseteq B \in \mathcal{C}_{n-1} \subseteq \mathcal{C}'$, which implies that $\underline{x} \in \bigcup \mathcal{C}'$. This contradicts $\underline{x} \in (\bigcup \mathcal{C}) \setminus (\bigcup \mathcal{C}')$, and we conclude that Claim 2 holds.

Since $\mathcal{C}'$ is countable and each pair of elements in $\mathcal{C}'$ is disjoint, we have

$$\int_{\bigcup\mathcal{C}'}|f-g|\,\mathrm{d}\nu=\sum_{A\in\mathcal{C}'}\int_A|f-g|\,\mathrm{d}\nu>\frac{1}{N}\sum_{A\in\mathcal{C}'}\nu(A)=\frac{1}{N}\nu\bigl(\bigcup\mathcal{C}'\bigr).$$

On the other hand, recall that the inequalities $|f(\underline{x})|\leqslant M$ and $|g(\underline{x})|\leqslant M$ hold for all $\underline{x}\in\mathcal{O}_\omega(T)$. We have

$$\int_{\bigcup\mathcal{C}'}|f-g|\,\mathrm{d}\nu\leqslant\int_{\mathcal{O}_\omega(T)}|f-g|\,\mathrm{d}\nu\leqslant 2M\cdot\nu(\mathcal{O}_\omega(T)\setminus K)\leqslant 2M\delta.$$

Since $I_N\subseteq\bigcup\mathcal{C}=\bigcup\mathcal{C}'$, we get

$$\nu(I_N)\leqslant\nu\bigl(\bigcup\mathcal{C}'\bigr)<N\int_{\bigcup\mathcal{C}'}|f-g|\,\mathrm{d}\nu\leqslant 2MN\delta.$$

Since $J_N\subseteq I_N\cup(\mathcal{O}_\omega(T)\setminus K)$, we get $\nu(J_N)\leqslant\nu(I_N)+\nu(\mathcal{O}_\omega(T)\setminus K)<(2MN+1)\delta$. As $\delta>0$ is chosen arbitrarily, taking $\delta\to 0^+$ we get $\nu(J_N)=0$, which holds for every $N\in\mathbb{N}$.

Recall that $\nu(S)=0$ and that all orbits in $\mathcal{O}_\omega(T)\setminus\bigl(S\cup\bigl(\bigcup_{N=1}^{+\infty}J_N\bigr)\bigr)$ satisfy properties (i) and (ii) in the statement of Proposition 6.8. Therefore, we conclude that ν-almost every point $\underline{x}\in\mathcal{O}_\omega(T)$ satisfies properties (i) and (ii). $\square$

6.3. **Rokhlin formulas.** This subsection is devoted to showing Proposition 6.9 and Theorem 6.11. We first use the Shannon–McMillan–Breiman Theorem and Proposition 6.8 to establish the formula in Proposition 6.9, a variant of the Rokhlin formula for measure-theoretic entropy of shift maps. Then we use this formula to establish the formula in Theorem 6.11, the Rokhlin formula for measure-theoretic entropy of transition probability kernels. Finally, in Remark 6.12, we point out that an equivalent form of the classical Rokhlin formula for forward expansive single-valued map (see for example, [PU10, Theorem 2.9.7]) can be deduced from Theorem 6.11. That is why we call the formulas in Proposition 6.9 and Theorem 6.11 the *Rokhlin formulas*. This subsection and Subsection 6.5 rely on Appendix A.2.

Let μ be a probability measure on some measurable space $(Y,\mathscr{B}(Y))$. If there exists a countable measurable set $A\in\mathscr{B}(Y)$ such that $\nu(A)=1$, then we set

$$H(\nu):=-\sum_{y\in A}\nu(\{y\})\log(\nu(\{y\})),\tag{6.4}$$

where we follow the convention that $0\log 0=0$. If $\nu(A)<1$ for all countable measurable set $A\in\mathscr{B}(Y)$, then we set $H(\nu)=+\infty$.

Let ν be an arbitrary Borel probability measure on $\mathcal{O}_\omega(T)$. Denote by $\widehat{\nu}$ the Borel probability measure on X^ω given by $\widehat{\nu}(A):=\nu(A\cap\mathcal{O}_\omega(T))$ for all $A\in\mathscr{B}(X^\omega)$.

Proposition 6.9. *Let T be a forward expansive correspondence on the compact metric space (X,d) with an expansive constant $\epsilon>0$, ν be a σ-invariant Borel probability measure on $\mathcal{O}_\omega(T)$, and $\mathcal{P}$ be a backward conditional transition probability kernel of*

$\widehat{\nu}$ from X^ω to X supported on $\mathcal{O}_2(T) \times X^\omega$. Then we have

$$h_\nu(\sigma) = \int_{\mathcal{O}_\omega(T)} H(\mathcal{P}_{\underline{x}}) \,\mathrm{d}\nu(\underline{x}).$$

See Definition A.14 for the notion of backward conditional transition probability kernels.

Proof. Choose a finite Borel measurable partition $\mathcal{A}$ of X with $\operatorname{mesh}\mathcal{A} < \epsilon$. Recall for each $n \in \mathbb{N}$, $\widetilde{\mathcal{A}}^n_T = \{B_1 \times \cdots \times B_n \times X^\omega \cap \mathcal{O}_\omega(T) : B_1, \dots, B_n \in \mathcal{A}\}$, and that for each $\underline{x} \in \mathcal{O}_\omega(T)$, $\widetilde{\mathcal{A}}^n_T(\underline{x})$ is the element in $\widetilde{\mathcal{A}}^n_T$ containing $\underline{x}$.

For each $n \in \mathbb{N}$, set $\widetilde{\mathcal{A}}^{1,n+1}_T := \bigl\{X \times A \cap \mathcal{O}_\omega(T) : A \in \widetilde{\mathcal{A}}^n_T\bigr\} = \sigma^{-1}\bigl(\widetilde{\mathcal{A}}^n_T\bigr)$, and for each $\underline{x} \in \mathcal{O}_\omega(T)$, denote by $\widetilde{\mathcal{A}}^{1,n+1}_T(\underline{x})$ the element in $\widetilde{\mathcal{A}}^{1,n+1}_T$ containing $\underline{x}$. We can check that $\widetilde{\mathcal{A}}^{1,n+1}_T(\underline{x}) = \sigma^{-1}\bigl(\widetilde{\mathcal{A}}^n_T(\sigma(\underline{x}))\bigr)$ holds for all $\underline{x} \in \mathcal{O}_\omega(T)$, thus $\nu\bigl(\widetilde{\mathcal{A}}^{1,n+1}_T(\underline{x})\bigr) = \nu \circ \sigma^{-1}\bigl(\widetilde{\mathcal{A}}^n_T(\sigma(\underline{x}))\bigr) = \nu\bigl(\widetilde{\mathcal{A}}^n_T(\sigma(\underline{x}))\bigr)$.

We could verify $\widetilde{\mathcal{A}}^n_T = \bigvee_{k=0}^{n-1} \sigma^{-k}\bigl(\widetilde{\mathcal{A}}^1_T\bigr)$ and $\widetilde{\mathcal{A}}^{1,n+1}_T = \bigvee_{k=1}^{n} \sigma^{-k}\bigl(\widetilde{\mathcal{A}}^1_T\bigr)$. By Corollary 6.7, we have $h_\nu(\sigma) = h_\nu\bigl(\sigma, \widetilde{\mathcal{A}}^n_T\bigr)$.

Applying the Shannon–McMillan–Breiman Theorem to the measure-preserving system $(\mathcal{O}_\omega(T), \mathscr{B}(\mathcal{O}_\omega(T)), \nu, \sigma)$ with the partition $\widetilde{\mathcal{A}}^1_T$, we get that for ν-almost every $\underline{x} \in \mathcal{O}_\omega(T)$, $\nu\bigl(\widetilde{\mathcal{A}}^{1,n+1}_T(\underline{x})\bigr) > 0$ holds for all $n \in \mathbb{N}$, the limit $\lim\limits_{n\to+\infty} \frac{\nu(\widetilde{\mathcal{A}}^{n+1}_T(\underline{x}))}{\nu(\widetilde{\mathcal{A}}^{1,n+1}_T(\underline{x}))}$ exists, and we have

$$h_\nu(\sigma) = h_\nu\bigl(\sigma, \widetilde{\mathcal{A}}^1_T\bigr) = \int_{\mathcal{O}_\omega(T)} -\log\Biggl(\lim_{n\to+\infty} \frac{\nu\bigl(\widetilde{\mathcal{A}}^{n+1}_T(\underline{x})\bigr)}{\nu\bigl(\widetilde{\mathcal{A}}^{1,n+1}_T(\underline{x})\bigr)}\Biggr) \,\mathrm{d}\nu(\underline{x}).$$

Note that $\sigma' \colon X^\omega \to X^\omega$ is the projection map from $X^\omega = X \times X^\omega$ onto X^ω. Applying (A.15) in Remark A.12 and writing

$$L_1 := \lim_{n\to+\infty} \frac{\nu\bigl(\widetilde{\mathcal{A}}^{n+1}_T(x_0, x_1, \dots)\bigr)}{\nu\bigl(\widetilde{\mathcal{A}}^{1,n+1}_T(x_0, x_1, \dots)\bigr)} \text{ and } L_2 := \lim_{n\to+\infty} \frac{\nu\bigl(\widetilde{\mathcal{A}}^{n+1}_T(x_0, x_1, \dots)\bigr)}{\nu\bigl(\widetilde{\mathcal{A}}^{n}_T(x_1, x_2, \dots)\bigr)}$$

we get

$$\begin{aligned} h_\nu(\sigma) &= \int_{X^\omega} \biggl(\int_X -\log L_1 \,\mathrm{d}\mathcal{P}_{(x_1,x_2,\dots)}(x_0)\biggr) \,\mathrm{d}(\widehat{\nu} \circ (\sigma')^{-1})(x_1, x_2, \dots) \\ &= \int_{\mathcal{O}_\omega(T)} \Biggl(\sum_{x_0 \in T^{-1}(x_1)} -\mathcal{P}_{(x_1,x_2,\dots)}(\{x_0\}) \log L_2\Biggr) \,\mathrm{d}\nu(x_1, x_2, \dots). \end{aligned} \tag{6.5}$$

Since $\sigma(\mathcal{O}_\omega(T)) = \mathcal{O}_\omega(T) \cap T(X) \times X^\omega$, we have $\sigma^{-1}(\mathcal{O}_\omega(T) \cap T(X) \times X^\omega) = \mathcal{O}_\omega(T)$, and thus $\nu(\mathcal{O}_\omega(T) \cap T(X) \times X^\omega) = \nu(\mathcal{O}_\omega(T)) = 1$. This allows us to assume $(x_1, x_2, \dots) \in \mathcal{O}_\omega(T) \cap T(X) \times X^\omega$ in the integrand in the right-hand side of (6.5). Now we fix an arbitrary orbit $(x_1, x_2, \dots) \in \mathcal{O}_\omega(T)$ with $x_1 \in T(X)$ and compute this integrand.

Fix an arbitrary $x_0 \in T^{-1}(x_1)$. Recall that $\mathcal{A}(x_0)$ refers to the element of $\mathcal{A}$ containing x_0. Since $\mathcal{P}$ is a backward conditional transition probability kernel of $\widehat{\nu}$

from X^ω to X supported by $\mathcal{O}_2(T)\times X^\omega$ and since $x_1\in T(X)$, the measure $\mathcal{P}_{(x_1,x_2,\dots)}$ is supported on $T^{-1}(x_1)$. Note that the diameter of $\mathcal{A}(x_0)$ is less than ϵ, an expansive constant for T, by Remark 6.2 we get that $\mathcal{A}(x_0)\cap T^{-1}(x_1)=\{x_0\}$. Consequently,

$$\mathcal{P}_{(x_1,x_2,\dots)}(\mathcal{A}(x_0))=\mathcal{P}_{(x_1,x_2,\dots)}(\{x_0\}). \tag{6.6}$$

Also, by the definition of the partitions $\widetilde{\mathcal{A}}^n_T$, we have $\widetilde{\mathcal{A}}^{n+1}_T(x_0,x_1,\dots)=\mathcal{A}(x_0)\times\mathcal{A}(x_1)\times\cdots\times\mathcal{A}(x_n)\times X^\omega\cap\mathcal{O}_\omega(T)=\mathcal{A}(x_0)\times\widetilde{\mathcal{A}}^n_T(x_1,x_2,\dots)\cap\mathcal{O}_\omega(T)$. Applying Proposition 6.8 to the Borel measurable function that assigns each $\underline{x}\in\mathcal{O}_\omega(T)$ the value $\mathcal{P}(\underline{x},\mathcal{A}(x_0))$, by Definition A.14 (b), we get

$$\begin{aligned}\mathcal{P}((x_1,x_2,\dots),\mathcal{A}(x_0))&=\lim_{n\to+\infty}\frac{\int_{\widetilde{\mathcal{A}}^n_T(x_1,x_2,\dots)}\mathcal{P}(\underline{x},\mathcal{A}(x_0))\,\mathrm{d}\nu(\underline{x})}{\nu\big(\widetilde{\mathcal{A}}^n_T(x_1,x_2,\dots)\big)}\\&=\lim_{n\to+\infty}\frac{\nu\big(\mathcal{A}(x_0)\times\widetilde{\mathcal{A}}^n_T(x_1,x_2,\dots)\cap\mathcal{O}_\omega(T)\big)}{\nu\big(\widetilde{\mathcal{A}}^n_T(x_1,x_2,\dots)\big)}\\&=\lim_{n\to+\infty}\frac{\nu\big(\widetilde{\mathcal{A}}^{n+1}_T(x_0,x_1,\dots)\big)}{\nu\big(\widetilde{\mathcal{A}}^n_T(x_1,x_2,\dots)\big)}.\end{aligned}\tag{6.7}$$

By (6.5), (6.6), and (6.7), we have

$$h_\nu(\sigma)=-\int_{\mathcal{O}_\omega(T)}\sum_{x_0\in T^{-1}(x_1)}\mathcal{P}_{(x_1,x_2,\dots)}(\{x_0\})\log\big(\mathcal{P}_{(x_1,x_2,\dots)}(\{x_0\})\big)\,\mathrm{d}\nu(x_1,x_2,\dots).$$

Therefore, by (6.4) we conclude $h_\nu(\sigma)=\int_{\mathcal{O}_\omega(T)}H(\mathcal{P}_{\underline{x}})\,\mathrm{d}\nu(\underline{x})$ as we want. □

Lemma 6.10. *Let T be a correspondence on a compact metric space (X,d), $\mathcal{Q}$ be a transition probability kernel on X supported by T, and μ be a $\mathcal{Q}$-invariant Borel probability measure on X. If $\mathcal{R}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^{[1]}$ from X to X supported on $\mathcal{O}_2(T)$, then the transition probability kernel $\widetilde{\mathcal{R}}$ from X^ω to X given by*

$$\widetilde{\mathcal{R}}((x_1,x_2,\dots),B)\coloneqq\mathcal{R}(x_1,B)\quad\text{for all }(x_1,x_2,\dots)\in X^\omega\text{ and }B\in\mathscr{B}(X)\tag{6.8}$$

is a backward conditional transition probability kernel of $\mu\mathcal{Q}^\omega$ from X^ω to X supported on $\mathcal{O}_2(T)\times X^\omega$.

Proof. To verify that $\widetilde{\mathcal{R}}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^\omega$ from $\mathcal{O}_\omega(T)$ to X supported on $\mathcal{O}_2(T)\times X^\omega$, we should check the properties (a) and (b) in Definition A.14.

First, because $\mathcal{R}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^{[1]}$ from X to X supported on $\mathcal{O}_2(T)$, by Definition A.14 (a), $\mathcal{R}_x$ is supported on $T^{-1}(x)$ for all $x\in T(X)$. For each $(x_1,x_2,\dots)\in T(x)\times X^\omega$, since $\mathcal{R}_{x_1}$ is supported on $T^{-1}(x_1)$, we have $\widetilde{\mathcal{R}}\big((x_1,x_2,\dots),T^{-1}(x_1)\big)=\mathcal{R}\big(x_1,T^{-1}(x_1)\big)=1$. Thus Definition A.14 (a) holds for $\widetilde{\mathcal{R}}$ as $T(X)\times X^\omega=\sigma'(\mathcal{O}_2(T)\times X^\omega)$.

By Remark A.15, we have $\mu\mathcal{Q}^{[1]}=\big(\mu\mathcal{R}^{[1]}\big)\circ\iota_2^{-1}$, so Lemma A.10 indicates $\mu\mathcal{Q}^{[n]}=\big(\mu\mathcal{R}^{[n]}\big)\circ\iota_{n+1}^{-1}$ for all $n\in\mathbb{N}$. Thus, by (5.7) and (A.4) in Lemma A.3, for every $n\in\mathbb{N}$

and every $A_0, A_1, \dots, A_n \in \mathscr{B}(X)$ we have

$$\begin{aligned}
\int_{A_1\times\cdots\times A_n\times X^\omega} \widetilde{\mathcal{R}}(\underline{x}, A_0)\,\mathrm{d}(\mu\mathcal{Q}^\omega)(\underline{x}) &= \int_{A_1\times\cdots\times A_n} \mathcal{R}(x_1, A_0)\,\mathrm{d}\big(\mu\mathcal{Q}^{[n-1]}\big)(x_1,\dots,x_n) \\
&= \int_{A_n\times\cdots\times A_1} \mathcal{R}(x_1, A_0)\,\mathrm{d}\big(\mu\mathcal{R}^{[n-1]}\big)(x_n,\dots,x_1) \\
&= \big(\mu\mathcal{R}^{[n]}\big)(A_n\times\cdots\times A_0) \\
&= \big(\mu\mathcal{Q}^{[n]}\big)(A_0\times\cdots\times A_n) \\
&= (\mu\mathcal{Q}^\omega)(A_0\times\cdots\times A_n\times X^\omega).
\end{aligned}$$

This is equivalent to the property (b) in Definition A.14 for $\widetilde{\mathcal{R}}$ by the Dynkin's π-λ Theorem, corresponding to (b2) in Remark A.12. Therefore, $\widetilde{\mathcal{R}}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^\omega$ from X^ω to X supported on $\mathcal{O}_2(T)\times X^\omega$. □

Theorem 6.11. *Let T be a forward expansive correspondence on a compact metric space (X, d), $\mathcal{Q}$ be a transition probability kernel on X supported by T, and μ be a $\mathcal{Q}$-invariant Borel probability measure on X. If $\mathcal{R}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^{[1]}$ from X to X supported on $\mathcal{O}_2(T)$, then we have*

$$h_\mu(\mathcal{Q}) = \int_X H(\mathcal{R}_x)\,\mathrm{d}\mu(x).$$

Proof. First, by Lemma 6.10, the transition probability kernel $\widetilde{\mathcal{R}}$ given by (6.8) is a backward conditional transition probability kernel of $\mu\mathcal{Q}^\omega$ from X^ω to X supported on $\mathcal{O}_2(T)\times X^\omega$.

Thus by Proposition 6.9 and Theorem 5.24,

$$\begin{aligned}
h_\mu(\mathcal{Q}) = h_{\mu\mathcal{Q}^\omega|_T}(\sigma) &= \int_{\mathcal{O}_\omega(T)} H\big(\widetilde{\mathcal{R}}_{(x_1,x_2,\dots)}\big)\,\mathrm{d}(\mu\mathcal{Q}^\omega|_T)(x_1,x_2,\dots) \\
&= \int_{\mathcal{O}_\omega(T)} H(\mathcal{R}_{x_1})\,\mathrm{d}(\mu\mathcal{Q}^\omega|_T)(x_1,x_2,\dots) \\
&= \int_X H(\mathcal{R}_{x_1})\,\mathrm{d}\mu(x_1),
\end{aligned}$$

where the last equality follows from taking $n=0$ in (5.7).

Therefore $h_\mu(\mathcal{Q}) = \int_X H(\mathcal{R}_x)\,\mathrm{d}\mu(x)$. □

Remark 6.12. If the forward expansive correspondence T in Theorem 6.11 is induced by a single-valued forward expansive map f, i.e., $T = \mathcal{C}_f$, we can conclude the following statement, which is equivalent to the Rokhlin formula (see for example, [PU10, Theorem 2.9.7])

Let X be a compact metric space, $f\colon X\to X$ be a forward expansive continuous map, and μ be an f-invariant Borel probability measure on X. If $\mathcal{R}$ is a transition probability kernel on X that satisfies $\mu\big(A\cap f^{-1}(B)\big) = \int_B \mathcal{R}(x, A)\,\mathrm{d}\mu(x)$ for all A, $B\in$

$\mathscr{B}(X)$, then we have

$$h_\mu(f) = \int_X H(\mathcal{R}_x)\,\mathrm{d}\mu(x). \tag{6.9}$$

6.4. **Proof of Theorem D.** We aim to establish Theorem D for general correspondences in this subsection.

Let (X,d) be a compact metric space. Recall that a transition probability kernel $\mathcal{Q}$ on X is *supported by* a correspondence T on X if $\mathcal{Q}(x,T(x)) = 1$ holds for all $x \in X$.

Lemma 6.13. *Let $\mathcal{Q}$ be a transition probability kernel on a compact metric space (X,d) supported by a correspondence T on X and μ be a Borel probability measure on X. Then the measure $\mu\mathcal{Q}^{[n-1]}$ is supported on $\mathcal{O}_n(T)$ for every $n \in \mathbb{N}$, and the measure $\mu\mathcal{Q}^{\omega}$ is supported on $\mathcal{O}_\omega(T)$.*

Proof. First we show that for each $n \in \mathbb{N}$ and each point $x \in X$, we have

$$\mathcal{Q}^{[n-1]}(x,\mathcal{O}_n(T)) = 1. \tag{6.10}$$

If $n = 1$, the equality (6.10) holds because $\mathcal{Q}^{[0]}(x,\mathcal{O}_1(T)) = \widehat{\mathrm{id}_X}(x,X) = 1$.

Now suppose that (6.10) holds for some $n \in \mathbb{N}$. By (5.4), we have

$$\begin{aligned}
&\mathcal{Q}^{[n]}(x,\mathcal{O}_{n+1}(T)) \\
&\quad= \int_{(x_1,\dots,x_n)\in X^n} \mathcal{Q}(x_n, \pi_{n+1}(x_1,\dots,x_n;\mathcal{O}_{n+1}(T)))\,\mathrm{d}\mathcal{Q}_x^{[n-1]}(x_1,\dots,x_n) \\
&\quad= \int_{(x_1,\dots,x_n)\in \mathcal{O}_n(T)} \mathcal{Q}(x_n, T(x_n))\,\mathrm{d}\mathcal{Q}_x^{[n-1]}(x_1,\dots,x_n) \\
&\quad= \int_{(x_1,\dots,x_n)\in \mathcal{O}_n(T)} \mathrm{d}\mathcal{Q}_x^{[n-1]}(x_1,\dots,x_n) \\
&\quad= \mathcal{Q}^{[n-1]}(x,\mathcal{O}_n(T)) \\
&\quad= 1,
\end{aligned}$$

where the second-to-last equality follows from $\mathcal{Q}(x_n,T(x_n)) = 1$ because $\mathcal{Q}$ is supported by T. Hence by induction, we get that the equality (6.10) holds for all $n \in \mathbb{N}$.

For each $n \in \mathbb{N}$, by (6.10) and Definition 5.6, we have

$$\big(\mu\mathcal{Q}^{[n-1]}\big)(\mathcal{O}_n(T)) = \int_X \mathcal{Q}^{[n-1]}(x,\mathcal{O}_n(T))\,\mathrm{d}\mu(x) = \int_X \mathrm{d}\mu = 1.$$

By (5.7), we get

$$\begin{aligned}
(\mu\mathcal{Q}^{\omega})(\mathcal{O}_\omega(T)) = (\mu\mathcal{Q}^{\omega})\Bigg(\bigcap_{n=1}^{+\infty} \mathcal{O}_n(T)\times X^{\omega}\Bigg) &= \lim_{n\to+\infty} (\mu\mathcal{Q}^{\omega})\big(\mathcal{O}_n(T)\times X^{\omega}\big) \\
&= \lim_{n\to+\infty} \big(\mu\mathcal{Q}^{[n-1]}\big)(\mathcal{O}_n(T)) = \lim_{n\to+\infty} 1 = 1.
\end{aligned}$$

Therefore $\mu\mathcal{Q}^{[n-1]}$ is supported on $\mathcal{O}_n(T)$ for every $n \in \mathbb{N}$, and $\mu\mathcal{Q}^{\omega}$ is supported on $\mathcal{O}_\omega(T)$. $\square$

Remark 6.14. Suppose that $\mathcal{Q}$ is supported by T. Denote by $\mu\mathcal{Q}^\omega|_T$ the restricted measure of $\mu\mathcal{Q}^\omega$ on $\mathcal{O}_\omega(T)$. Since $(\mu\mathcal{Q}^\omega)(\mathcal{O}_\omega(T)) = 1$ and $\mathcal{O}_\omega(T)$ is forward-invariant under $\sigma'\colon X^\omega \to X^\omega$, the measure-preserving system $(\mathcal{O}_\omega(T), \mathscr{B}(\mathcal{O}_\omega(T)), \mu\mathcal{Q}^\omega|_T, \sigma)$ is isomorphic to $\left(X^\omega, \mathscr{B}(X^\omega), \mu\mathcal{Q}^\omega, \sigma'\right)$, so their entropies are equal. Thus we can rewrite Theorem 5.24 as

$$h_\mu(\mathcal{Q}) = h_{\mu\mathcal{Q}^\omega|_T}(\sigma). \tag{6.11}$$

Recall the projection maps $\widetilde{\pi}_1$, $\widetilde{\pi}_2$, and $\widetilde{\pi}_{12}$ given in (2.5). Let T be a correspondence on a compact metric space (X, d). In Proposition A.11, if $X_1 = X_2 = X$ and $M = \mathcal{O}_2(T)$, then for a Borel probability measure μ on X and a transition probability kernel $\mathcal{Q}$ on X, the property (a) in Proposition A.11 means that $\mathcal{Q}$ is supported by T, and the property (b) means that $\mu\mathcal{Q}^{[1]} = \nu$. Proposition A.11 also indicates $\mu = \nu \circ \widetilde{\pi}_1^{-1}$. Consequently, we get the following proposition.

Proposition 6.15. *Let T be a correspondence on a compact metric space (X, d) and ν be a Borel probability measure on X^2 supported on $\mathcal{O}_2(T)$. Then there exists a transition probability kernel $\mathcal{Q}$ on X supported by T such that $(\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]} = \nu$.*

Moreover, if two transition probability kernels $\mathcal{Q}$ and $\mathcal{Q}'$ on X satisfy $\nu = (\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]} = (\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}'^{[1]}$, then for each $A \in \mathscr{B}(X)$, $\mathcal{Q}(x, A) = \mathcal{Q}'(x, A)$ holds for $\nu \circ \widetilde{\pi}_1^{-1}$-almost every $x \in X$.

Lemma 6.16. *Let T be a correspondence on a compact metric space X and ν be a σ-invariant Borel probability measure on $\mathcal{O}_\omega(T)$. Then there exists a transition probability kernel $\mathcal{Q}$ on X supported by T such that $\nu \circ \widetilde{\pi}_1^{-1} \in \mathcal{M}(X, \mathcal{Q})$ and $(\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]} = \nu \circ \widetilde{\pi}_{12}$.*

Proof. The measure $\nu \circ \widetilde{\pi}_{12}^{-1}$ is a Borel probability measure on X^2 supported on $\mathcal{O}_2(T)$ because the image of $\widetilde{\pi}_{12}$ lies in $\mathcal{O}_2(T)$. By Proposition 6.15, we can choose a transition probability kernel $\mathcal{Q}$ on X supported by T such that $\nu \circ \widetilde{\pi}_{12}^{-1} = ((\nu \circ \widetilde{\pi}_{12}^{-1}) \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]} = (\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]}$. Since ν is σ-invariant and $\left((\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]}\right) \circ \widetilde{\pi}_2^{-1} = (\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}$ from (A.10), we have

$$\begin{aligned}(\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q} &= \left((\nu \circ \widetilde{\pi}_1^{-1})\mathcal{Q}^{[1]}\right) \circ \widetilde{\pi}_2^{-1} = (\nu \circ \widetilde{\pi}_{12}^{-1}) \circ \widetilde{\pi}_2^{-1} \\ &= \nu \circ \widetilde{\pi}_2^{-1} = (\nu \circ \sigma^{-1}) \circ \widetilde{\pi}_1^{-1} = \nu \circ \widetilde{\pi}_1^{-1}.\end{aligned}$$

Therefore, $\nu \circ \widetilde{\pi}_1^{-1} = \mu$ is $\mathcal{Q}$-invariant. □

Now let us give the proof of Theorem D.

Proof of Theorem D. (i) Since σ is a continuous transformation on the compact metric space $\mathcal{O}_\omega(T)$, by the Bogolyubov–Krylov Theorem (see for example, [PU10, Theorem 3.1.8]), we can choose a Borel probability measure ν on $\mathcal{O}_\omega(T)$ that is invariant under σ. Then Theorem D (i) follows from Lemma 6.16.

(ii) Let $\mathcal{Q}$ be an arbitrary transition probability kernel on X supported by T and μ be an arbitrary $\mathcal{Q}$-invariant probability measure on X.

By Remark 5.8 and (A.10), we have $\mathcal{Q}_x^{[1]} \circ \widetilde{\pi}_2^{-1} = \left(\delta_x \mathcal{Q}^{[1]}\right) \circ \widetilde{\pi}_2^{-1} = \delta_x \mathcal{Q} = \mathcal{Q}_x$.

We apply the (classical) Variational Principle to the dynamical system $(\mathcal{O}_\omega(T), \sigma)$:

$$P\big(\sigma, \widetilde{\phi}\big) = \sup_{\nu} \bigg\{ h_\nu(\sigma) + \int_{\mathcal{O}_\omega(T)} \widetilde{\phi} \, \mathrm{d}\nu \bigg\},$$

where the measure ν ranges over all σ-invariant Borel probability measures on $\mathcal{O}_\omega(T)$.

Since $\mu\mathcal{Q}^\omega|_T$ is a σ-invariant probability measure on $\mathcal{O}_\omega(T)$ for arbitrary μ and $\mathcal{Q}$, we get

$$P\big(\sigma, \widetilde{\phi}\big) \geqslant \sup_{\mathcal{Q},\mu} \bigg\{ h_{\mu\mathcal{Q}^\omega|_T}(\sigma) + \int_{\mathcal{O}_\omega(T)} \widetilde{\phi} \, \mathrm{d}(\mu\mathcal{Q}^\omega|_T) \bigg\}.$$

Recall $P(T, \phi) = P\big(\sigma, \widetilde{\phi}\big)$ from Theorem 4.9 and $h_\mu(\mathcal{Q}) = h_{\mu\mathcal{Q}^\omega|_T}(\sigma)$ from (6.11). We can rewrite the inequality above as

$$P(T, \phi) \geqslant \sup_{\mathcal{Q},\mu} \bigg\{ h_\mu(\mathcal{Q}) + \int_{\mathcal{O}_\omega(T)} \widetilde{\phi} \, \mathrm{d}(\mu\mathcal{Q}^\omega|_T) \bigg\}.$$

Since $\mathcal{Q}_x(T(x)) = 1$ for all $x \in X$ and since $\mu\mathcal{Q}^\omega(\mathcal{O}_\omega(T)) = 1$ from Lemma 6.13, the equality (A.11) can be written as

$$\int_{\mathcal{O}_\omega(T)} \widetilde{\phi} \, \mathrm{d}(\mu\mathcal{Q}^\omega) = \int_X \int_{T(x_1)} \phi(x_1, x_2) \, \mathrm{d}\mathcal{Q}_{x_1}(x_2) \, \mathrm{d}\mu(x_1).$$

Therefore we get

$$P(T, \phi) \geqslant \sup_{\mathcal{Q},\mu} \bigg\{ h_\mu(\mathcal{Q}) + \int_X \int_{T(x_1)} \phi(x_1, x_2) \, \mathrm{d}\mathcal{Q}_{x_1}(x_2) \, \mathrm{d}\mu(x_1) \bigg\},$$

i.e., we complete the proof of statement (ii). $\square$

In addition, the following result is useful in Subsection 3.1, and its proof uses the similar techniques in this subsection.

Proposition 6.17. *Let T be a correspondences on a compact metric space X, $Y \in \mathcal{F}(X)$ such that $T|_Y$ is a correspondence on Y, and $\phi \in C(\mathcal{O}_2(X), \mathbb{R})$. Assume that for each Borel probability measure μ on X with the property that there exists a transition probability kernel $\mathcal{Q}$ on X such that μ is $\mathcal{Q}$-invariant, we have $\mu(Y) = 1$. Then $P(T, \phi) = P(T|_Y, \phi|_{\mathcal{O}_2(T|_Y)})$.*

Proof. Denote by σ_X the shift map on $\mathcal{O}_\omega(T)$, by $\sigma_Y \coloneqq \sigma_X|_{Y^\omega \cap \mathcal{O}_\omega(X)}$ the shift map on $\mathcal{O}_\omega(T|_Y) = Y^\omega \cap \mathcal{O}_\omega(X)$, by $\widetilde{\phi}_X \in C(\mathcal{O}_\omega(T), \mathbb{R})$ given by $\widetilde{\phi}_X(x_1, x_2, \dots) \coloneqq \phi(x_1, x_2)$ for all $(x_1, x_2, \dots) \in \mathcal{O}_\omega(T)$, and by $\widetilde{\phi}_Y \in C(\mathcal{O}_\omega(T|_Y), \mathbb{R})$ given by $\widetilde{\phi}_Y(x_1, x_2, \dots) \coloneqq \phi(x_1, x_2)$ for all $(x_1, x_2, \dots) \in \mathcal{O}_\omega(T|_Y)$. By Theorem 4.9, we have $P(T, \phi) = P\big(\sigma_X, \widetilde{\phi}_X\big)$ and $P(T|_Y, \phi|_{\mathcal{O}_2(T|_Y)}) = P\big(\sigma_Y, \widetilde{\phi}_Y\big)$.

Fix an arbitrary σ_X-invariant Borel probability measure ν on $\mathcal{O}_\omega(T)$. By Lemma 6.16, the measure μ on X given by $\mu(A) \coloneqq \nu(A \times X^\omega \cap \mathcal{O}_\omega(T))$ for all $A \in \mathscr{B}(X)$ is $\mathcal{Q}$-invariant for some transition probability kernel $\mathcal{Q}$ on X, so $\mu(Y) = 1$, i.e., $\nu(Y \times X^\omega \cap \mathcal{O}_\omega(T)) = 1$. Since ν is σ_X-invariant, we have $\nu(X^n \times Y \times X^\omega \cap \mathcal{O}_\omega(T)) = \nu(\sigma_X^{-n}(Y \times X^\omega \cap \mathcal{O}_\omega(T))) = \nu(Y \times X^\omega \cap \mathcal{O}_\omega(T)) = 1$ holds for all $n \in \mathbb{N}_0$. Consequently, we have $\nu(\mathcal{O}_\omega(T|_Y)) = \nu(Y^\omega \cap \mathcal{O}_\omega(T)) = \nu(\cap_{n=0}^{+\infty}(X^n \times Y \times X^\omega \cap \mathcal{O}_\omega(T))) = 1$.

We have proved that $\nu(\mathcal{O}_\omega(T|_Y)) = 1$ holds for all σ_X-invariant Borel probability measure ν on $\mathcal{O}_\omega(T)$. Since $\widetilde{\phi}_Y = \widetilde{\phi}_X|_{\mathcal{O}_\omega(T|_Y)}$, applying the classical Variational Principle (B.7) for σ_X with the potential $\widetilde{\phi}_X$ and σ_Y with the potential $\widetilde{\phi}_Y$, we get $P\big(\sigma_X, \widetilde{\phi}_X\big) = P\big(\sigma_Y, \widetilde{\phi}_Y\big)$. Recall $P(T, \phi) = P\big(\sigma_X, \widetilde{\phi}_X\big)$ and $P(T|_Y, \phi|_{\mathcal{O}_2(T|_Y)}) = P\big(\sigma_Y, \widetilde{\phi}_Y\big)$, so we conclude $P(T, \phi) = P(T|_Y, \phi|_{\mathcal{O}_2(T|_Y)})$. $\square$

6.5. **Proof of Theorem A.** We aim to proceed with the proof of Theorem A for forward expansive correspondences in this subsection.

For a correspondence T on a compact metric space (X, d), recall that $\widetilde{\pi}_{12} \colon \mathcal{O}_\omega(T) \to X^2$ is the projection given by $\widetilde{\pi}_{12}(x_1, x_2, \dots) = (x_1, x_2)$. For a Borel probability measure ν on $\mathcal{O}_\omega(T)$, set $\nu_{12} \coloneqq \nu \circ \widetilde{\pi}_{12}^{-1}$, a Borel probability measure on X^2. Note that ν_{12} is supported on $\mathcal{O}_2(T)$ because the image of $\widetilde{\pi}_{12}$ lies in $\mathcal{O}_2(T)$. The following proposition is useful to prove Theorem A.

Proposition 6.18. *Let T be a forward expansive correspondence on a compact metric space (X, d) with an expansive constant $\epsilon > 0$ and ν be a σ-invariant Borel probability measure on $\mathcal{O}_\omega(T)$. If a transition probability kernel $\mathcal{Q}$ on X supported by T and a $\mathcal{Q}$-invariant Borel probability measure μ on X satisfying $\nu_{12} = \mu\mathcal{Q}^{[1]}$, then $h_\nu(\sigma) \leqslant h_\mu(\mathcal{Q})$.*

Proof. Recall $\widehat{\nu}$ be the measure on X^ω given by $\widehat{\nu}(A) = \nu(A \cap \mathcal{O}_\omega(T))$ for all $A \in \mathscr{B}(X^\omega)$. Let $\mathcal{P}$ be a backward conditional transition probability kernel of $\widehat{\nu}$ from X^ω to X supported on $\mathcal{O}_2(T) \times X^\omega$ and $\mathcal{S}$ be a forward conditional transition probability kernel of $\widehat{\nu}$ from X to X^ω supported on $\mathcal{O}_\omega(T)$. Proposition 6.9 indicates

$$h_\nu(\sigma) = \int_{\mathcal{O}_\omega(T)} H(\mathcal{P}_{\underline{x}}) \,\mathrm{d}\nu(\underline{x}).$$

By Corollary A.4, $\nu_{12} = \mu\mathcal{Q}^{[1]}$ implies $\mu = \nu_{12} \circ \widetilde{\pi}_1^{-1} = (\nu \circ \widetilde{\pi}_{12}^{-1}) \circ \widetilde{\pi}_1^{-1} = \nu \circ \widetilde{\pi}_1^{-1}$. Applying (A.15) in Remark A.12 (b3) for $\mathcal{S}$ we get

$$h_\nu(\sigma) = \int_X \bigg(\int_{\mathcal{O}_\omega(T)} H(\mathcal{P}_{(x_1, x_2, \dots)}) \,\mathrm{d}\mathcal{S}_{x_1}(x_2, x_3, \dots) \bigg) \,\mathrm{d}\mu(x_1). \tag{6.12}$$

Recall from Remark 6.2 that $T^{-1}(x_1)$, on which the measure $\mathcal{P}_{(x_1, x_2, \dots)}$ are supported, is a finite set. Since the map $x \mapsto x \log x$ is a convex function for $x \in [0, 1]$, by (6.4) and Jensen's inequality we have for each $x_1 \in X$,

$$\begin{aligned} &\int_{\mathcal{O}_\omega(T)} H(\mathcal{P}_{(x_1, x_2, \dots)}) \,\mathrm{d}\mathcal{S}_{x_1}(x_2, x_3, \dots) \\ &\quad = - \sum_{x_0 \in T^{-1}(x_1)} \int_{\mathcal{O}_\omega(T)} \mathcal{P}((x_1, x_2, \dots), \{x_0\}) \log(\mathcal{P}((x_1, x_2, \dots), \{x_0\})) \,\mathrm{d}\mathcal{S}_{x_1}(x_2, x_3, \dots) \\ &\quad \leqslant - \sum_{x_0 \in T^{-1}(x_1)} \mathcal{R}(x_1, \{x_0\}) \log(\mathcal{R}(x_1, \{x_0\})), \end{aligned} \tag{6.13}$$

where $\mathcal{R}(x_1,\{x_0\}) \coloneqq \int_{\mathcal{O}_\omega(T)} \mathcal{P}((x_1,x_2,\dots),\{x_0\})\,\mathrm{d}\mathcal{S}_{x_1}(x_2,x_3,\dots)$ for all $(x_0,x_1) \in \mathcal{O}_2(T)$. Moreover, for each $x_1 \in X$ and $A \in \mathscr{B}(X)$, define

$$\mathcal{R}(x_1,A) \coloneqq \int_{\mathcal{O}_\omega(T)} \mathcal{P}((x_1,x_2,\dots),A)\,\mathrm{d}\mathcal{S}_{x_1}(x_2,x_3,\dots). \tag{6.14}$$

We can check that this $\mathcal{R}$ is a transition probability kernel on X. Now we verify that $\mathcal{R}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^{[1]}$ from X to X supported on $\mathcal{O}_2(T)$.

First, recall that $\mathcal{P}$ is a backward conditional transition probability kernel of $\widehat{\nu}$ from X^ω to X supported on $\mathcal{O}_2(T) \times X^\omega$. By Definition A.14 (a), we have $\mathcal{P}((x_1,x_2,\dots),T^{-1}(x_1)) = 1$ for all $(x_1,x_2,\dots) \in T(X) \times X^\omega$. By (6.14), we have $\mathcal{R}(x_1,T^{-1}(x_1)) = 1$ for all $x_1 \in T(X)$.

Second, applying (A.15) for $\mathcal{S}$, we have

$$\begin{aligned}\int_B \mathcal{R}(x_1,A)\,\mathrm{d}\mu(x_1) &= \int_B \biggl(\int_{\mathcal{O}_\omega(T)} \mathcal{P}((x_1,x_2,\dots),A)\,\mathrm{d}\mathcal{S}_{x_1}(x_2,x_3,\dots)\biggr)\,\mathrm{d}\mu(x_1) \\ &= \int_{B\times X^\omega \cap \mathcal{O}_\omega(T)} \mathcal{P}((x_1,x_2,\dots),A)\,\mathrm{d}\nu(x_1,x_2,\dots).\end{aligned}$$

Recall that $\mathcal{P}$ is a backward conditional transition probability kernel of $\widehat{\nu}$ from X^ω to X supported on $\mathcal{O}_2(T) \times X^\omega$. Definition A.14 (b) implies

$$\nu(A\times B\times X^\omega \cap \mathcal{O}_\omega(T)) = \int_{B\times X^\omega\cap\mathcal{O}_\omega(T)} \mathcal{P}((x_1,x_2,\dots),A)\,\mathrm{d}\nu(x_1,x_2,\dots).$$

Hence, $\int_B \mathcal{R}(x_1,A)\,\mathrm{d}\mu(x_1) = \nu(A\times B\times X^\omega\cap\mathcal{O}_\omega(T)) = \nu_{12}(A\times B) = \bigl(\mu\mathcal{Q}^{[1]}\bigr)(A\times B)$. Applying the Dynkin's π-λ Theorem (see the equivalence between the properties (b) and (b1) in Remark A.12), we conclude that $\mathcal{R}$ is a backward conditional transition probability kernel of $\mu\mathcal{Q}^{[1]}$ from X to X supported on $\mathcal{O}_2(T)$ by Definition A.14. Thus Theorem 6.11 indicates $h_\mu(\mathcal{Q}) = \int_X H(\mathcal{R}_x)\,\mathrm{d}\mu(x)$. By (6.13), we get

$$\begin{aligned}h_\mu(\mathcal{Q}) &= \int_X \sum_{x_0\in T^{-1}(x_1)} -\mathcal{R}(x_1,\{x_0\})\log(\mathcal{R}(x_1,\{x_0\}))\,\mathrm{d}\mu(x_1) \\ &\geqslant \int_X \biggl(\int_{\mathcal{O}_\omega(T)} H\bigl(\mathcal{P}_{(x_1,x_2,\dots)}\bigr)\,\mathrm{d}\mathcal{S}_{x_1}(x_2,x_3,\dots)\biggr)\,\mathrm{d}\mu(x_1).\end{aligned}$$

Therefore by (6.12) we conclude $h_\nu(\sigma) \leqslant h_\mu(\mathcal{Q})$. □

With all the preparations in previous subsections, we are now ready to prove Theorem A.

Recall that a pair $(\mu,\,\mathcal{Q})$ consisting of a transition probability kernel $\mathcal{Q}$ on X supported by T and a $\mathcal{Q}$-invariant Borel probability measure μ on X is called an equilibrium state for the correspondence T and the potential function ϕ if it satisfies the equality (1.1).

Proof of Theorem A (1) and (2). By Proposition D.1, the forward expansiveness of T implies that the shift map $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ is forward expansive. By applying

[PU10, Theorem 3.5.6] for the forward expansive dynamical system $(\mathcal{O}_\omega(T),\sigma)$ with the potential $\widetilde{\phi}\colon \mathcal{O}_\omega(T)\to\mathbb{R}$, we can choose a σ-invariant Borel probability measure ν on $\mathcal{O}_\omega(T)$ such that the following equality holds:

$$P\bigl(\sigma,\widetilde{\phi}\bigr)=h_\nu(\sigma)+\int_{\mathcal{O}_\omega(T)}\widetilde{\phi}\,\mathrm{d}\nu.$$

We choose a transition probability kernel $\mathcal{Q}$ on X supported by T and a $\mathcal{Q}$-invariant Borel probability measure μ on X such that $\nu_{12}=\mu\mathcal{Q}^{[1]}$ (the existence of this choice is ensured by Proposition 6.15). Then Proposition 6.18 indicates that $h_\nu(\sigma)\leqslant h_\mu(\mathcal{Q})$.

By Theorem 4.9 and (A.11), we have

$$\begin{aligned}P(T,\phi)&=P\bigl(\sigma,\widetilde{\phi}\bigr)\\&=h_\nu(\sigma)+\int_{\mathcal{O}_\omega(T)}\widetilde{\phi}\,\mathrm{d}\nu\\&\leqslant h_\mu(\mathcal{Q})+\int_{\mathcal{O}_2(T)}\phi(x_1,x_2)\,\mathrm{d}\nu_{12}\\&=h_\mu(\mathcal{Q})+\int_{\mathcal{O}_2(T)}\phi(x_1,x_2)\,\mathrm{d}\bigl(\mu\mathcal{Q}^{[1]}\bigr)\\&=h_\mu(\mathcal{Q})+\int_X\int_{T(x_1)}\phi(x_1,x_2)\,\mathrm{d}\mathcal{Q}_{x_1}(x_2)\,\mathrm{d}\mu(x_1).\end{aligned}$$

By Theorem D, we have $P(T,\phi)\geqslant h_\mu(\mathcal{Q})+\int_X\int_{T(x_1)}\phi(x_1,x_2)\,\mathrm{d}\mathcal{Q}_{x_1}(x_2)\,\mathrm{d}\mu(x_1)$. Thus we have $P(T,\phi)=h_\mu(\mathcal{Q})+\int_X\int_{T(x_1)}\phi(x_1,x_2)\,\mathrm{d}\mathcal{Q}_{x_1}(x_2)\,\mathrm{d}\mu(x_1)$. Together with Theorem D, we manage to prove Theorem A except for $P(T,\phi)\in\mathbb{R}$. Now we show it.

First, Remark 4.7 indicates that $P(T,\phi)>-\infty$.

Recall from Remark 6.2 that there exists $M\in\mathbb{N}$ such that $\#T(x)\leq M$ for all $x\in X$. Suppose that a transition probability kernel $\mathcal{R}$ on X satisfies the two properties (a) and (b2) in Lemma 6.10. Then Theorem 6.11 indicates that $h_\mu(\mathcal{Q})=\int_X H(\mathcal{R}_x)\,\mathrm{d}\mu(x)$. Since for each $x\in X$, the transition probability kernel $\mathcal{R}_x$ is supported on $T^{-1}(x)$ (property (a) in Lemma 6.10), we have $H(\mathcal{R}_x)=-\sum_{y\in T^{-1}(x)}\mathcal{R}_x(\{y\})\log(\mathcal{R}_x(\{y\}))\leqslant\log M$ for all $x\in X$. As a result, $h_\mu(\mathcal{Q})\leqslant\log M$, and thus $P(T,\phi)=h_\mu(\mathcal{Q})+\int_X\int_{T(x_1)}\phi(x_1,x_2)\,\mathrm{d}\mathcal{Q}_{x_1}(x_2)\,\mathrm{d}\mu(x_1)\leqslant\log M+\|\phi\|_\infty<+\infty$. $\square$

Let T be a forward expansive correspondence on a compact metric space X. If the potential function is identically zero, then Theorem A (2) suggests that there exists a transition probability kernel $\mathcal{Q}$ supported by T and a $\mathcal{Q}$-invariant probability measure μ on X such that $h(T)=h_\mu(\mathcal{Q})$. One can show that $h(T)$ and $h_\mu(\mathcal{Q})$ are both non-negative, so only in the case that $h(T)>0$ is the equality $h(T)=h_\mu(\mathcal{Q})$ non-trivial. There have been some results that show $h(T)>0$ for some kinds of correspondences T, for example, [PV17, Theorem C] and [RT18, Theorem 3.3]. Moreover, under their restrictions on T, we conclude $h_\mu(\mathcal{Q})>0$.

7. Thermodynamic formalism for correspondences

In this section, we develop thermodynamic formalism in two different settings for forward expansive correspondence T on a compact metric space X with a continuous potential function $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$.

In the first version, we assume that T has the specification property (see Definition 7.1) and that ϕ is Bowen summable (see Definition 7.3), then the Variational Principle holds, the equilibrium state exists and is unique in the sense of Theorem B, and the unique equilibrium state can be obtained by the eigenvectors of the Ruelle operator and its adjoint operator (see Theorem B for precise statements).

In the second version, we assume that T is distance-expanding (see Definition 7.5), open (see Definition 7.11), and strongly transitive (see Definition 7.12) and ϕ is Hölder continuous, then similar results hold and in addition, we get some equidistribution properties (see Theorem C for precise statements).

7.1. **Specification property and Bowen summability.** First, we introduce the specification property for correspondences. The notion of specification for correspondences or set-valued maps has been discussed by B. E. Raines, T. Tennant [RT18], as well as W. Cordeiro and M. J. Pacífico [CP16]. But in order to ensure Appendix B.3 (ii) and Proposition D.3, we give a definition for this notion with subtle differences from theirs and slightly stronger than the specification property given in [CP16, Definition 5.1].

Definition 7.1 (Specification property)**.** We say that a correspondence T on a compact metric space (X, d) has the *specification property* if, for an arbitrary $\epsilon > 0$, there exists $M \in \mathbb{N}$ depending only on ϵ with the following property:

For arbitrary $n \in \mathbb{N}$, $x_0^1, \dots, x_0^n \in X$, $m_1, \dots, m_n, p_1, \dots, p_n \in \mathbb{N}$ with $p_j > M$ for every $j \in (n]$, and an orbit $\bigl(x_0^j, x_1^j, \dots, x_{m_j-1}^j\bigr) \in \mathcal{O}_{m_j}(T)$ for every $j \in (n]$, there exists an orbit $z = (z_0, z_1, \dots) \in \mathcal{O}_\omega(T)$ such that $d\bigl(z_{m(j-1)+i}, x_i^j\bigr) < \epsilon$ for all $j \in (n]$ and $i \in [m_j - 1]$, where $m(j) \coloneqq \sum_{k=1}^{j}(m_k + p_k)$.

Recall D. Ruelle's definition of specification property of a continuous map from [Ru92, Section 1]:

Definition 7.2 (Ruelle's specification property)**.** We say that a continuous map $f\colon X \to X$ on a compact space (X, d) has the *specification property* if, for an arbitrary $\epsilon > 0$, there exists $M \in \mathbb{N}$ depending only on ϵ with the following property:

For arbitrary $n \in \mathbb{N}$, $x_1, \dots, x_n \in X$, and $m_1, \dots, m_n, p_1, \dots, p_n \in \mathbb{N}$ with $p_j > M$ for every $j \in (n]$, there exists a point $z \in X$ such that $d(f^{m(j-1)+i}(z), f^i(x_j)) < \epsilon$ for all $j \in (n]$ and $i \in [m_j - 1]$, where $m(j) \coloneqq \sum_{k=1}^{j}(m_k + p_k)$.

Proposition D.3 states that the specification property of a correspondence T in the sense of Definition 7.1 implies the specification property of the corresponding shift map $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ in the sense of Definition 7.2. This proposition corresponds to [RT18, Theorem 4.1], with a similar proof. For the convenience of our reader, we include a proof in Appendix D due to the subtle differences between Definition 7.1 and the definition of specification property for correspondences in [RT18].

Definition 7.3 (Bowen summability)**.** Let T be a forward expansive correspondence on a compact metric space (X, d). For a bounded Borel measurable function $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$, denote

$$\begin{aligned} K_{\phi,T}(\delta, n) \coloneqq \sup\Biggl\{ \Biggl| \sum_{k=1}^{n} (\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1})) \Biggr| : \\ (x_1, \dots, x_{n+1}), (y_1, \dots, y_{n+1}) \in \mathcal{O}_{n+1}(T) \\ \text{satisfying } d(x_k, y_k) < \delta \text{ for all } k \in (n+1] \Biggr\} \end{aligned} \tag{7.1}$$

for each $n \in \mathbb{N}$ and each $\delta > 0$.

Choose an expansive constant $\epsilon > 0$ of T, write $K_{\phi,T}(\epsilon) \coloneqq \sup\{K_{\phi,T}(\epsilon, n) : n \in \mathbb{N}\}$, and define $\mathcal{V}_T \coloneqq \{\phi : K_{\phi,T}(\epsilon) < +\infty\}$. Functions in $\mathcal{V}_T$ are called *Bowen summable* with respect to T.

The notation $\mathcal{V}_T$ above does not contain ϵ because it does not depend on ϵ, which we will prove in Proposition 7.4.

Proposition 7.4. *Let T be a forward expansive correspondence on a compact metric space (X, d), ϵ_1, $\epsilon_2 > 0$ be two expansive constants of T with $\epsilon_1 < \epsilon_2$, and $\phi \in B(\mathcal{O}_2(T), \mathbb{R})$. There exists $L \in \mathbb{N}$ such that for each $n \in \mathbb{N}$ with $n > L$, we have*

$$K_{\phi,T}(\epsilon_1, n) \leqslant K_{\phi,T}(\epsilon_2, n) \leqslant K_{\phi,T}(\epsilon_1, n - L) + 2L\|\phi\|_\infty. \tag{7.2}$$

Proof. First, since $\epsilon_1 < \epsilon_2$, if $(x_1, \dots, x_{n+1}), (y_1, \dots, y_{n+1}) \in \mathcal{O}_{n+1}(T)$ satisfy $d(x_k, y_k) < \epsilon_1$ for every $k \in (n+1]$, then $d(x_k, y_k) < \epsilon_2$ for every $k \in (n+1]$. Thus by (7.1), we have $K_{\phi,T}(\epsilon_1, n) \leqslant K_{\phi,T}(\epsilon_2, n)$.

Now we focus on the second inequality in (7.2).

Applying Lemma 6.5, we can choose $L \in \mathbb{N}$ with the following property:

For each $n \in \mathbb{N}$ greater than L, if two orbits $(x_1, \dots, x_{n+1}), (y_1, \dots, y_{n+1}) \in \mathcal{O}_{n+1}(T)$ satisfy $d(x_k, y_k) < \epsilon_2$ for every $k \in (n+1]$, then $d(x_k, y_k) < \epsilon_1$ holds for every $k \in (n+1-L]$. Since

$$\begin{aligned} &\Biggl| \sum_{k=1}^{n} (\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1})) \Biggr| \\ &\qquad \leqslant \Biggl| \sum_{k=1}^{n-L} (\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1})) \Biggr| + \sum_{k=n-L+1}^{n} (|\phi(x_k, x_{k+1})| + |\phi(y_k, y_{k+1})|) \\ &\qquad \leqslant \Biggl| \sum_{k=1}^{n-L} (\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1})) \Biggr| + 2L\|\phi\|_\infty, \end{aligned}$$

by (7.1) we get $K_{\phi,T}(\epsilon_2, n) \leqslant K_{\phi,T}(\epsilon_1, n - L) + 2L\|\phi\|_\infty$. □

Definition 7.5 (Distance-expanding)**.** Let T be a correspondence on a compact metric space (X, d). We say that T is *distance-expanding* if there exist $\lambda > 1$, $\eta > 0$, and $n \in \mathbb{N}$ with the property that for each $x, y \in X$, if $d(x, y) \leqslant \eta$, then

$$\inf\{d(x', y') : x' \in T^n(x),\ y' \in T^n(y)\} \geqslant \lambda d(x, y).$$

Remark 7.6. Note that if a correspondence T is distance-expanding, then it must be forward expansive, and thus we can say whether a bounded function $\phi \in B(\mathcal{O}_2(T), \mathbb{R})$ is Bowen summable.

Proposition 7.7. *Let T be a correspondence on a compact metric space X and $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$ be a bounded Borel measurable function. If T is distance-expanding and ϕ is Hölder continuous with respect to the metric d_2 on $\mathcal{O}_2(T)$, then ϕ is Bowen summable.*

Proof. If T is distance-expanding and ϕ is Hölder continuous, suppose that there are several constants $\lambda > 1$, $\eta > 0$, $n \in \mathbb{N}$, $\alpha \in (0,1)$, and $C > 0$ satisfying $|\phi(x_1, x_2) - \phi(y_1, y_2)| \leqslant C \cdot d_2((x_1, x_2), (y_1, y_2))^\alpha$ for all (x_1, x_2), $(y_1, y_2) \in \mathcal{O}_2(T)$ and $\inf\{d(x', y') : x' \in T^n(x),\, y' \in T^n(y)\} \geqslant \lambda d(x, y)$ for all x, $y \in X$ with $d(x, y) \leqslant \eta$.

Fix an arbitrary $q \in \mathbb{N}$. If $(x_1, \dots, x_{qn+1})$, $(y_1, \dots, y_{qn+1}) \in \mathcal{O}_{qn+1}(T)$ satisfy $d(x_k, y_k) < \eta$ for all $k \in (qn + 1]$, then for each $j \in ((q-1)n + 1]$, we have $d(x_{j+n}, y_{j+n}) \geqslant \inf\{d(x', y') : x' \in T^n(x_j),\, y' \in T^n(y_j)\} \geqslant \lambda d(x_j, y_j)$. Thereby, for every $p \in (q]$ and $r \in [n-1]$, we have $d(x_{pn-r}, y_{pn-r}) \leqslant \lambda^{p-q} \cdot d(x_{qn-r}, y_{qn-r}) < \lambda^{p-q}\eta$ and similarly $d(x_{pn-r+1}, y_{pn-r+1}) < \lambda^{p-q}\eta$, consequently,

$$\begin{aligned}
\left|\sum_{k=1}^{qn} (\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1}))\right| &\leqslant \sum_{k=1}^{qn} |\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1})| \\
&\leqslant \sum_{k=1}^{qn} C \cdot \max\{d(x_k, y_k), d(x_{k+1}, y_{k+1})\}^\alpha \\
&< \sum_{p=1}^{q} \sum_{r=0}^{n-1} C \cdot \lambda^{p-q}\eta \\
&< Cn\eta \frac{\lambda}{\lambda - 1}.
\end{aligned}$$

Fix an arbitrary $m \in \mathbb{N}, m \geqslant n$. Suppose $m = qn + r$, where $q \in \mathbb{N}$ and $r \in [n-1]$. If $(x_1, \dots, x_{m+1})$, $(y_1, \dots, y_{m+1}) \in \mathcal{O}_{m+1}(T)$ satisfy $d(x_k, y_k) < \eta$ for all $k \in (m+1]$, then $(x_1, \dots, x_{qn+1}), (y_1, \dots, y_{qn+1}) \in \mathcal{O}_{qn+1g}(T)$ satisfy $d(x_k, y_k) < \eta$ for all $k \in (qn + 1]$. By the estimates above, we have

$$\left|\sum_{k=1}^{qn} (\phi(x_k, x_{k+1}) - \phi(y_k, y_{k+1}))\right| < Cn\eta \frac{\lambda}{\lambda - 1}.$$

As a result, we have

$$\begin{aligned}
&\left|\sum_{k=1}^{m}(\phi(x_k,x_{k+1})-\phi(y_k,y_{k+1}))\right| \\
&\qquad \leqslant \left|\sum_{k=1}^{qn}(\phi(x_k,x_{k+1})-\phi(y_k,y_{k+1}))\right| + \left|\sum_{k=qn+1}^{qn+r}(\phi(x_k,x_{k+1})-\phi(y_k,y_{k+1}))\right| \\
&\qquad < Cn\eta\frac{\lambda}{\lambda-1}+r\|\phi\|_\infty \\
&\qquad < Cn\eta\frac{\lambda}{\lambda-1}+n\|\phi\|_\infty .
\end{aligned}$$

By (7.1), we get $K_{\phi,T}(\eta,m)\leqslant Cn\eta\frac{\lambda}{\lambda-1}+n\|\phi\|_\infty$. Hence

$$K_{\phi,T}(\eta)=\sup\{K_{\phi,T}(\eta,m) : m\in\mathbb{N}\}\leqslant Cn\eta\frac{\lambda}{\lambda-1}+n\|\phi\|_\infty<+\infty.$$

Therefore, by Definition 7.3, we conclude that ϕ is Bowen summable. □

Recall the notion of Bowen summability of $\varphi\colon X\to\mathbb{R}$ with a continuous map $f\colon X\to X$ from [Ru92, Section 1]:

Definition 7.8 (Bowen summability)**.** Let (X,d) be a compact metric space and $f\colon X\to X$ be a forward expansive continuous map. For a bounded Borel measurable function $\varphi\colon X\to\mathbb{R}$, denote

$$K_{\varphi,f}(\delta,n):=\sup\left\{\left|\sum_{k=0}^{n-1}\bigl(\varphi\bigl(f^k(x)\bigr)-\varphi\bigl(f^k(y)\bigr)\bigr)\right| : x,\,y\in X \text{ with } y\in B_x(\epsilon,n)\right\}, \tag{7.3}$$

for each $n\in\mathbb{N}$ and each $\delta>0$, where $B_x(\epsilon,n)$ is the *Bowen ball* given by

$$B_x(\epsilon,n):=\bigl\{y\in X : d\bigl(f^k(x),f^k(y)\bigr)<\delta \text{ for every } k\in[n-1]\bigr\}. \tag{7.4}$$

Choose an expansive constant ϵ of f, we write $K_{\varphi,f}(\epsilon):=\sup\{K_{\varphi,f}(\epsilon,n) : n\in\mathbb{N}\}$, and define $\mathcal{V}_f:=\{\varphi : K_{\varphi,f}(\epsilon)<+\infty\}$. Functions in $\mathcal{V}_f$ are called *Bowen summable* with respect to f.

The notation $\mathcal{V}_f$ does not contain ϵ because it does not depend on ϵ (see [Ru92, Section 1]).

Proposition 7.9. *Let T be a correspondence on a compact metric space (X,d). If a function $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ is Bowen summable with respect to T, then the $\widetilde{\phi}\colon\mathcal{O}_\omega(T)\to\mathbb{R}$ is Bowen summable with respect to the shift map $\sigma\colon\mathcal{O}_\omega(T)\to\mathcal{O}_\omega(T)$.*

Note that we do not require ϕ to be continuous.

Proof. Choose a number $\epsilon>0$ small enough such that ϵ is an expansive constant for T and that $\tilde{\epsilon}:=\frac{\epsilon}{2(1+\epsilon)}$ is an expansive constant for the shift map σ.

Suppose that $\underline{x} = (x_1, x_2, \dots)$, $\underline{y} = (y_1, y_2, \dots) \in \mathcal{O}_\omega(T)$ satisfy $\underline{y} \in B_{\underline{x}}(\tilde{\epsilon}, n+1)$, then for every $k \in [n]$,

$$\tilde{\epsilon} > d_\omega(\sigma^k(\underline{x}), \sigma^k(\underline{y})) = d((x_{k+1}, x_{k+2}, \dots), (y_{k+1}, y_{k+2}, \dots)) \geqslant \frac{d(x_{k+1}, y_{k+1})}{2(1 + d(x_{k+1}, y_{k+1}))}.$$

This implies that $d(x_{k+1},\, y_{k+1}) < \epsilon$ for every $k \in [n]$. By (7.3) and (7.1), we get $K_{\widetilde{\phi},\sigma}(\tilde{\epsilon}, n+1) \leqslant K_{\phi,T}(\epsilon, n)$. Since ϕ is Bowen summable with respect to T, we have

$$\begin{aligned} K_{\widetilde{\phi},\sigma}(\tilde{\epsilon}) &= \sup\bigl\{K_{\widetilde{\phi},\sigma}(\tilde{\epsilon}, n+1) : n \in \mathbb{N}_0\bigr\} \\ &\leqslant \max\bigl\{K_{\widetilde{\phi},\sigma}(\tilde{\epsilon}, 1),\, \sup\{K_{\phi,T}(\epsilon, n) : n \in \mathbb{N}\}\bigr\} \\ &\leqslant \max\{2\|\phi\|_\infty,\, K_{\phi,T}(\epsilon)\} \\ &< +\infty. \end{aligned}$$

Therefore $\widetilde{\phi}$ is Bowen summable with respect to the shift map σ. □

7.2. **Forward expansive correspondences with the specification property.** Our target in this subsection is to prove Theorem B. The key tool for our proof is the Ruelle–Perron–Frobenius Theorem for the shift map. We first recall some definitions and propositions from [RT18] without explanations.

Let Y be a compact metric space, $f\colon Y \to Y$ be a forward expansive continuous map with specification property, and $\psi\colon Y \to \mathbb{R}$ be a Bowen summable continuous function.

We recall the definition of the Ruelle operator $\mathcal{L}_\psi$ on Borel measurable functions on Y given by

$$\mathcal{L}_\psi(\Phi)(x) \coloneqq \sum_{y \in f^{-1}(x)} \Phi(y) \exp(\psi(y)). \tag{7.5}$$

The operator $\mathcal{L}_\psi$ is linear and maps the spaces of bounded Borel functions onto itself. The act of $\mathcal{L}_\psi$ on continuous functions determines completely the deal operator $\mathcal{L}_\psi^*$, a bounded linear map on finite Borel measures on Y. In other words, for a finite Borel measure ν on Y, the equality

$$\int_Y \Phi \,\mathrm{d}\mathcal{L}_\psi^*(\nu) = \int_Y \mathcal{L}_\psi(\Phi)\,\mathrm{d}\nu \tag{7.6}$$

holds for all continuous function $\Phi\colon Y \to \mathbb{R}$. This implies that (7.6) holds for all bounded Borel measurable functions $\Phi\colon Y \to \mathbb{R}$.

If $A \subseteq Y$ is a Borel set satisfying that $f|_A$ is injective, then we have

$$\mathcal{L}_\psi^*(\nu)(A) = \int_Y \mathbb{1}_A \,\mathrm{d}\mathcal{L}_\psi^*(\nu) = \int_Y \mathcal{L}_\psi(\mathbb{1}_A)\,\mathrm{d}\nu = \int_{f(A)} \exp\psi \circ (f|_A)^{-1}\,\mathrm{d}\nu. \tag{7.7}$$

If there is a non-zero Borel measure ν on Y such that $\mathcal{L}_\psi^*(\nu) = \lambda\nu$, then $\mathcal{L}_\psi$ defines an operator on $L^1(\nu)$.

We recall [RT18, Theorem 2.1], i.e., the Ruelle–Perron–Frobenius Theorem, as follows.

Proposition 7.10. *Let Y be a compact metric space, $f\colon Y\to Y$ be a forward expansive continuous map with specification property, and $\psi\colon Y\to\mathbb{R}$ be a Bowen summable continuous function. The following statements hold:*

(i) *There is a unique eigenvector ν (up to a multiplicative constant) of $\mathcal{L}_\psi^*$ acting on finite Borel measures on Y:*
$$\mathcal{L}_\psi^*(\nu)=\lambda\nu.$$
Furthermore, the eigenvalue $\lambda=\exp(P(f,\psi))$ and ν is a Gibbs state[5] *for ψ.*

(ii) *There is a unique non-negative eigenfunction $\Phi\in L^1(\nu)$ (up to a multiplicative constant) of $\mathcal{L}_\psi$ acting on $L^1(\nu)$:*
$$\mathcal{L}_\psi(\Phi)=\lambda\Phi,\quad \Phi\geqslant 0.$$
Furthermore, $\lambda=\exp(P(f,\psi))$, $\log\Phi$ is essentially bounded, and $\Phi\nu$ is the only equilibrium state for ψ.

(iii) *Denote by $\mathbf{1}_Y\colon Y\to\mathbb{R}$ the function that assigns each point $y\in Y$ the constant value 1, we have*
$$\lim_{n\to+\infty}\exp(-nP(f,\psi))\cdot\mathcal{L}_\psi^n(\mathbf{1}_Y)=\Phi\ \text{ in } L^1(\nu).$$

Now suppose that T is a forward expansive correspondence with the specification property on a compact metric space X and that $\phi\colon\mathcal{O}_2(T)\to\mathbb{R}$ is a Bowen summable continuous potential function.

By Theorem 4.9, (6.11), and (A.11), the following equality (7.8) concerning the dynamical system $(\mathcal{O}_\omega(T),\sigma)$ is equivalent to (1.2).

$$P\bigl(\sigma,\widetilde{\phi}\bigr)=h_{\mu_\phi\mathcal{Q}^\omega|_T}(\sigma)+\int_{\mathcal{O}_\omega(T)}\widetilde{\phi}\,\mathrm{d}(\mu_\phi\mathcal{Q}^\omega|_T).\tag{7.8}$$

By Proposition D.1, the forward expansiveness of T implies the forward expansiveness of σ. By Proposition D.3, the specification property of T implies the specification property of σ. By Proposition 7.9, the Bowen summability of ϕ with respect to T implies the Bowen summability of $\widetilde{\phi}$ with respect to σ.

Thereby, the dynamical system $(\mathcal{O}_\omega(T),\sigma)$ is forward expansive and has the specification property, and the continuous function $\widetilde{\phi}\colon\mathcal{O}_\omega(T)\to\mathbb{R}$ is Bowen summable with respect to σ. As a result, we can apply Proposition 7.10 for σ and $\widetilde{\phi}$. With this approach, we are now ready to prove assertions (i), (ii), (iii) and the uniqueness of $(\mu_\phi,\mathcal{Q})$ in Theorem B in turn.

[5]That ν is a Gibbs state for ψ means that there is a number $c>0$ such that for all $x\in Y$ and $n\in\mathbb{N}$,
$$\exp\biggl(\sum_{k=0}^{n-1}\bigl(\psi\bigl(f^k(x)\bigr)-nP(f,\psi)-c\bigr)\biggr)\leqslant\nu(B_x(\epsilon,n))\leqslant\exp\biggl(\sum_{k=0}^{n-1}\bigl(\psi\bigl(f^k(x)\bigr)-nP(f,\psi)+c\bigr)\biggr),$$
where $B_x(\epsilon,n)$ is the Bowen ball given in (7.4).

Proof of Theorem B, (i). By Proposition 7.10 (i), we can choose a Borel probability measure ν on $\mathcal{O}_\omega(T)$ with

$$\mathcal{L}^*_{\widetilde{\phi}}(\nu) = \lambda \cdot \nu, \tag{7.9}$$

where $\lambda = \exp\bigl(P\bigl(\sigma, \widetilde{\phi}\bigr)\bigr)$.

Set $\nu_{12} = \nu \circ \widetilde{\pi}_{12}^{-1}$, a Borel probability measure on X^2 supported on $\mathcal{O}_2(T)$ and $m_\phi \coloneqq \nu_{12} \circ \widetilde{\pi}_1^{-1} = (\nu \circ \widetilde{\pi}_{12}^{-1}) \circ \widetilde{\pi}_1^{-1} = \nu \circ \widetilde{\pi}_1^{-1}$, where $\widetilde{\pi}_{12}, \widetilde{\pi}_1$, and $\widetilde{\pi}_1$ are the projection maps given in (2.5). By Proposition 6.15, we can choose a transition probability kernel $\mathcal{Q}$ on X supported by T such that $m_\phi \mathcal{Q}^{[1]} = \nu_{12}$.

We will prove $\mathcal{L}^*_{\widetilde{\phi}}(m_\phi \mathcal{Q}^\omega|_T) = \lambda \cdot m_\phi \mathcal{Q}^\omega|_T$, or equivalently, for each $n \in \mathbb{N} \setminus \{1\}$ and arbitrary Borel sets $A_1, \dots, A_n \in \mathscr{B}(X)$ with $\operatorname{diam} A_1$ less than ϵ, an expansive constant for T, we have

$$\begin{aligned} &\mathcal{L}^*_{\widetilde{\phi}}\bigl(m_\phi \mathcal{Q}^\omega|_T\bigr)(A_1 \times \cdots \times A_n \times X^\omega \cap \mathcal{O}_\omega(T)) \\ &\qquad = \lambda \cdot (m_\phi \mathcal{Q}^\omega|_T)(A_1 \times \cdots \times A_n \times X^\omega \cap \mathcal{O}_\omega(T)). \end{aligned} \tag{7.10}$$

For each $x_2 \in T(A_1)$, there exists $x_1 \in A_1$ such that $x_2 \in T(x_1)$. If there is another $x_1' \in A_1$ such that $x_2 \in T(x_1')$ and $x_1 \neq x_1'$, we can choose $(x_2, x_3, \dots) \in \mathcal{O}_\omega(T)$, and then we have $(x_1, x_2, x_3, \dots) \in \mathcal{O}_\omega(T)$ and $(x_1', x_2, x_3, \dots) \in \mathcal{O}_\omega(T)$. As the forward expansiveness of T with ϵ as an expansive constant, we get either $x_1 = x_1'$, or $d(x_1, x_1') \geqslant \epsilon$. We have assumed $x_1 \neq x_1'$, so we have $d(x_1, x_1') \geqslant \epsilon$, which contradicts the assumptions $x_1, x_1' \in A_1$ and $\operatorname{diam} A_1 < \epsilon$. Thus $A_1 \cap T^{-1}(x_2)$ is a singleton. This allows us to define a map $J \colon T(A_1) \to A_1$ satisfying $A_1 \cap T^{-1}(x_2) = \{J(x_2)\}$ for all $x_2 \in T(A_1)$. The map J is a Borel map because its graph is closed in $T(A_1) \times A_1$.

If two orbits $\underline{x}^{(1)}, \underline{x}^{(2)} \in A_1 \times X^\omega \cap \mathcal{O}_\omega(T)$ satisfy $\sigma\bigl(\underline{x}^{(1)}\bigr) = \sigma\bigl(\underline{x}^{(2)}\bigr)$, then $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ are of the form $\bigl(x_1^{(1)}, x_2, x_3, \dots\bigr) \in \mathcal{O}_\omega(T)$ and $\bigl(x_1^{(2)}, x_2, x_3, \dots\bigr) \in \mathcal{O}_\omega(T)$, respectively, where $x_1^{(1)}, x_1^{(2)} \in A_1$, $x_2 \dots, x_n \in X$. Since $x_2 \in T\bigl(x_1^{(1)}\bigr) \subseteq T(A_1)$ and $x_1^{(1)}, x_1^{(2)} \in A_1 \cap T^{-1}(x_2)$, we have $x_1^{(1)} = J(x_2) = x_1^{(2)}$. Thus σ is injective on $A_1 \times X^\omega \cap \mathcal{O}_\omega(T)$ and we have $(\sigma|_{A_1 \times X^\omega \cap \mathcal{O}_\omega(T)})^{-1}(x_2, x_3, \dots) = (J(x_2), x_2, x_3, \dots)$ for all $(x_2, x_3, \dots) \in \sigma(A_1 \times X^\omega \cap \mathcal{O}_\omega(T))$. By (7.7), we have

$$\begin{aligned} &\mathcal{L}^*_{\widetilde{\phi}}(m_\phi \mathcal{Q}^\omega|_T)(A_1 \times \cdots \times A_n \times X^\omega \cap \mathcal{O}_\omega(T)) \\ &\qquad = \int_{\sigma(A_1 \times \cdots \times A_n \times X^\omega \cap \mathcal{O}_\omega(T))} \exp\bigl(\widetilde{\phi} \circ \bigl(\sigma|_{A_1 \times X^\omega \cap \mathcal{O}_\omega(T)}\bigr)^{-1}\bigr) \,\mathrm{d}(m_\phi \mathcal{Q}^\omega|_T) \\ &\qquad = \int_{(T(A_1) \cap A_2) \times A_3 \times \cdots \times A_n \times X^\omega} \exp(\phi(J(x_2), x_2)) \,\mathrm{d}(m_\phi \mathcal{Q}^\omega)(x_2, x_3, \dots) \\ &\qquad = \int_{(T(A_1) \cap A_2) \times A_3 \times \cdots \times A_n} \exp(\phi(J(x_2), x_2)) \,\mathrm{d}\bigl(m_\phi \mathcal{Q}^{[n-2]}\bigr)(x_2, \dots, x_n). \end{aligned} \tag{7.11}$$

In addition, we have

$$\begin{aligned} &(m_\phi \mathcal{Q}^\omega|_T)(A_1 \times \cdots \times A_n \times X^\omega \cap \mathcal{O}_\omega(T)) \\ &\qquad = (m_\phi \mathcal{Q}^\omega)(A_1 \times \cdots \times A_n \times X^\omega) = \bigl(m_\phi \mathcal{Q}^{[n-1]}\bigr)(A_1 \times \cdots \times A_n). \end{aligned} \tag{7.12}$$

By (7.11) and (7.12), the equality (7.10) is equivalent to

$$\int_{(T(A_1)\cap A_2)\times A_3\times\cdots\times A_n} \exp(\phi(J(x_2),x_2))\,\mathrm{d}\big(m_\phi\mathcal{Q}^{[n-2]}\big)(x_2,\dots,x_n) \\ = \lambda\cdot\big(m_\phi\mathcal{Q}^{[n-1]}\big)(A_1\times\cdots\times A_n). \tag{7.13}$$

We prove (7.13) by induction on n.

If $n=2$, we rewrite (7.13) as

$$\int_{T(A_1)\cap A_2} \exp(\phi(J(x_2),x_2))\,\mathrm{d}m_\phi(x_2) = \lambda\cdot\big(m_\phi\mathcal{Q}^{[1]}\big)(A_1\times A_2). \tag{7.14}$$

To prove (7.14), we come back to the property of ν. The equality (7.9) implies

$$\begin{aligned}\int_{\sigma(A_1\times A_2\times X^\omega\cap\mathcal{O}_\omega(T))} & \exp\Big(\widetilde{\phi}\circ\big(\sigma|_{A_1\times X^\omega\cap\mathcal{O}_\omega(T)}\big)^{-1}\Big)\,\mathrm{d}\nu \\ &= \lambda\cdot\nu(A_1\times A_2\times X^\omega\cap\mathcal{O}_\omega(T)) \\ &= \lambda\cdot\nu_{12}(A_1\times A_2\cap\mathcal{O}_2(T)) \\ &= \lambda\cdot\big(m_\phi\mathcal{Q}^{[1]}\big)(A_1\times A_2).\end{aligned}$$

Moreover, recall $m_\phi=\nu\circ\widetilde{\pi}_1^{-1}$. We have

$$\begin{aligned}\int_{\sigma(A_1\times A_2\times X^\omega\cap\mathcal{O}_\omega(T))} & \exp\Big(\widetilde{\phi}\circ\big(\sigma|_{A_1\times X^\omega\cap\mathcal{O}_\omega(T)}\big)^{-1}\Big)\,\mathrm{d}\nu \\ &= \int_{(T(A_1)\cap A_2)\times X^\omega\cap\mathcal{O}_\omega(T)} \exp(\phi(J(x_2),x_2))\,\mathrm{d}\nu(x_2,x_3,\dots) \\ &= \int_{T(A_1)\cap A_2} \exp(\phi(J(x_2),x_2))\,\mathrm{d}m_\phi(x_2).\end{aligned}$$

Hence the equality (7.14) holds, i.e., the equality (7.13) holds for $n=2$.

Suppose that (7.13) holds for $n-1$, $n\geqslant 3$, which means

$$\int_{(T(A_1)\cap A_2)\times A_3\times\cdots\times A_{n-1}} \exp(\phi(J(x_2),x_2))\,\mathrm{d}\big(m_\phi\mathcal{Q}^{[n-3]}\big)(x_2,\dots,x_{n-1}) \\ = \lambda\cdot\big(m_\phi\mathcal{Q}^{[n-2]}\big)(A_1\times\cdots\times A_{n-1}).$$

This and (A.4) in Lemma A.3 imply that

$$\begin{aligned}\lambda\cdot & \big(m_\phi\mathcal{Q}^{[n-1]}\big)(A_1\times\cdots\times A_n) \\ &= \lambda\int_{A_1\times\cdots\times A_{n-1}} \mathcal{Q}(x_{n-1},A_n)\,\mathrm{d}\big(m_\phi\mathcal{Q}^{[n-2]}\big)(x_1,\dots,x_{n-1}) \\ &= \int_{(T(A_1)\cap A_2)\times A_3\times\cdots\times A_{n-1}} \mathcal{Q}(x_{n-1},A_n)e^{\phi(J(x_2),x_2)}\,\mathrm{d}\big(m_\phi\mathcal{Q}^{[n-3]}\big)(x_2,\dots,x_{n-1}) \\ &= \int_{(T(A_1)\cap A_2)\times A_3\times\cdots\times A_n} e^{\phi(J(x_2),x_2)}\,\mathrm{d}\big(m_\phi\mathcal{Q}^{[n-2]}\big)(x_2,\dots,x_n).\end{aligned}$$

Hence (7.13) holds for n, and therefore $\mathcal{L}^*_{\widetilde{\phi}}\big(m_\phi\mathcal{Q}^\omega|_T\big)=\lambda\cdot m_\phi\mathcal{Q}^\omega|_T$. $\square$

Proof of Theorem B, (ii). Let $v\colon X \to \mathbb{R}$ be a non-negative bounded Borel measurable function. For each $(x_1, x_2, \dots) \in \mathcal{O}_\omega(T)$, we have

$$\begin{aligned}\mathcal{L}_{\widetilde{\phi}}\widetilde{v}(x_1, x_2, \dots) &= \sum_{(x_0, x_1, \dots) \in \sigma^{-1}(x_1, x_2, \dots)} \widetilde{v}(x_0, x_1, \dots) \exp(\widetilde{\phi}(x_0, x_1, \dots)) \\ &= \sum_{x_0 \in T^{-1}(x_1)} v(x_0)\phi(x_0, x_1),\end{aligned}$$

which indicates that $\mathcal{L}_{\widetilde{\phi}}\widetilde{v}(x_1, x_2, \dots)$ only depends on x_1. Consequently, there exists a function $w\colon X \to \mathbb{R}$ such that $\widetilde{w} = \mathcal{L}_{\widetilde{\phi}}\widetilde{v}$. Then one can check the non-negativeness, boundedness, and Borel measurability of w.

The statement above implies that

$$\exp\bigl(-nP\bigl(\sigma, \widetilde{\phi}\bigr)\bigr) \cdot \mathcal{L}^n_{\widetilde{\phi}}(\mathbf{1}_{\mathcal{O}_\omega(T)}) = \exp\bigl(-nP\bigl(\sigma, \widetilde{\phi}\bigr)\bigr) \cdot \mathcal{L}^n_{\widetilde{\phi}}\bigl(\widetilde{\mathbf{1}_X}\bigr)$$

is of the form $\widetilde{u}_n$ for some non-negative bounded Borel measurable function $u_n\colon X \to \mathbb{R}$ for each $n \in \mathbb{N}$. By Proposition 7.10 (ii) and (iii), $\{\widetilde{u}_n\}_{n\in\mathbb{N}}$ converges to Φ in $L^1(m_\phi \mathcal{Q}^\omega|_T)$, where Φ is the only non-negative eigenfunction of $\mathcal{L}_{\widetilde{\phi}}$ acting on $L^1(m_\phi \mathcal{Q}^\omega|_T)$.

By taking $n = 0$ in (5.7), we get that the sequence $\{u_n\}_{n\in\mathbb{N}}$ converges in $L^1(m_\phi)$. Suppose that u_n converges to $u_\phi \in L^1(m_\phi)$ as $n \to +\infty$ in $L^1(m_\phi)$. Then $\widetilde{u}_n$ converges to $\widetilde{u}_\phi$ as $n \to +\infty$ in $L^1(m_\phi \mathcal{Q}^\omega|_T)$. Thus $\Phi = \widetilde{u}_\phi$ in $L^1(m_\phi)$. By Proposition 7.10 (ii), we have $\mathcal{L}_{\widetilde{\phi}}(\widetilde{u}_\phi) = \lambda\widetilde{u}_\phi$. $\square$

Proof of Theorem B, (iii). By Proposition 7.10 (ii), $\widetilde{u}_\phi(m_\phi \mathcal{Q}^\omega|_T)$, a Borel probability measure on $\mathcal{O}_\omega(T)$, is the only equilibrium state for σ.

Set $\mu_\phi := u_\phi m_\phi$. For an arbitrary Borel set $M \in \mathscr{B}(X^\omega)$, we have

$$\begin{aligned}&\widetilde{u}_\phi(m_\phi \mathcal{Q}^\omega|_T)(M \cap \mathcal{O}_\omega(T)) \\ &\quad= \int_{M\cap\mathcal{O}_\omega(T)} \widetilde{u}_\phi(x_1, x_2, \dots)\,\mathrm{d}(m_\phi \mathcal{Q}^\omega|_T)(x_1, x_2, \dots) \\ &\quad= \int_M u_\phi(x_1)\,\mathrm{d}(m_\phi \mathcal{Q}^\omega)(x_1, x_2, \dots) \\ &\quad= \int_X \biggl(\int_{X^\omega} \mathbb{1}_M(x_1, x_2, \dots) \cdot u_\phi(x_1)\,\mathrm{d}\mathcal{Q}^\omega_x(x_1, x_2, \dots)\biggr)\,\mathrm{d}m_\phi(x) \text{ (by Proposition 5.9)} \\ &\quad= \int_X \biggl(\int_{X^\omega} \mathbb{1}_M(x_1, x_2, \dots) \cdot u_\phi(x)\,\mathrm{d}\mathcal{Q}^\omega_x(x_1, x_2, \dots)\biggr)\,\mathrm{d}m_\phi(x) \qquad \text{(by Lemma A.8)} \\ &\quad= \int_X u_\phi(x) \cdot \mathcal{Q}^\omega(x, M)\,\mathrm{d}m_\phi(x) \\ &\quad= \int_X \mathcal{Q}^\omega(x, M)\,\mathrm{d}(u_\phi m_\phi)(x) \\ &\quad= ((u_\phi m_\phi)\mathcal{Q}^\omega)(M) \\ &\quad= ((u_\phi m_\phi)\mathcal{Q}^\omega|_T)(M \cap \mathcal{O}_\omega(T)).\end{aligned}$$

Hence $\mu_\phi \mathcal{Q}^\omega|_T = (u_\phi m_\phi)\mathcal{Q}^\omega|_T = \widetilde{u}_\phi(m_\phi \mathcal{Q}^\omega|_T)$ is the only equilibrium state for σ, and therefore (1.2) holds for $(\mu_\phi, \mathcal{Q})$. $\square$

We have finished constructing an equilibrium state $(\mu_\phi, \mathcal{Q})$ for the correspondence T and the potential function ϕ. Now we show that it is unique in the sense of Theorem B.

Proof of Theorem B, uniqueness of the equilibrium state. Recall that (1.2) is equivalent to that $\mu\mathcal{Q}^\omega|_T$ is an equilibrium state for $\widetilde{\phi}$ in the dynamical system $(\mathcal{O}_\omega(T), \sigma)$. Since $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ is a forward expansive continuous map with specification property and $\widetilde{\phi}\colon \mathcal{O}_\omega(T) \to \mathbb{R}$ is a Bowen summable continuous function, Proposition 7.10 (ii) says that the equilibrium state for $\widetilde{\phi}$ in the dynamical system $(\mathcal{O}_\omega(T), \sigma)$ is unique.

Suppose that both $(\mu, \mathcal{Q})$ and $(\mu', \mathcal{Q}')$ are equilibrium states for the correspondence T and the potential function ϕ, then both $\mu\mathcal{Q}^\omega|_T$ and $\mu'(\mathcal{Q}')^\omega|_T$ are equilibrium states for σ and $\widetilde{\phi}$, and thus $\mu\mathcal{Q}^\omega|_T = \mu'(\mathcal{Q}')^\omega|_T$. Thereby, we have $\mu\mathcal{Q}^\omega = \mu'(\mathcal{Q}')^\omega$ by Lemma 6.13. By (5.7), we have $\mu\mathcal{Q}^{[1]} = \mu'(\mathcal{Q}')^{[1]}$. By Proposition A.11, we conclude $\mu = \mu'$ and that for μ-almost $x \in X$ and all $A \in \mathscr{B}(X)$, the equality $\mathcal{Q}(x, A) = \mathcal{Q}'(x, A)$ holds. $\square$

7.3. **Open, distance-expanding, strongly transitive correspondences.** In this subsection, we will prove Theorem C, which provides another version of conditions that ensure the Variational Principle, the existence and uniqueness of the equilibrium state, and some equidistribution properties of the unique equilibrium state. We first introduce some notions.

Definition 7.11 (Openness)**.** Let T be a correspondence on a compact metric space $(X, \ , d)$. We say that T is *open*, if, for each open subset $U \subseteq X$, $T(U)$ is an open subset of X.

Propositions D.4 and D.5 indicate that the openness of the correspondence T and of the corresponding shift map $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ are equivalent.

Definition 7.12 (Strong transitivity)**.** We say that a correspondence T on a compact metric space X is *strongly transitive* if $\bigcup_{n=1}^{+\infty} T^{-n}(x)$ is dense in X for every $x \in X$.

Remark. We call this property to be *strongly transitive* because if $T = \mathcal{C}_f$ for some continuous map $f\colon X \to X$, then this property is slightly stronger than topological transitivity, see Appendix B.3 (v).

Definition 7.13 (Topological exactness)**.** Let T be a correspondence on a compact metric space X. We say that T is *topologically exact*[6] if for every non-empty open subset $U \subseteq X$, there exists $N \in \mathbb{N}$ such that $T^N(U) = X$.

---

[6]C. Siqueira and D Smania [SS17, Section 4.4] called this property locally eventually onto for $\mathbf{f}_c$ on what they called "hyperbolic repellers".

Proposition 7.14. *Let X be a compact metric space and T be a correspondence on (X, d). If a function $\phi\colon \mathcal{O}_2(T) \to \mathbb{R}$ is α-Hölder continuous with respect to the metric d_2 on $\mathcal{O}_2(T)$, then the function $\widetilde{\phi}\colon \mathcal{O}_\omega(T) \to \mathbb{R}$ is α-Hölder continuous with respect to the metric d_ω on $\mathcal{O}_\omega(T)$.*

Proof. Suppose that ϕ is α-Hölder continuous with respect to the metric d_2 on $\mathcal{O}_2(T)$ and that a constant $C > 0$ satisfy

$$|\phi(x_1, x_2) - \phi(y_1, y_2)| \leqslant C \cdot d_2((x_1, x_2), (y_1, y_2))^\alpha$$

for all (x_1, x_2), $(y_1, y_2) \in \mathcal{O}_2(T)$. Then for arbitrary $\underline{x} = (x_1, x_2, \dots)$ and $\underline{y} = (y_1, y_2, \dots)$ in $\mathcal{O}_\omega(T)$, we have

$$\bigl|\widetilde{\phi}(\underline{x}) - \widetilde{\phi}(\underline{y})\bigr| = |\phi(x_1, x_2) - \phi(y_1, y_2)| \leqslant C \cdot \max\{d(x_1, y_1), d(x_2, y_2)\}^\alpha.$$

Since

$$\begin{aligned} d_\omega(\underline{x}, \underline{y}) &\geqslant \frac{d(x_1, y_1)}{2(1 + d(x_1, y_1))} + \frac{d(x_2, y_2)}{4(1 + d(x_2, y_2))} \\ &\geqslant \frac{d(x_1, y_1)}{4(1 + \operatorname{diam} X)} + \frac{d(x_2, y_2)}{4(1 + \operatorname{diam} X)} \\ &\geqslant \frac{\max\{d(x_1, y_1), d(x_2, y_2)\}}{4(1 + \operatorname{diam} X)}, \end{aligned}$$

where $\operatorname{diam} X = \sup\{d(z, w) : z,\, w \in X\} < +\infty$, we have

$$\bigl|\widetilde{\phi}(\underline{x}) - \widetilde{\phi}(\underline{y})\bigr| \leqslant C \cdot (4(1 + \operatorname{diam} X))^\alpha \cdot d_\omega(\underline{x}, \underline{y})^\alpha.$$

Therefore we conclude that $\widetilde{\phi}$ is α-Hölder continuous with respect to the metric d_ω on $\mathcal{O}_\omega(T)$. $\square$

Let T be a correspondence on a compact metric space X, recall $\mathcal{O}_{-n}(x) = \mathcal{O}_{n+1}(T) \cap X^n \times \{x\} = \{(y_0, y_1, \dots, y_n) \in \mathcal{O}_{n+1}(T) : y_n = x\}$ for all $n \in \mathbb{N}$ and $x \in X$ from Theorem C. If T is forward expansive, then the set $\mathcal{O}_{-n}(x)$ is finite for all $n \in \mathbb{N}$ and $x \in X$, ensured by the fact shown in Remark 6.2 that $T^{-1}(y) = \{z \in X : y \in T(z)\}$ is a finite set for all $y \in X$.

The proofs of Theorem C (i), (ii), and (iii) are similar to the proofs of Theorem B (i), (ii), and (iii), and the proofs of Theorem C (a) and (b) are similar, so now we sketch the proofs of Theorem C (i), (ii), (iii), and (b) and give a detailed proof of Theorem C (a).

First, to prove Theorem C (i), (ii), and (iii), we should notice the property of the corresponding shift map $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ if an open, strongly transitive, distance-expanding correspondence T on X is given: that σ is open (by Proposition D.4), topologically transitive (by Proposition D.6), and distance-expanding (by Proposition D.8). Also, by Proposition 7.14, the lifted potential function $\widetilde{\phi}\colon \mathcal{O}_\omega(T) \to \mathbb{R}$ is α-Hölder continuous if $\phi\colon X \to \mathbb{R}$ is α-Hölder continuous. Thereby, under the setting of Theorem C, we can apply the following version of the Ruelle–Perron–Frobenius Theorem for the continuous map σ and the potential function $\widetilde{\phi}$.

Proposition 7.15. *Let Y be a compact metric space, $f\colon Y\to Y$ be an open, topologically transitive, distance-expanding continuous map, and $\psi\colon Y\to\mathbb{R}$ be an α-Hölder continuous function with respect to the metric on Y, where $\alpha\in(0,1)$. Then the following statements hold:*

(i) *There is a unique eigenvector ν (up to a multiplicative constant) of $\mathcal{L}_\psi^*$ acting on finite Borel measures on Y:*

$$\mathcal{L}_\psi^*(\nu)=\lambda\nu.$$

Furthermore, the eigenvalue $\lambda=\exp(P(f,\psi))$ and ν is a Gibbs state for ψ.

(ii) *There is a unique positive α-Hölder eigenfunction Φ (up to a multiplicative constant) of $\mathcal{L}_\psi$:*

$$\mathcal{L}_\psi(\Phi)=\lambda\Phi,\quad \Phi>0.$$

Furthermore, $\lambda=\exp(P(f,\psi))$ and $\Phi\nu$ is the only equilibrium state for ψ.

(iii) *If we denote by $\mathbf{1}_Y\colon Y\to\mathbb{R}$ the function that assigns each point $y\in Y$ the constant value 1, then the sequence $\exp(-nP(f,\psi))\cdot\mathcal{L}_\psi^n$ converges uniformly to Φ as $n\to+\infty$.*

In addition, the backward orbits under f are equidistributed with respect to the measure $\Phi\nu$. More precisely, the following statements hold:

(a) *For each $y\in Y$, the following sequence of Borel probability measures on Y*

$$\frac{1}{\sum\limits_{z\in f^{-n}(y)}\exp\Bigl(\sum\limits_{i=0}^{n-1}\psi(f^i(z))\Bigr)}\sum_{z\in f^{-n}(y)}\frac{\sum_{j=0}^{n}\delta_{f^j(z)}}{n+1}\exp\Biggl(\sum_{i=0}^{n-1}\psi(f^i(z))\Biggr),\ n\in\mathbb{N},$$

converges to $\Phi\nu$ in the weak topology as n tends to $+\infty$.*

(b) *If, moreover, f is topologically exact, then for each $y\in Y$, the following sequence of Borel probability measures on Y*

$$\frac{1}{\sum\limits_{z\in f^{-n}(y)}\exp\Bigl(\sum\limits_{i=0}^{n-1}\psi(f^i(z))\Bigr)}\sum_{z\in f^{-n}(y)}\delta_z\exp\Biggl(\sum_{i=0}^{n-1}\psi(f^i(z))\Biggr),\ n\in\mathbb{N},$$

converges to ν in the weak topology as n tends to $+\infty$.*

This proposition is summarized from [PU10, Chapter 5]. In detail, statement (i) comes from [PU10, Theorem 5.2.8, Propositions 5.2.11, and 5.1.1], statement (ii) comes from [PU10, Proposition 5.1.5, 5.3.1, 5.2.10, Theorems 5.3.2, and 5.6.2], statement (iii) comes from [PU10, Section 5.4, (5.4.2)], statement (a) comes from [PU10, Remark 4.4.4], and statement (b) comes from [PU10, Section 5.4, (5.4.4)].

By applying Proposition 7.15 (i) for σ and $\widetilde{\phi}$, we can get the unique Borel probability measure ν on $\mathcal{O}_\omega(T)$ with $\mathcal{L}_{\widetilde{\phi}}^*(\nu)=\exp\bigl(P\bigl(\sigma,\widetilde{\phi}\bigr)\bigr)\cdot\nu$. The proof of Theorem B (i) indicates that ν is of the form $m_\phi\mathcal{Q}^\omega|_T$, where m_ϕ is a Borel probability measure on X and $\mathcal{Q}$ is a transition probability kernel on X. Consequently, Theorem C (i) follows.

By applying Proposition 7.15 (ii) for σ and $\widetilde{\phi}$, we can get the unique α-Hölder function $\Phi\colon \mathcal{O}_\omega(T) \to \mathbb{R}$ with $\mathcal{L}_{\widetilde{\phi}}(\Phi) = \exp\bigl(P\bigl(\sigma, \widetilde{\phi}\bigr)\bigr) \cdot \Phi$. Applying Proposition 7.15 (iii), we can see $\Phi = \widetilde{u}_\phi$ for some function $u_\phi \in L^1(m_\phi)$ following the proof of Theorem B (ii). In addition, suppose that T is continuous in the sense of Definition 4.3, we aim to prove that u_ϕ is continuous. Fix an arbitrary $\epsilon > 0$. By Proposition 7.15 (ii), $\widetilde{u}_\phi$ is Hölder continuous, so we can choose $\delta > 0$ such that $|\widetilde{u}_\phi(\underline{x}) - \widetilde{u}_\phi(\underline{y})| < \epsilon$ holds for all $\underline{x}$, $\underline{y} \in \mathcal{O}_\omega(T)$ with $d_\omega(\underline{x}, \underline{y}) < 2\delta$. Choose $n \in \mathbb{N}$ satisfying $2^{-n} < \delta$. Set $\delta_n \coloneqq \delta$. According to Definition 4.3, the compactness of X implies that $T\colon X \to \mathcal{F}(X)$ is uniformly continuous, i.e., for each $\eta > 0$, there exists $\eta' > 0$ satisfying $d_H(T(x), T(y)) < \eta$ for all x, $y \in X$ with $d(x, y) < \eta'$, where d_H is the Hausdorff distance on $\mathcal{F}(X)$. This allows us to choose $\delta_{n-1}, \delta_{n-2}, \dots, \delta_1$ sequentially with the property that for each $k \in (n-1]$,

(1) $d_H(T(x), T(y)) < \delta_{k+1}$ for all x, $y \in X$ with $d(x, y) < \delta_k$,

(2) $0 < \delta_k < \delta_{k+1}$.

Fix arbitrary x_1, $y_1 \in X$ with $d(x_1, y_1) < \delta_1$. By induction on k, we can see that we can choose $x_2 \in T(x_1)$, $y_2 \in T(y_1)$, $\dots$, $x_n \in T(x_{n-1})$, $y_n \in T(y_{n-1}) \in X$ such that $d(x_k, y_k) < \delta_k < \delta_n = \delta$ for all $k \in (n]$. Furthermore, we choose x_{n+1}, y_{n+1}, x_{n+2}, y_{n+2}, $\dots \in X$ such that $\underline{x} \coloneqq (x_1, \dots, x_n, x_{n+1}, \dots) \in \mathcal{O}_\omega(T)$ and $\underline{y} \coloneqq (y_1, \dots, y_n, y_{n+1}, \dots) \in \mathcal{O}_\omega(T)$. We have

$$d_\omega(\underline{x}, \underline{y}) = \sum_{k=1}^{+\infty} \frac{1}{2^k} \frac{d(x_k, y_k)}{1 + d(x_k, y_k)} \leqslant \sum_{k=1}^{n} \frac{1}{2^k}\delta + \sum_{k=n+1}^{+\infty} \frac{1}{2^k} < \delta + 2^{-n} < 2\delta.$$

This implies $|u_\phi(x_1) - u_\phi(y_1)| = |\widetilde{u}_\phi(\underline{x}) - \widetilde{u}_\phi(\underline{y})| < \epsilon$. Since ϵ is chosen arbitrarily, we conclude that u_ϕ is continuous. Theorem C (ii) follows.

We have proved $(u_\phi m_\phi)\mathcal{Q}^\omega|_T = \widetilde{u}_\phi(m_\phi \mathcal{Q}^\omega|_T)$ in the proof of Theorem B (iii). This equality and Proposition 7.15 (ii) imply Theorem C (iii).

In the proof of the uniqueness of the equilibrium state in Theorem B, we have shown that if the equilibrium state for the shift map σ and the potential $\widetilde{\phi}$ is unique, then the equilibrium state for the correspondence T and the potential ϕ is unique in the sense of Theorem B. The uniqueness of the equilibrium state in the setting of Theorem C also follows by the uniqueness of the equilibrium state for the shift map σ and the potential $\widetilde{\phi}$ (see Proposition 7.15 (ii)) in the same way as the proof of the uniqueness of the equilibrium state in Theorem B.

Now we give a detailed proof of Theorem C (a).

Proof of Theorem C (a). We have pointed out that $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ is an open, topologically transitive, distance-expanding continuous map, and that $\widetilde{\phi}\colon \mathcal{O}_\omega(T) \to \mathbb{R}$ is a Hölder continuous function. This allows us to apply Proposition 7.15 (a) to the shift map σ and the potential function $\widetilde{\phi}$:

For each $\underline{x} \in \mathcal{O}_\omega(T)$, the following sequence of Borel probability measures on $\mathcal{O}_\omega(T)$

$$\frac{1}{\sum\limits_{\underline{z}\in\sigma^{-n}(\underline{x})} \exp\Bigl(\sum\limits_{i=0}^{n-1} \widetilde{\phi}(\sigma^i(\underline{z}))\Bigr)} \sum_{\underline{z}\in\sigma^{-n}(\underline{x})} \frac{\sum_{j=0}^{n} \delta_{\sigma^j(\underline{z})}}{n+1} \exp\biggl(\sum_{i=0}^{n-1} \widetilde{\phi}(\sigma^i(\underline{z}))\biggr), \ n \in \mathbb{N},$$

converges to $\mu_\phi \mathcal{Q}^\omega|_T$ in the weak* topology as n tends to $+\infty$.

If we consider the projection of the sequence onto the first coordinate, i.e., we consider each item composing $\widetilde{\pi}_1^{-1}$, then we can get Theorem C (a). $\square$

Notice that by Proposition D.7, we can apply Proposition 7.15 (b) to the shift map σ and the potential function $\widetilde{\phi}$ under the assumption that T is topologically exact. Then we consider the projection of the measure sequence, and then we can get Theorem C (b) in the same way as the proof of Theorem C (a).

Appendix A. Transition probability kernels

We have recalled the definition of transition probability kernels and other related notions in Subsections 5.1 and 5.2. We discuss further about transition probability kernels in this appendix.

A.1. **Proofs of basic properties.** In this appendix, we prove the lemmas and propositions in Subsections 5.1 and 5.2 and check that Definitions 5.11, 5.14, and 5.18 are well-defined.

Proof of Lemma 5.5. For every $y \in Y$, by the bounded convergence theorem,

$$\lim_{n\to+\infty} \mathcal{Q}f_n(y) = \lim_{n\to+\infty} \int_X f_n(x) \,\mathrm{d}\mathcal{Q}_y(x) = \int_X f(x) \,\mathrm{d}\mathcal{Q}_y(x) = \mathcal{Q}f(y).$$

Since $|\mathcal{Q}f(y)| = \bigl|\int_X f(x) \,\mathrm{d}\mathcal{Q}_y(x)\bigr| \leqslant \int_X \|f\|_\infty \,\mathrm{d}\mathcal{Q}_y(x) \leqslant \|f\|_\infty$ holds for every $y \in Y$, we have $\|\mathcal{Q}f\|_\infty \leqslant \|f\|_\infty$.

For every $A \in \mathscr{M}(X)$, denote by $\mathbb{1}_A \colon X \to \mathbb{R}$ the characteristic function on A. For each $y \in Y$, we have $\mathcal{Q}\mathbb{1}_A(y) = \int_X \mathbb{1}_A(x) \,\mathrm{d}\mathcal{Q}_y(x) = \mathcal{Q}(y, A)$. By Definition 5.1, the map $\mathcal{Q}\mathbb{1}_A \colon y \mapsto \mathcal{Q}(y, A)$ is measurable. Consequently, by (5.1), for an arbitrary simple function $g = a_1\mathbb{1}_{A_1} + a_2\mathbb{1}_{A_2} + \cdots + a_n\mathbb{1}_{A_n}$, where $a_1, \dots, a_n \in \mathbb{R}$ and $A_1, \dots, A_n \in \mathscr{M}(X)$, the function $\mathcal{Q}g = a_1\mathcal{Q}\mathbb{1}_{A_1} + \cdots + a_n\mathcal{Q}\mathbb{1}_{A_n}$ is measurable.

Choose a sequence of uniformly bounded simple functions $f_n \colon X \to \mathbb{R}$ that converges pointwise to f as $n \to +\infty$. We have shown that $\mathcal{Q}f_n$ converges pointwise to $\mathcal{Q}f$ as $n \to \infty$. By the measurability of $\mathcal{Q}f_n$ for all $n \in \mathbb{N}$, we conclude that $\mathcal{Q}f$ is measurable. $\square$

Proof of Proposition 5.9. Choose a sequence of uniformly bounded simple functions $f_n \colon X \to \mathbb{R}$, $n \in \mathbb{N}$, convergent pointwise to f as $n \to +\infty$. By Lemma 5.5, $\mathcal{Q}f_n$ is uniformly bounded and convergent pointwise to $\mathcal{Q}f$ as $n \to +\infty$. Thus

$$\lim_{n\to+\infty} \int_X f_n \,\mathrm{d}(\mu\mathcal{Q}) = \int_X f \,\mathrm{d}(\mu\mathcal{Q}), \text{ and } \lim_{n\to+\infty} \int_Y \mathcal{Q}f_n \,\mathrm{d}\mu = \int_Y \mathcal{Q}f \,\mathrm{d}\mu.$$

Thereby, we can reduce (5.2) to the case where f is a simple function. Moreover, the linearity of $\mathcal{Q}$ as an operator on all bounded measurable functions on X, i.e., (5.1), allows us to further reduce (5.2) to the case where f is a characteristic function. Let us verify (5.2) when $f = \mathbb{1}_B$, the characteristic function of an arbitrary measurable set $B \in \mathscr{M}(X)$:

$$\int_X \mathbb{1}_B \,\mathrm{d}(\mu\mathcal{Q}) = (\mu\mathcal{Q})(B) = \int_Y \mathcal{Q}(y, B)\,\mathrm{d}\mu(y) = \int_Y \mathcal{Q}\mathbb{1}_B \,\mathrm{d}\mu.$$

Hence, (5.2) holds when f is an arbitrary characteristic function, and therefore it holds when f is a bounded measurable function. □

Lemma A.1. *The map $\mathcal{Q}'\mathcal{Q}$ given in Definition 5.11 is indeed a transition probability kernel from Z to X.*

Proof. First, fix an arbitrary $z \in Z$, the map that assigns each measurable set $A \in \mathscr{M}(X)$ the value $(\mathcal{Q}'\mathcal{Q})(z, A)$ is the probability measure $\mathcal{Q}'_z\mathcal{Q}$ on X. Then we shall check that for every $A \in \mathscr{M}(X)$, the function that assigns each $z \in Z$ the value $(\mathcal{Q}'\mathcal{Q})(z, A)$ is measurable. Fix an arbitrary measurable set $A \in \mathscr{M}(X)$. For each $z \in Z$ we have

$$(\mathcal{Q}'(\mathcal{Q}\mathbb{1}_A))(z) = \int_Y \mathcal{Q}\mathbb{1}_A(y)\,\mathrm{d}\mathcal{Q}'_z(y) = \int_Y \mathcal{Q}(y, A)\,\mathrm{d}\mathcal{Q}'_z(y) = (\mathcal{Q}'_z\mathcal{Q})(A) = (\mathcal{Q}'\mathcal{Q})(z, A).$$

Thus the function that assigns $z \in Z$ the value $(\mathcal{Q}'\mathcal{Q})(z, A)$ is $\mathcal{Q}'(\mathcal{Q}\mathbb{1}_A)$, which is measurable by Lemma 5.5. Therefore we conclude that $\mathcal{Q}'\mathcal{Q}$ is indeed a transition probability kernel from Z to X. □

Proof of Lemma 5.13. For every bounded measurable funciton $g\colon Y \to \mathbb{R}$, by Proposition 5.9 and Definition 5.4, we have

$$\int_Y g\,\mathrm{d}(\mu\mathcal{Q}') = \int_Z \mathcal{Q}'g\,\mathrm{d}\mu(z) = \int_Z \biggl(\int_Y g\,\mathrm{d}\mathcal{Q}'_z\biggr)\,\mathrm{d}\mu(z).$$

For each measurable set A in X, applying the equality above we have

$$\begin{aligned}
(\mu(\mathcal{Q}'\mathcal{Q}))(A) &= \int_Z (\mathcal{Q}'\mathcal{Q})(z, A)\,\mathrm{d}\mu(z) \\
&= \int_Z (\mathcal{Q}'_z\mathcal{Q})(A)\,\mathrm{d}\mu(z) \\
&= \int_Z \biggl(\int_Y \mathcal{Q}(y, A)\,\mathrm{d}\mathcal{Q}'_z(y)\biggr)\,\mathrm{d}\mu(z) \\
&= \int_Y \mathcal{Q}(y, A)\,\mathrm{d}(\mu\mathcal{Q}')(y) \\
&= ((\mu\mathcal{Q}')\mathcal{Q})(A).
\end{aligned}$$

Therefore, $\mu(\mathcal{Q}'\mathcal{Q}) = (\mu\mathcal{Q}')\mathcal{Q}$. □

Lemma A.2. *The map $\mathcal{Q}^{[n]}$ defined in Definition 5.14 is indeed a transition probability kernel from $(X, \mathscr{M}(X))$ to $(X^{n+1}, \mathscr{M}(X^{n+1}))$.*

Proof. First, $\mathcal{Q}^{[0]} = \widehat{\mathrm{id}_X}$ is a transition probability kernel from X to X.

Suppose that we have varified that $\mathcal{Q}^{[n-1]}$ is a transition probability kernel from X to X^n for some $n \in \mathbb{N}$. Now we focus on proving that $\mathcal{Q}^{[n]}$ is a transition probability kernel from X to X^{n+1}.

For each $A_{n+1} \in \mathscr{M}(X^{n+1})$, write $L_{A_{n+1}}(\underline{y}) \coloneqq \mathcal{Q}(x_n, \pi_{n+1}(x_1, \dots, x_n; A_{n+1}))$ for all $\underline{y} = (x_1, \dots, x_n) \in X^n$. We claim that $L_{A_{n+1}}$ is a measurable function on $(X^n, \mathscr{M}(X^n))$ for every $A_{n+1} \in \mathscr{M}(X^{n+1})$.

Indeed, we consider the case $A_{n+1} = B_1 \times \cdots \times B_n \times B_{n+1}$ first, where $B_1, \dots, B_{n+1} \in \mathscr{M}(X)$ are arbitrary measurable sets. In this case,

$$\pi_{n+1}(x_1, \dots, x_n; A_{n+1}) = \begin{cases} B_{n+1}, & \text{if } x_i \in B_i \text{ for all } i \in (n]\,, \\ \emptyset, & \text{otherwise} \end{cases}$$

holds for all $(x_1, \dots, x_n) \in X^n$, and thus $L_{A_{n+1}}(x_1, \dots, x_n) = \mathcal{Q}(x_n, B_{n+1}) \prod_{i=1}^{n} \mathbb{1}_{B_i}(x_i)$. Since $\mathcal{Q}(x_n, B_{n+1})$ as a function of x_n is measurable on X and since $\mathbb{1}_{B_i}$ is measurable on X for all $i \in (n]$, the function $L_{A_{n+1}}(x_1, \dots, x_n)$, their product, is measurable on X^n.

Denote by $\mathcal{D}$ a subset of $\mathscr{M}(X^{n+1})$ consisting of all measurable sets $A \in \mathscr{M}(X^{n+1})$ satisfying that L_A is measurable on $(X^n, \mathscr{M}(X^n))$. Then the claim is equivalent to $\mathcal{D} = \mathscr{M}(X^{n+1})$. We have shown $\mathcal{G} \coloneqq \{B_1 \times \cdots \times B_{n+1} : B_1, \dots, B_{n+1} \in \mathscr{M}(X)\} \subseteq \mathcal{D}$.

Let $(x_1, \dots, x_n) \in X^n$ and $A \in \mathscr{M}(X^{n+1})$. From (5.3) we can see

$$\pi_{n+1}(x_1, \dots, x_n; A^c) = \pi_{n+1}(x_1, \dots, x_n; A)^c.$$

Thus we have $L_{A^c} = 1 - L_A$. This indicates that $A \in \mathcal{D}$ implies $A^c \in \mathcal{D}$. In other words, $\mathcal{D}$ is closed under complement.

Let $(x_1, \dots, x_n) \in X^n$ and a countable collection of mutually disjoint measurable sets $A_1, A_2, \dots \in \mathscr{M}(X^{n+1})$ be arbitrary. From (5.3) we can see

$$\pi_{n+1}\bigl(x_1, \dots, x_n; \bigcup_{i=1}^{+\infty} A_i\bigr) = \bigcup_{i=1}^{+\infty} \pi_{n+1}(x_1, \dots, x_n; A_i),$$

and thus $L_{\bigcup_{i=1}^{+\infty} A_i} = \sum_{i=1}^{+\infty} L_{A_i}$. This shows that $\mathcal{D}$ is closed under countable disjoint union.

Thus $\mathcal{D}$ is a Dynkin system. Note that $\mathcal{G} = \{B_1 \times \cdots \times B_{n+1} : B_1, \dots, B_n \in \mathscr{M}(X)\}$ is a π-system because it is non-empty and closed under intersection. By the Dynkin's π-λ Theorem, $\mathcal{G} \subseteq \mathcal{D}$ implies that $\mathcal{D}$ contains the σ-algebra generated by $\mathcal{G}$, i.e., $\mathscr{M}(X^{n+1}) \subseteq \mathcal{D}$. Since $\mathcal{D} \subseteq \mathscr{M}(X^{n+1})$, we conclude $\mathcal{D} = \mathscr{M}(X^{n+1})$. The claim is therefore established.

This measurability ensures that the integral in (5.4) is well-defined. Moreover, $L_{A_{n+1}}$ is bounded because its range is contained in $[0,1]$, so by Definition 5.4 we have $\mathcal{Q}^{[n]}(x, A_{n+1}) = \int_{\underline{y} \in X^n} L_{A_{n+1}}(\underline{y}) \,\mathrm{d}\mathcal{Q}_x^{[n-1]}(\underline{y}) = \bigl(\mathcal{Q}^{[n-1]} L_{A_{n+1}}\bigr)(x)$. By Lemma 5.5, $\mathcal{Q}^{[n]}(x, A_{n+1}) = \mathcal{Q}^{[n-1]} L_{A_{n+1}}(x)$ is measurable as a function on $x \in X$.

Now we fix an arbitrary point $x \in X$, we shall show that $\mathcal{Q}^{[n]}(x, \cdot)$, the map that assigns each measurable set $A_{n+1} \in \mathscr{M}(X^{n+1})$ the value $\mathcal{Q}^{[n]}(x, A_{n+1})$, is a probability measure on $(X^{n+1}, \mathscr{M}(X^{n+1}))$.

By (5.1) and Lemma 5.5, if measurable sets $A_1, A_2, \dots \in \mathscr{M}(X^{n+1})$ are mutually disjoint, then

$$\begin{aligned}\mathcal{Q}^{[n]}\Bigg(x, \bigcup_{i=1}^{+\infty} A_i\Bigg) &= \big(\mathcal{Q}^{[n-1]} L_{\bigcup_{i=1}^{+\infty} A_i}\big)(x) \\ &= \Bigg(\mathcal{Q}^{[n-1]}\Bigg(\sum_{i=1}^{+\infty} L_{A_i}\Bigg)\Bigg)(x) \\ &= \sum_{i=1}^{+\infty} \big(\mathcal{Q}^{[n-1]} L_{A_i}\big)(x) \\ &= \sum_{i=1}^{+\infty} \mathcal{Q}^{[n]}(x, A_i).\end{aligned}$$

Moreover, because

$$\mathcal{Q}(x_n, \pi_{n+1}(x_1, \dots, x_n; X^{n+1})) = \mathcal{Q}(x_n, X) = 1$$

holds for all $(x_1, \dots, x_n) \in X^n$, by (5.4) we have

$$\mathcal{Q}^{[n]}(x, X^{n+1}) = \int_{(x_1,\dots,x_n)\in X^n} \mathrm{d}\mathcal{Q}_x^{[n-1]}(x_1, \dots, x_n) = 1.$$

Hence $\mathcal{Q}^{[n]}(x, \cdot)$ is a probability measure on $(X^{n+1}, \mathscr{M}(X^{n+1}))$.

We have checked the two conditions in Definition 5.1 and conclude that $\mathcal{Q}^{[n]}$ is a transition probability kernel from X to X^{n+1}. By induction on n, therefore, we conclude that $\mathcal{Q}^{[n]}$ is indeed a transition probability kernel from X to X^{n+1} for all $n \in \mathbb{N}_0$. $\square$

Proof of Lemma 5.17. If $m = 0$, (5.5) is trivial. We only consider $m \in \mathbb{N}$ below.

Through induction on $m \in \mathbb{N}$, we can reduce (5.5) to the case in which $m = 1$, i.e., $\mathcal{Q}^{[n]}(x, A) = \mathcal{Q}^{[n+1]}(x, A \times X)$. We check this equality using (5.4), the inductively definition of $\mathcal{Q}^{[n+1]}$:

$$\begin{aligned}&\mathcal{Q}^{[n+1]}(x, A \times X) \\ &\quad= \int_{X^{n+1}} \mathcal{Q}(x_{n+1}, \pi_{n+2}(x_1, \dots, x_{n+1}; A \times X)) \,\mathrm{d}\mathcal{Q}_x^{[n]}(x_1, \dots, x_{n+1}) \\ &\quad= \int_A \mathcal{Q}(x_{n+1}, X) \,\mathrm{d}\mathcal{Q}_x^{[n]}(x_1, \dots, x_{n+1}) + \int_{A^c} \mathcal{Q}(x_{n+1}, \emptyset) \,\mathrm{d}\mathcal{Q}_x^{[n]}(x_1, \dots, x_n) \\ &\quad= \mathcal{Q}_x^{[n]}(A) \\ &\quad= \mathcal{Q}^{[n]}(x, A).\end{aligned}$$

By induction, therefore, (5.5) holds for all $m \in \mathbb{N}$. $\square$

Lemma A.3. *The transition probability kernel $\mathcal{Q}^\omega$ satisfying (5.6) in Definition 5.18 exists and is unique.*

Proof. By the Kolmogrov extension theorem, Lemma 5.17 indicates that for each $x \in X$, there exists a unique probability measure μ_x on $(X^\omega, \mathscr{M}(X^\omega))$ satisfying

$$\mu_x(A \times X^\omega) = \mathcal{Q}^{[n]}(x, A) \tag{A.1}$$

for arbitrary measurable set $A \in \mathscr{M}(X^{n+1})$ and $n \in \mathbb{N}_0$.

Write $\mathscr{B} := \{A \times X^\omega : A \in \mathscr{M}(X^n) \text{ for some } n \in \mathbb{N}\}$, a π-system on X. Denote by $\mathcal{D}'$ the set consisting of all measurable sets $B \in \mathscr{M}(X^\omega)$ satisfying that the function which assigns each point $x \in X$ the value $\mu_x(B)$ is measurable. The equality (A.1) implies $\mathscr{B} \subseteq \mathcal{D}'$.

Fix an arbitrary $x \in X$. Since $\mu_x(B^c) = 1 - \mu_x(B)$ holds for an arbitrary measurable set $B \in \mathscr{M}(X^\omega)$ and $\mu_x\bigl(\bigcup_{i=1}^{+\infty} B_i\bigr) = \sum_{i=1}^{+\infty} \mu_x(B_i)$ holds for an arbitrary collection of mutually disjoint measurable sets $B_1, B_2, \dots \in \mathscr{M}(X^\omega)$, the collection $\mathcal{D}'$ of some measurable subsets of X^ω is closed under complement and countable disjoint union. Thus $\mathcal{D}'$ is a Dynkin system. Since $\mathscr{M}(X^\omega)$ is the σ-algebra generated by the π-system $\mathscr{B}$ and $\mathscr{B}$ is contained in $\mathcal{D}'$, we get $\mathscr{M}(X^\omega) \subseteq \mathcal{D}'$ by the Dynkin's π-λ Theorem. Thus, $\mathcal{D}' = \mathscr{M}(X^\omega)$, i.e., for every $B \in \mathscr{M}(X^\omega)$, the function that assigns each $x \in X$ the value $\mu_x(B)$ is measurable. Hence, if we set $\mathcal{Q}^\omega(x, B) = \mu_x(B)$ for each $x \in X$ and $B \in \mathscr{M}(X^\omega)$, it is a transition probability kernel from X to X^ω satisfying (5.6) in Definition 5.18.

The uniqueness of the transition probability kernel $\mathcal{Q}^\omega$ from X to X^ω that satisfies (5.6) is because for each $x \in X$, (5.6) determines the probability measure $\mathcal{Q}^\omega_x$ by determining its value on the algebra $\mathcal{K} := \{A \times X^\omega : A \in X^n \text{ for some } n \in \mathbb{N}\}$, by which $\mathscr{M}(X^\omega)$ is the σ-algebra generated. □

Now we establish some more properties of transition probability kernels.

Corollary A.4. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$ and $\mu \in \mathcal{P}(X)$. We have*

$$\bigl(\mu\mathcal{Q}^{[1]}\bigr) \circ \widetilde{\pi}_1^{-1} = \mu. \tag{A.2}$$

Proof. If we take $n = 0$ and $m = 1$ in Lemma 5.17, then we can get $\mathcal{Q}^{[1]}(x, A \times X) = \mathcal{Q}^{[0]}(x, A) = \widehat{\mathrm{id}_X}(x, A)$ for all $x \in X$ and $A \in \mathscr{M}(X)$. Thus by Definition 5.6, we have

$$\begin{aligned}
\bigl(\mu\mathcal{Q}^{[1]}\bigr)(A \times X) &= \int_X \mathcal{Q}^{[1]}(x, A \times X) \,\mathrm{d}\mu(x) \\
&= \int_X \widehat{\mathrm{id}_X}(x, A) \,\mathrm{d}\mu(x) \\
&= \int_X \mathbb{1}_A(x) \,\mathrm{d}\mu(x) \\
&= \mu(A)
\end{aligned}$$

for all $A \in \mathscr{M}(X)$. Therefore, we get $\bigl(\mu\mathcal{Q}^{[1]}\bigr) \circ \widetilde{\pi}_1^{-1} = \mu$. □

Lemma A.5. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $n \in \mathbb{N}$, and $\mu \in \mathcal{P}(X)$. If $B \in \mathscr{M}(X^{n+1})$, then*

$$\big(\mu\mathcal{Q}^{[n]}\big)(B) = \int_{X^n} \mathcal{Q}(x_{n+1}, \pi_{n+1}(x_2, \dots, x_{n+1}; B)) \,\mathrm{d}\big(\mu\mathcal{Q}^{[n-1]}\big)(x_2, \dots, x_{n+1}). \tag{A.3}$$

If $A_0, A_1, \dots, A_n \in \mathscr{M}(X)$, then

$$\big(\mu\mathcal{Q}^{[n]}\big)(A_0 \times \cdots \times A_n) = \int_{A_0 \times \cdots \times A_{n-1}} \mathcal{Q}(x_n, A_n) \,\mathrm{d}\big(\mu\mathcal{Q}^{[n-1]}\big)(x_1, \dots, x_n). \tag{A.4}$$

Proof. By Proposition 5.9 and Definitions 5.4, 5.14, and 5.6,

$$\begin{aligned}
&\int_{X^n} \mathcal{Q}(x_{n+1}, \pi_{n+1}(x_2, \dots, x_{n+1}; B)) \,\mathrm{d}\big(\mu\mathcal{Q}^{[n-1]}\big)(x_2, \dots, x_{n+1}) \\
&\qquad = \int_X \bigg(\int_{X^n} \mathcal{Q}(x_n, \pi_{n+1}(x_1, \dots, x_n; B)) \,\mathrm{d}\mathcal{Q}_y^{[n-1]}(x_1, \dots, x_n) \bigg) \,\mathrm{d}\mu(y) \\
&\qquad = \int_X \mathcal{Q}^{[n]}(y, B) \,\mathrm{d}\mu(y) \\
&\qquad = \big(\mu\mathcal{Q}^{[n]}\big)(B).
\end{aligned}$$

Therefore, (A.3) holds, and (A.4) follows by taking $B = A_0 \times \cdots \times A_n$ in (A.3). $\square$

Lemma A.6. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $n \in \mathbb{N}$, and B be a measurable set in $\mathscr{M}(X^n)$. For each $x \in X$, we have*

$$\big(\mathcal{Q}\mathcal{Q}^{[n-1]}\big)(x, B) = \mathcal{Q}^{[n]}(x, X \times B). \tag{A.5}$$

In particular, for each $x \in X$ and each $B \in \mathscr{M}(X)$, we have

$$\mathcal{Q}(x, B) = \mathcal{Q}^{[1]}(x, X \times B). \tag{A.6}$$

Moreover, for each $x \in X$ and each measurable set $A \in \mathscr{M}(X^\omega)$, we have

$$(\mathcal{Q}\mathcal{Q}^\omega)(x, A) = \mathcal{Q}^\omega(x, X \times A). \tag{A.7}$$

Proof. We prove (A.5) by induction on n.

If $n = 1$, by Definitions 5.11 and 5.6, we have

$$\big(\mathcal{Q}\mathcal{Q}^{[0]}\big)(x, B) = \big(\mathcal{Q}_x \widehat{\mathrm{id}_X}\big)(x, B) = \int_X \widehat{\mathrm{id}_X}(y, B) \,\mathrm{d}\mathcal{Q}_x(y) = \mathcal{Q}_x(B) = \mathcal{Q}(x, B).$$

Moreover, by (5.4), we have

$$\mathcal{Q}^{[1]}(x, X \times B) = \int_X \mathcal{Q}(y, \pi_1(y; X \times B)) \,\mathrm{d}\mathcal{Q}_x^{[0]}(y) = \int_X \mathcal{Q}(y, B) \,\mathrm{d}\delta_x(y) = \mathcal{Q}(x, B),$$

where δ_x is the Dirac measure on $(X, \mathscr{M}(X))$ at x. Thus $\big(\mathcal{Q}\mathcal{Q}^{[0]}\big)(x, B) = \mathcal{Q}^{[1]}(x, X \times B)$.

Suppose that (A.5) holds for some $n \in \mathbb{N}$. Let $x \in X$ and $B \in \mathscr{M}(X^{n+1})$ be arbitrary. Since $\pi_{n+2}(x_1, x_2, \dots, x_{n+1}; X \times B) = \pi_{n+1}(x_2, \dots, x_{n+1}; B)$, by (5.4) we have

$$\mathcal{Q}^{[n+1]}(x, X \times B) = \int_{X^{n+1}} \mathcal{Q}(x_{n+1}, \pi_{n+1}(x_2, \dots, x_{n+1}; B)) \,\mathrm{d}\mathcal{Q}_x^{[n]}(x_1, x_2, \dots, x_{n+1}).$$

Note by the induction hypothesis that $\mathcal{Q}_x^{[n]}(X\times C) = \big(\mathcal{Q}\mathcal{Q}^{[n-1]}\big)_x(C) = \big(\mathcal{Q}_x\mathcal{Q}^{[n-1]}\big)(C)$ holds for all $C \in \mathscr{M}(X^n)$, which means that $\mathcal{Q}_x\mathcal{Q}^{[n-1]}$ is the projection of $\mathcal{Q}_x^{[n]}$ from X^{n+1} onto the last n coordinates. This leads to the following:

$$\mathcal{Q}^{[n+1]}(x, X \times B) = \int_{X^n} \mathcal{Q}(x_{n+1}, \pi_{n+1}(x_2, \dots, x_{n+1}; B)) \,\mathrm{d}\big(\mathcal{Q}_x\mathcal{Q}^{[n-1]}\big)(x_2, \dots, x_{n+1})$$

Hence by (A.3) in Lemma A.5 and Definition 5.11, we conclude $\mathcal{Q}^{[n+1]}(x, X\times B) = \big(\mathcal{Q}_x\mathcal{Q}^{[n]}\big)(B) = \big(\mathcal{Q}\mathcal{Q}^{[n]}\big)(x, B)$. Therefore, (A.5) follows for all $n \in \mathbb{N}$.

Applying the Dynkin's π-λ Theorem, we can reduce (A.7) to the cases where A belongs to $\{B \times X^\omega : B \in \mathscr{M}(X^n) \text{ for some } n \in \mathbb{N}\}$, a π-system by which $\mathscr{M}(X^\omega)$ is the σ-algebra generated.

Let $n \in \mathbb{N}$ and $B \in \mathscr{M}(X^n)$ be arbitrary. By (5.7), (A.5), and (5.6), we have

$$\begin{aligned}(\mathcal{Q}\mathcal{Q}^\omega)(x, B \times X^\omega) &= (\mathcal{Q}_x\mathcal{Q}^\omega)(B \times X^\omega)\\ &= \big(\mathcal{Q}_x\mathcal{Q}^{[n-1]}\big)(B)\\ &= \big(\mathcal{Q}\mathcal{Q}^{[n-1]}\big)(x, B)\\ &= \mathcal{Q}^{[n]}(x, X \times B)\\ &= \mathcal{Q}^\omega(x, X \times B \times X^\omega).\end{aligned}$$

Therefore (A.7) holds for all measurable sets $A \in \mathscr{M}(X^\omega)$ by the Dynkin's π-λ Theorem. $\square$

We have the following corollary by Lemma A.6 and Definition 5.6.

Corollary A.7. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $n \in \mathbb{N}$, $B \in \mathscr{M}(X^n)$, and μ be a probability measure on $(X, \mathscr{M}(X))$. Then*

$$\big(\mu\mathcal{Q}\mathcal{Q}^{[n-1]}\big)(B) = \big(\mu\mathcal{Q}^{[n]}\big)(X \times B). \tag{A.8}$$

Moreover, we have $(\mu\mathcal{Q}\mathcal{Q}^\omega)(B) = (\mu\mathcal{Q}^\omega)(X \times B)$. Additionally, if μ is $\mathcal{Q}$-invariant, then

$$(\mu\mathcal{Q}^\omega)(B) = (\mu\mathcal{Q}^\omega)(X \times B). \tag{A.9}$$

If we take $n = 1$ in (A.8), we get

$$\big(\mu\mathcal{Q}^{[1]}\big) \circ \widetilde{\pi}_2 = \mu\mathcal{Q}. \tag{A.10}$$

Lemma A.8. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $n \in \mathbb{N}_0$, and $f \colon X^{n+2} \to \mathbb{R}$ be a measurable function, then we have $\mathcal{Q}_y^{[n]}(\{y\}\times X^n) = 1$ and*

$$\begin{aligned}&\int_{X^{n+1}} f(x_0, x_0, x_1, \dots, x_n) \,\mathrm{d}\mathcal{Q}_y^{[n]}(x_0, \dots, x_n)\\ &\qquad = \int_{X^{n+1}} f(y, x_0, x_1, \dots, x_n) \,\mathrm{d}\mathcal{Q}_y^{[n]}(x_0, \dots, x_n)\end{aligned}$$

for all $y \in X$.

Proof. First, by Lemma 5.5, for each $y \in X$ we have

$$\mathcal{Q}_y^{[n]}(\{y\} \times X^n) = \mathcal{Q}^{[0]}(y, \{y\}) = \widehat{\mathrm{id}_X}(y, \{y\}) = 1.$$

Hence the following equality

$$\begin{aligned}\int_{X^{n+1}} f(x_0, x_0, \dots, x_n) \,\mathrm{d}\mathcal{Q}_y^{[n]}(x_0, \dots, x_n) &= \int_{\{y\}\times X^n} f(x_0, x_0, \dots, x_n) \,\mathrm{d}\mathcal{Q}_y^{[n]}(x_0, \dots, x_n) \\ &= \int_{\{y\}\times X^n} f(y, x_0, \dots, x_n) \,\mathrm{d}\mathcal{Q}_y^{[n]}(x_0, \dots, x_n) \\ &= \int_{X^{n+1}} f(y, x_0, \dots, x_n) \,\mathrm{d}\mathcal{Q}_y^{[n]}(x_0, \dots, x_n)\end{aligned}$$

holds for all $y \in X$. □

Lemma A.9. *Let $\mathcal{Q}$ be a transition probability kernel on a measurable space $(X, \mathscr{M}(X))$, $\phi \in B(X^2, \mathbb{R})$, and $\mu \in \mathcal{P}(X)$. Then we have*

$$\begin{aligned}\int_{X^\omega} \phi(x_1, x_2) \,\mathrm{d}(\mu\mathcal{Q}^\omega)(x_1, x_2, \dots) &= \int_{X^2} \phi \,\mathrm{d}\big(\mu\mathcal{Q}^{[1]}\big) \\ &= \int_X \int_X \phi(x_1, x_2) \,\mathrm{d}\mathcal{Q}_{x_1}(x_2) \,\mathrm{d}\mu(x_1).\end{aligned} \tag{A.11}$$

Proof. By taking $n = 1$ in (5.7), we get $\int_{X^\omega} \phi(x_1, x_2) \,\mathrm{d}(\mu\mathcal{Q}^\omega)(x_1, x_2, \dots) = \int_{X^2} \phi \,\mathrm{d}\big(\mu\mathcal{Q}^{[1]}\big)$.

Moreover, by Proposition 5.9 and Definition 5.4, we have

$$\int_{X^2} \phi \,\mathrm{d}\big(\mu\mathcal{Q}^{[1]}\big) = \int_X \mathcal{Q}^{[1]}\phi \,\mathrm{d}\mu = \int_X \int_{X^2} \phi(x_1, x_2) \,\mathrm{d}\mathcal{Q}_y^{[1]}(x_1, x_2) \,\mathrm{d}\mu(y) \tag{A.12}$$

By Lemma A.8 and (A.6) in Lemma A.6,

$$\int_{X^2} \phi(x_1, x_2) \,\mathrm{d}\mathcal{Q}_y^{[1]}(x_1, x_2) = \int_{X^2} \phi(y, x_2) \,\mathrm{d}\mathcal{Q}_y^{[1]}(x_1, x_2) = \int_X \phi(y, x_2) \,\mathrm{d}\mathcal{Q}_y(x_2). \tag{A.13}$$

Therefore, by (A.12) and (A.12), we have

$$\int_{X^2} \phi \,\mathrm{d}\big(\mu\mathcal{Q}^{[1]}\big) = \int_X \int_X \phi(x_1, x_2) \,\mathrm{d}\mathcal{Q}_{x_1}(x_2) \,\mathrm{d}\mu(x_1),$$

establishing (A.11). □

Recall for each $n \in \mathbb{N}$, $\iota_n(x_1, x_2, \dots, x_n) = (x_n, x_{n-1}, \dots, x_1)$ for all $(x_1, \dots, x_n) \in X^n$.

Lemma A.10. *Let $\mathcal{Q}, \mathcal{R}$ be transition probability kernels on a measurable space $(X, \mathscr{M}(X))$ and $\mu \in \mathcal{P}(X)$. If $\mu\mathcal{Q}^{[1]} = \big(\mu\mathcal{R}^{[1]}\big) \circ \iota_2^{-1}$, then $\mu \in \mathcal{M}(X, \mathcal{Q}) \cap \mathcal{M}(X, \mathcal{R})$ and $\mu\mathcal{Q}^{[n]} = \big(\mu\mathcal{R}^{[n]}\big) \circ \iota_{n+1}^{-1}$ for all $n \in \mathbb{N}$.*

Proof. Since $\mu\mathcal{Q}^{[1]} = \big(\mu\mathcal{R}^{[1]}\big) \circ \iota_2^{-1}$, we have $\big(\mu\mathcal{Q}^{[1]}\big) \circ \widetilde{\pi}_1 = \big(\mu\mathcal{R}^{[1]}\big) \circ \widetilde{\pi}_2$ and $\big(\mu\mathcal{Q}^{[1]}\big) \circ \widetilde{\pi}_2 = \big(\mu\mathcal{R}^{[1]}\big) \circ \widetilde{\pi}_1$. Thus by (A.2) and (A.10), we have $\mu = \mu\mathcal{R}$ and $\mu\mathcal{Q} = \mu$, i.e., $\mu \in \mathcal{M}(X, \mathcal{Q}) \cap \mathcal{M}(X, \mathcal{R})$.

Now we prove $\mu\mathcal{Q}^{[n]} = \big(\mu\mathcal{R}^{[n]}\big) \circ \iota_{n+1}^{-1}$ by induction on $n \in \mathbb{N}$. The case $n = 0$ holds by hypothesis.

Suppose $\mu\mathcal{Q}^{[n-1]} = \bigl(\mu\mathcal{R}^{[n-1]}\bigr) \circ \iota_n^{-1}$ holds for some $n \in \mathbb{N}$, $n \geqslant 2$. In order to show $\mu\mathcal{Q}^{[n]} = \bigl(\mu\mathcal{R}^{[n]}\bigr) \circ \iota_{n+1}^{-1}$, it is enough to prove $\bigl(\mu\mathcal{Q}^{[n]}\bigr)(A_0 \times \cdots \times A_n) = \bigl(\mu\mathcal{R}^{[n]}\bigr)(A_n \times \cdots \times A_0)$ for all $A_0, A_1, \dots, A_n \in \mathscr{M}(X)$.

Fix arbitrary $A_0, \dots, A_n \in \mathscr{M}(X)$. In this proof, we write $A_i^n := A_i \times \cdots \times A_n$, $A_n^i := A_n \times \cdots \times A_i$, $\underline{x}_i^n := (x_i, \dots, x_n)$, and $\underline{x}_n^i := (x_n, \dots, x_i)$ for $i = 0$ and $i = 1$.

$$
\begin{aligned}
&\bigl(\mu\mathcal{R}^{[n]}\bigr)(A_n^0)\\
&\quad= \int_{A_n^1} \mathcal{R}(x_1, A_0)\,\mathrm{d}\bigl(\mu\mathcal{R}^{[n-1]}\bigr)(\underline{x}_n^1) && \text{(by (A.4))}\\
&\quad= \int_{A_1^n} \mathcal{R}(x_1, A_0)\,\mathrm{d}\bigl(\mu\mathcal{Q}^{[n-1]}\bigr)(\underline{x}_1^n) && \text{(by } \mu\mathcal{Q}^{[n-1]} = \bigl(\mu\mathcal{R}^{[n-1]}\bigr) \circ \iota_{n-1}^{-1}\text{)}\\
&\quad= \int_X \int_{A_1^n} \mathcal{R}_{x_1}(A_0)\,\mathrm{d}\mathcal{Q}_y^{[n-1]}(\underline{x}_1^n)\,\mathrm{d}\mu(y) && \text{(by Proposition 5.9 and Definition 5.4)}\\
&\quad= \int_X \mathcal{R}_y(A_0)\mathcal{Q}_y^{[n-1]}(A_1^n)\,\mathrm{d}\mu(y) && \text{(by Lemma A.8)}\\
&\quad= \int_X \int_{A_0} \mathcal{Q}_y^{[n-1]}(A_1^n)\,\mathrm{d}\mathcal{R}_y(x_0)\,\mathrm{d}\mu(y)\\
&\quad= \int_{X \times A_0} \mathcal{Q}_y^{[n-1]}(A_1^n)\,\mathrm{d}\bigl(\mu\mathcal{R}^{[1]}\bigr)(y, x_0) && \text{(by (A.11))}\\
&\quad= \int_{A_0 \times X} \mathcal{Q}_y^{[n-1]}(A_1^n)\,\mathrm{d}\bigl(\mu\mathcal{Q}^{[1]}\bigr)(x, y) && \text{(by } \mu\mathcal{Q}^{[1]} = \bigl(\mu\mathcal{R}^{[1]}\bigr) \circ \iota_2^{-1}\text{)}\\
&\quad= \int_{A_0} \int_X \mathcal{Q}_y^{[n-1]}(A_1^n)\,\mathrm{d}\mathcal{Q}_x(y)\,\mathrm{d}\mu(x) && \text{(by (A.11))}\\
&\quad= \int_{A_0} \bigl(\mathcal{Q}\mathcal{Q}^{[n-1]}\bigr)(x, A_1^n)\,\mathrm{d}\mu(x) && \text{(by Definitions 5.6 and 5.11)}\\
&\quad= \int_{A_0} \mathcal{Q}^{[n]}(x, X \times A_1^n)\,\mathrm{d}\mu(x) && \text{(by (A.5))}\\
&\quad= \int_X \mathcal{Q}^{[n]}(x, A_0 \times A_1^n)\,\mathrm{d}\mu(x) && \text{(by } \mathcal{Q}^{[n]}(x, \{x\} \times X^n) = 1 \text{ in Lemma A.8)}\\
&\quad= \bigl(\mu\mathcal{Q}^{[n]}\bigr)(A_0^n) && \text{(by Definition 5.6).}
\end{aligned}
$$

Hence we conclude $\mu\mathcal{Q}^{[n]} = \bigl(\mu\mathcal{R}^{[n]}\bigr) \circ \iota_{n+1}^{-1}$, and therefore, Lemma A.10 follows. ☐

A.2. **Conditional transition probability kernels.** In Section 7, we always need to address the following question: for a probability measure ν on X^2, how to find a probability measure μ on X and a transition probability kernel $\mathcal{Q}$ on X such that $\nu = \mu\mathcal{Q}^{[1]}$? This appendix is devoted to discussing related theories about conditional transition probability kernels.

Proposition A.11. *Let X_1 and X_2 be compact metric spaces, M be a non-empty closed subset of $X_1 \times X_2$, ν be a Borel probability measure on $X_1 \times X_2$ supported on M (i.e., $\nu(M) = 1$), and $\kappa \colon X_1 \times X_2 \to X_1$ be the projection map given by $\kappa(x_1, x_2) = x_1$*

for all $x_1 \in X_1$ and $x_2 \in X_2$. There exists a Borel probability measure μ on X_1 and a transition probability kernel $\mathcal{Q}$ from X_1 to X_2 with the following properties:

(a) *For each $x_1 \in \kappa(M)$, we have*
$$\mathcal{Q}(x_1, \{x_2 \in X_2 : (x_1, x_2) \in M\}) = 1.$$

(b) *For each $C \in \mathscr{B}(X_1 \times X_2)$, we have*
$$\nu(C) = \int_{X_1} \mathcal{Q}(x_1, \{x_2 \in X_2 : (x_1, x_2) \in C\}) \, \mathrm{d}\mu(x_1).$$

Moreover, μ must be $\nu \circ \kappa^{-1}$, and $\mathcal{Q}$ is unique in the sense that if both μ, $\mathcal{Q}$ and μ, $\mathcal{Q}'$ satisfy the properties (a) and (b), then for μ-almost every $x_1 \in X_1$ and all $B \in \mathscr{B}(X_2)$, the equality $\mathcal{Q}(x_1, B) = \mathcal{Q}'(x_1, B)$ holds.

Remark A.12. We list three properties equivalent to the property (b) in Proposition A.11 for the Borel probability measure μ on X_1 and the transition probability kernel $\mathcal{Q}$ from X_1 to X_2:

(b1) For each $A \in \mathscr{B}(X_1)$ and each $B \in \mathscr{B}(X_2)$, the following equality holds:
$$\nu(A \times B) = \int_A \mathcal{Q}(x_1, B) \, \mathrm{d}\mu(x_1). \tag{A.14}$$

(b2) There exist some π-systems $\mathfrak{A}_1 \subseteq \mathscr{B}(X_1)$ and $\mathfrak{A}_2 \subseteq \mathscr{B}(X_2)$ with the following property:
   (i) The σ-algebra generated by $\mathfrak{A}_i$ is $\mathscr{B}(X_i)$ for $i = 1$ and 2.
   (ii) For each $A \in \mathfrak{A}_1$ and each $B \in \mathfrak{A}_2$, the equality (A.14) holds.

(b3) For each lower bounded Borel measurable function $f \colon X_1 \times X_2 \to \mathbb{R} \cup \{+\infty\}$, we have
$$\int_{X_1 \times X_2} f(x_1, x_2) \, \mathrm{d}\nu(x_1, x_2) = \int_{X_1} \left(\int_{X_2} f(x_1, x_2) \, \mathrm{d}\mathcal{Q}_{x_1}(x_2) \right) \mathrm{d}\mu(x_1). \tag{A.15}$$

The equivalence of properties (b), (b1), and (b2) can be verified by the Dynkin's π-λ Theorem. Clearly (b3) implies (b). We explain why (b) implies (b3):

Suppose (b) holds for μ and $\mathcal{Q}$. Property (b) implies that the equality (A.15) holds when f is a characteristic function of an arbitrary Borel subset of $X_1 \times X_2$, and thus by (5.1) and Lemma 5.5 the equality (A.15) holds when f is an arbitrary simple function on $X_1 \times X_2$. Because each lower bounded Borel measurable function on $X_1 \times X_2$ can be pointwise approached by an increasing sequence of bounded simple functions, the equality (A.15) holds for all lower bounded Borel measurable functions $f \colon X_1 \times X_2 \to \mathbb{R} \cup \{+\infty\}$.

Let us now proceed with the proof of Proposition A.11.

Proof of Proposition A.11. Denote by $\mathscr{B}_1$ the σ-algebra on $X_1 \times X_2$ given by $\mathscr{B}_1 \coloneqq \{A \times X_2 : A \in \mathscr{B}(X_1)\}$. By [GS74, Theorem I.3.3], there exists a function $\mathcal{Q}_2$ defined on $X_1 \times X_2 \times \mathscr{B}(X_1 \times X_2)$ satisfying:

(i) Fix an arbitrary Borel set $B \in \mathscr{B}(X_1 \times X_2)$, the map that assigns each point $(x_1, x_2) \in X_1 \times X_2$ the value $\mathcal{Q}_2(x_1, x_2, B)$ is $\mathscr{B}_1$-measurable.

(ii) Fix an arbitrary point $(x_1, x_2) \in X_1 \times X_2$, the map that assigns each Borel set $B \in \mathscr{B}(X_1 \times X_2)$ the value $\mathcal{Q}_2(x_1, x_2, B)$ is a probability measure on $X_1 \times X_2$.

(iii) For each $B \in \mathscr{B}(X_1 \times X_2)$ and each $B' \in \mathscr{B}_1$, the following equality holds:

$$\int_{B'} \mathcal{Q}_2(x_1, x_2, B) \,\mathrm{d}\nu(x_1, x_2) = \nu(B \cap B'). \tag{A.16}$$

The condition (i) implies that $\mathcal{Q}_2(x_1, x_2, B)$ does not depend on x_2, so we set $\mathcal{Q}_1(x_1, B) \coloneqq \mathcal{Q}_2(x_1, x_2, B)$. Let $\mu = \nu \circ \kappa^{-1}$. For each $A \in \mathscr{B}(X_1)$ and each $B \in \mathscr{B}(X_1 \times X_2)$, we can rewrite (A.16) as

$$\int_{A} \mathcal{Q}_1(x_1, B) \,\mathrm{d}\mu(x_1) = \nu(B \cap A \times X_2). \tag{A.17}$$

By taking $B = A^c \times X_2$ in (A.17) we get $\int_A \mathcal{Q}_1(x_1, A^c \times X_2) \,\mathrm{d}\mu(x_1) = \nu(\emptyset) = 0$, which implies $\mathcal{Q}_1(x_1, A^c \times X_2) = 0$ holds for μ-almost every $x_1 \in A$. In other words, for μ-almost every $x_1 \in X_1$, either $x_1 \notin A$ or $\mathcal{Q}_1(x_1, A^c \times X_2) = 0$.

Choose a countable topological basis $\{A_i\}_{i \in \mathbb{N}}$ for X_1. Then for μ-almost every $x_1 \in X_1$, we have $\mathcal{Q}_1(x_1, A_i^c \times X_2) = 0$ for all $i \in \mathbb{N}$ satisfying $x_1 \in A_i$. By condition (ii), the following equality

$$\mathcal{Q}_1(x_1, \{x_1\}^c \times X_2) = \mathcal{Q}_1\Bigl(x_1, \bigcup_{A_i \ni x_1} (A_i^c \times X_2)\Bigr) = 0 \text{ for } \mu\text{-almost every } x_1 \in X_1.$$

holds for μ-almost every $x_1 \in X_1$. Hence for μ-almost every $x_1 \in X_1$, we have

$$\mathcal{Q}_1(x_1, \{x_1\} \times X_2) = 1. \tag{A.18}$$

Since ν is supported on M, by taking $A = X_1$ and $B = M$ in (A.17) we get

$$\int_{X_1} \mathcal{Q}_1(x_1, M) \,\mathrm{d}\mu(x_1) = \nu(M) = 1.$$

This implies that $\mathcal{Q}_1(x_1, M) = 1$ holds for μ-almost every $x_1 \in X_1$. This property together with (A.18) implies that for μ-almost every $x_1 \in X_1$, we have

$$\mathcal{Q}_1(x_1, \{x_1\} \times X_2 \cap M) = 1. \tag{A.19}$$

Suppose that (A.19) holds for $x_1 \in J$, where $J \in \mathscr{B}(X_1)$ and $\mu(J) = 1$. If $x_1 \in J$, then $\{x_1\} \times X_2 \cap M$ must be non-empty, and thus $x_1 \in \kappa(M)$. Consequently, $J \subseteq \kappa(M)$. Fix a point $x_2 \in X_2$ and a Borel measurable map $f \colon \kappa(M) \to X_2$ such that $(x_1, f(x_1)) \in M$ for every $x_1 \in \kappa(M)$. The existence of f is guaranteed by [MA99, Lemma 1.1]. For each $x_1 \in X_1$ and each $A \in \mathscr{B}(X_2)$, define

$$\mathcal{Q}(x_1, A) \coloneqq \begin{cases} \mathcal{Q}_1(x_1, \{x_1\} \times A), & \text{if } x_1 \in J, \\ \delta_{f(x_1)}(A), & \text{if } x_1 \in \kappa(M) \setminus J, \\ \delta_{x_2}(A), & \text{if } x_1 \in X_1 \setminus \kappa(M), \end{cases}$$

where δ_y refers to the Dirac measure on X_2 at a point $y \in X_2$. Fix an arbitrary Borel set $A \in \mathscr{B}(X_2)$. By condition (i), the function that assigns each point $x_1 \in J$ the value $\mathcal{Q}(x_1, A) = \mathcal{Q}_1(x_1, \{x_1\} \times A)$ is Borel measurable. Moreover, the measurability of f ensures that the function that assigns each point $x_1 \in \kappa(M) \setminus J$ the value $\mathcal{Q}(x_1, A) = \delta_{f(x_1)}(A)$ is Borel measurable. Furthermore, the the function that assigns each point $x_1 \in X_1 \setminus \kappa(M)$ the value $\mathcal{Q}(x_1, A) = \delta_{x_2}(A)$ is clearly Borel measurable. Consequently, the function that assigns each point $x_1 \in X_1$ the value $\mathcal{Q}(x_1, A)$ is Borel measurable.

Fix an arbitrary point $x_1 \in X_1$. If $x_1 \in J$, then by (A.19), $\mathcal{Q}(x_1, \cdot)$, the map that assigns each Borel set $A \in \mathscr{B}(X_2)$ the value $\mathcal{Q}(x_1, A) = \mathcal{Q}_1(x_1, \{x_1\} \times A)$ is a probability measure on X_2 supported on $\{x_2 \in X_2 : (x_1, x_2) \in M\}$. If $x_1 \in \kappa(M) \setminus J$, since $(x_1, f(x_1)) \in M$, the map $\mathcal{Q}(x_1, \cdot) = \delta_{f(x_1)}$ is a probability measure on X_2 supported on $\{x_2 \in X_2 : (x_1, x_2) \in M\}$. If $x_1 \in X_1 \setminus \kappa(M)$, the map $\mathcal{Q}(x_1, \cdot) = \delta_{x_2}$ is a probability measure on X_2. Hence, $\mathcal{Q}$ is a transition probability kernel from X_1 to X_2 with the property (a) in Proposition A.11. Now we verify the property (b1) in Remark A.12 for μ and $\mathcal{Q}$, which is equivalent to the property (b) in Proposition A.11.

By (A.18), we have for each $x_1 \in J$ and each $B \in \mathscr{B}(X_2)$,

$$\mathcal{Q}_1(x_1, X_1 \times B) = \mathcal{Q}_1(x_1, X_1 \times B \cap \{x_1\} \times X_2) = \mathcal{Q}_1(x_1, \{x_1\} \times B) = \mathcal{Q}(x_1, B).$$

Fix arbitrary Borel sets $A \in \mathscr{B}(X_1)$ and $B \in \mathscr{B}(X_2)$. Since $\mu(J) = 1$, applying (A.17) we can get

$$\int_A \mathcal{Q}(x_1, B) \, \mathrm{d}\mu(x_1) = \int_A \mathcal{Q}_1(x_1, X_1 \times B) \, \mathrm{d}\mu(x_1) = \nu(X_1 \times B \cap A \times X_2) = \nu(A \times B).$$

Hence we conclude that this μ and $\mathcal{Q}$ satisfy the properties (a) and (b) in Proposition A.11.

Now we check the uniqueness of μ and $\mathcal{Q}$.

Suppose that a transition probability kernel $\mathcal{Q}$ from X_1 to X_2 and a Borel probability measure μ on X_1 satisfy the properties (a) and (b). First, fix an arbitrary $A \in \mathscr{B}(x_1)$. by taking $C = A \times X_2$ in the property (b) we get

$$\nu \circ \kappa^{-1}(A) = \nu(A \times X_2) = \int_A \mathcal{Q}(x_1, X_2) \, \mathrm{d}\mu(x_1) = \int_A \mathrm{d}\mu(x_1) = \mu(A),$$

so $\mu = \nu \circ \kappa^{-1}$. Suppose that a transition probability kernel $\mathcal{Q}'$ from X_1 to X_2 together with μ also satisfies the properties (a) and (b). Then the property (b) implies that $\int_A \mathcal{Q}(x_1, B) \, \mathrm{d}\mu(x_1) = \int_A \mathcal{Q}'(x_1, B) \, \mathrm{d}\mu(x_1)$ holds for arbitrary Borel sets $A \in \mathscr{B}(X_1)$ and $B \in \mathscr{B}(X_2)$, which indicates that $\mathcal{Q}(x_1, B) = \mathcal{Q}'(x_1, B)$ holds for all $B \in \mathscr{B}(X_2)$ μ-almost every $x_1 \in X_1$.

Choose a countable topological basis $\mathscr{B}$ for X_2 and denote by $\mathfrak{T}(\mathscr{B})$ the collection of all finite intersections of elements in $\mathscr{B}$, which is also countable. We can see that $\mathfrak{T}(\mathscr{B})$ is a π-system and that the σ-algebra generated by $\mathfrak{T}(\mathscr{B})$ is $\mathscr{B}(X_2)$. Since the countability of $\mathfrak{T}(\mathscr{B})$, for μ-almost every $x_1 \in X_1$, the equality $\mathcal{Q}(x_1, B) = \mathcal{Q}'(x_1, B)$ holds for all $B \in \mathfrak{T}(\mathscr{B})$, and thereby holds for all $B \in \mathscr{B}(X_2)$ by the Dynkin's π-λ Theorem. $\square$

Let X_1 and X_2 be compact metric spaces, M be a non-empty closed subset of $X_1 \times X_2$, and ν be a Borel probability measure on $X_1 \times X_2$ supported on M. Denote by $\kappa_1 \colon X_1 \times X_2 \to X_1$ and $\kappa_2 \colon X_1 \times X_2 \to X_2$ the projection maps given by $\kappa_1(x_1, x_2) = x_1$ and $\kappa_2(x_1, x_2) = x_2$, respectively, for all $x_1 \in X_1$ and $x_2 \in X_2$. Proposition A.11 ensures the notions defined in the following two definitions always exist.

Definition A.13. If a transition probability kernel $\mathcal{Q}$ from X_1 to X_2 and the Borel probability measure $\mu = \nu \circ \kappa_1^{-1}$ on X_1 satisfy the two properties (a) and (b) in Proposition A.11, then $\mathcal{Q}$ is called a *forward conditional transition probability kernel of ν from X_1 to X_2 supported on M.*

Definition A.14. A transition probability kernel $\mathcal{Q}$ from X_2 to X_1 is called a *backward conditional transition probability kernel of ν from X_2 to X_1 supported on M* if it satisfies the following two properties:

(a) For each $x_2 \in \kappa_2(M)$, we have
$$\mathcal{Q}(x_2, \{x_1 \in X_1 : (x_1, x_2) \in M\}) = 1.$$
(b) For each $C \in \mathscr{B}(X_1 \times X_2)$, we have
$$\nu(C) = \int_{X_2} \mathcal{Q}(x_2, \{x_1 \in X_2 : (x_1, x_2) \in C\}) \,\mathrm{d}\big(\nu \circ \kappa_2^{-1}\big)(x_2).$$

Remark A.15. If $X_1 = X_2$, then by (A.11) in the case where ϕ is a characteristic function of a measurable subset of X^2, we can see that property (b) in Proposition A.11 is equivalent to $\nu = \mu \mathcal{Q}^{[1]}$. Similarly, (b) in Definition A.14 is equivalent to $\nu \circ \iota_2^{-1} = (\nu \circ \kappa_2^{-1}) \mathcal{Q}^{[1]}$, where $\iota_2(x, y) = (y, x)$ for all $(x, y) \in X^2$.

Appendix B. Single-valued maps

In this appendix, we focus on a degenerate case when the correspondence is induced by a single-valued map, and show that our theory is compatible with the classical ergodic theory for single-valued maps. In particular, we will explain why the conjectured (1.1) naturally arises and coincides with the Variational Principle for single-valued maps when the correspondence is induced by a single-valued map.

B.1. **Transition probability kernels and measure-theoretic entropy.** Let $(X, \mathscr{B}(X))$ and $(Y, \mathscr{B}(y))$ be measurable spaces and $F \colon Y \to X$ be a measurable map.

Definition B.1. Let $F \colon Y \to X$ be a measurable map. The transition probability kernel $\widehat{F}$ induced by F is defined as
$$\widehat{F}(y, A) := \mathbb{1}_{F^{-1}(A)}(y) = \begin{cases} 1, & \text{if } F(y) \in A, \\ 0, & \text{if } F(y) \notin A \end{cases}$$
for all $y \in Y$ and $A \in \mathscr{B}(X)$.

Remark. In this case, we have $\widehat{F}_y = \delta_{F(y)}$ for each $y \in Y$, where $\delta_{F(y)}$ refers to the Dirac measure on X at the point $F(y)$.

Let $F\colon Y \to X$ be a measurable map and $f\colon X \to \mathbb{R}$ be a measurable function. For each $y \in Y$, we have $\widehat{F}f(y) = \int_X f(x)\,\mathrm{d}\widehat{F}_y(x) = \int_X f(x)\,\mathrm{d}\delta_{F(y)}(x) = f(F(y))$, so $\widehat{F}f = f \circ F$.

Suppose that μ is a probability measure on $(Y, \mathscr{B}(Y))$. For each $A \in \mathscr{B}(X)$, we have $\bigl(\mu\widehat{F}\bigr)(A) = \int_Y \widehat{F}(y, A)\,\mathrm{d}\mu(y) = \int_Y \mathbb{1}_{F^{-1}(A)}(y)\,\mathrm{d}\mu(y) = \mu(F^{-1}(A))$, so

$$\mu\widehat{F} = \mu \circ F^{-1}. \tag{B.1}$$

This leads to the following lemma.

Lemma B.2. *Let $F\colon X \to X$ be a measurable map on a measurable space $(X, \mathscr{B}(X))$. A probability measure on X is $\widehat{F}$-invariant if and only if it is F-invariant.*

Let $(X, \mathscr{B}(X))$, $(Y, \mathscr{B}(Y))$, and $(Z, \mathscr{B}(Z))$ be measurable spaces and $F_1\colon Y \to X$ and $F_2\colon Z \to Y$ be measurable maps. From Definition B.1 and (B.1), we have for each $z \in Z$ and $A \in \mathscr{B}(X)$, $\bigl(\widehat{F}_2\widehat{F}_1\bigr)(z, A) = \bigl(\delta_{F_2(z)}\widehat{F}_1\bigr)(A) = \delta_{F_2(z)}(F_1^{-1}(A)) = \mathbb{1}_{F_2^{-1}\circ F_1^{-1}(A)}(z) = \widehat{F_1 \circ F_2}(z, A)$, so $\widehat{F}_2\widehat{F}_1 = \widehat{F_1 \circ F_2}$.

Let $(X, \mathscr{B}(X))$ be a measurable space and $F\colon X \to X$ be a measurable map. Then for each $n \in \mathbb{N}_0$ and arbitrary measurable sets $B_0, B_1, \dots, B_n \in \mathscr{B}(X)$, we have

$$\begin{aligned}\widehat{F}^{[n]}(x, B_0 \times B_1 \times \cdots \times B_n) &= \delta_{(x, F(x), \dots, F^n(x))}(B_0 \times B_1 \times \cdots \times B_n)\\ &= \mathbb{1}_{B_0 \cap F^{-1}(B_1) \cap \cdots \cap F^{-n}(B_n)}(x),\end{aligned}$$

which can be verified by induction on n based on Definition 5.14. This property and Definition 5.6 imply that for an arbitrary probability measure μ on X, we have

$$\bigl(\mu\widehat{F}^{[n]}\bigr)(B_0 \times B_1 \times \cdots \times B_n) = \mu(B_0 \cap F^{-1}(B_1) \cap \cdots \cap F^{-n}(B_n)). \tag{B.2}$$

Let μ be an F-invariant probability measure on X. By Lemma B.2, the mearure μ is also $\widehat{F}$-invatiant.

We recall some conventions from [PU10, Chapter 2]:

Let $\mathcal{A}$ be a finite measurable partition of X and $n \in \mathbb{N}$. The finite measurable partition $F^{-n}(\mathcal{A})$ is given by

$$F^{-1}(\mathcal{A}) \coloneqq \bigl\{F^{-1}(A) : A \in \mathcal{A}\bigr\}, \text{ and } F^{-n}(\mathcal{A}) \coloneqq F^{-1}\bigl(F^{-(n-1)}(\mathcal{A})\bigr).$$

The entropy $h_\mu(F, \mathcal{A})$ is given by

$$h_\mu(F, \mathcal{A}) \coloneqq \lim_{n \to +\infty} \frac{1}{n} H_\mu\bigl(\mathcal{A} \vee F^{-1}(\mathcal{A}) \vee \cdots \vee F^{-(n-1)}(\mathcal{A})\bigr). \tag{B.3}$$

By (B.2), we have $H_{\mu\widehat{F}^{[n-1]}}(\mathcal{A}^n) = H_\mu\bigl(\mathcal{A} \vee F^{-1}(\mathcal{A}) \vee \cdots \vee F^{-(n-1)}(\mathcal{A})\bigr)$. Hence

$$h_\mu\bigl(\widehat{F}, \mathcal{A}\bigr) = h_\mu(F, \mathcal{A}). \tag{B.4}$$

Recall $h_\mu(F) \coloneqq \sup_{\mathcal{A}} h_\mu(F, \mathcal{A})$ from [PU10, Chapter 2], where $\mathcal{A}$ ranges over all finite measurable partitions of X. By (B.4), we have

$$h_\mu(F) = h_\mu\bigl(\widehat{F}\bigr). \tag{B.5}$$

B.2. **Correspondences and topological pressure.** For a continuous map $f\colon X \to X$ on a compact metric space (X,d), denote by $\mathcal{C}_f\colon X \to \mathcal{F}(X)$ the map that assigns each point $x \in X$ the closed subset $\{f(x)\}$ of X. It turns out that $\mathcal{C}_f$ is a correspondence on X. Let us compute its topological pressure.

Let $\varphi\colon X \to \mathbb{R}$ be a continuous function. We recall the definition of the topological pressure $P(f,\varphi)$ from [PU10, Section 3.3]:

For each $n \in \mathbb{N}$, we say that a subset $E \subseteq X$ is (n,ϵ)-separated in (X,d) if the set $\{(x, f(x), \dots, f^{n-1}(x)) : x \in E\}$ is ϵ-separated in $(\mathcal{O}_n(\mathcal{C}_f), d_n)$; and we say that a subset $F \subseteq X$ is (n,ϵ)-spanning in (X,d) if the set $\{(x, f(x), \dots, f^{n-1}(x)) : x \in F\}$ is ϵ-spanning in $(\mathcal{O}_n(\mathcal{C}_f), d_n)$. The topological pressure $P(f,\varphi)$ is given by

$$\begin{aligned} P(f,\varphi) &\coloneqq \lim_{\epsilon\to 0^+} \limsup_{n\to+\infty} \frac{1}{n} \log\Biggl(\sup_{E_n(\epsilon)} \sum_{x\in E_n(\epsilon)} \exp\Biggl(\sum_{j=0}^{n-1} \varphi(f^j(x)) \Biggr)\Biggr) \\ &= \lim_{\epsilon\to 0^+} \limsup_{n\to+\infty} \frac{1}{n} \log\Biggl(\inf_{F_n(\epsilon)} \sum_{x\in F_n(\epsilon)} \exp\Biggl(\sum_{j=0}^{n-1} \varphi(f^j(x)) \Biggr)\Biggr), \end{aligned} \tag{B.6}$$

where $E_n(\epsilon)$ ranges over all (n,ϵ)-separated subsets in X and $F_n(\epsilon)$ ranges over all (n,ϵ)-spanning subsets in X.

We will show that $P(f,\varphi)$ and $P(\mathcal{C}_f, \widehat{\varphi})$ are equal, where $\widehat{\varphi}\colon \mathcal{O}_2(\mathcal{C}_f) \to \mathbb{R}$ is a function induced by φ (see (2.4) for its precise definition), and thus the topological pressure of correspondences generalizes the topological pressure of single-valued continuous maps.

Proposition B.3. *Let $f\colon X \to X$ be a continuous transformation on a compact metric space (X,d) and $\varphi\colon X \to \mathbb{R}$ be a continuous function. Then $P(f,\phi) = P(\mathcal{C}_f, \widehat{\varphi})$.*

Proof. Fix an arbitrary $n \in \mathbb{N}$. Since for each $x \in X$, $\mathcal{C}_f(x) = \{f(x)\}$ is a singleton, we can see that $(x_1, x_2, \dots, x_{n+1}) \in \mathcal{O}_{n+1}(\mathcal{C}_f)$ depends on x_1 in the way that $x_i = f^{i-1}(x_1)$ for $i \in \{2, \dots, n+1\}$. Thus the map Φ_{n+1} that assigns each point $x \in X$ the orbit $(x, f(x), \dots, f^n(x)) \in \mathcal{O}_{n+1}(\mathcal{C}_f)$ is a bijection from X to $\mathcal{O}_{n+1}(\mathcal{C}_f)$. Recall that a subset $E \subseteq X$ is $(n+1,\epsilon)$-separated in (X,d) if and only if $\Phi_{n+1}(E) = \{(x, f(x), \dots, f^n(x)) : x \in E\}$ is ϵ-separated in $(\mathcal{O}_{n+1}(\mathcal{C}_f), d_{n+1})$, so

$$\begin{aligned} &\sup_{E_{n+1}(\epsilon)} \sum_{x\in E_{n+1}(\epsilon)} \exp\Biggl(\sum_{j=0}^{n} \varphi(f^j(x)) \Biggr) \\ &\quad= \sup_{E_{n+1}(\epsilon)} \sum_{(x_1,\dots x_{n+1})\in\Phi_{n+1}(E_{n+1}(\epsilon))} \exp\Biggl(\sum_{j=1}^{n} \widehat{\varphi}(x_j, x_{j+1}) + \varphi(x_{n+1}) \Biggr) \\ &\quad= \sup_{E} \sum_{\underline{x}=(x_1,\dots,x_{n+1})\in E} \exp(S_n\widehat{\varphi}(\underline{x}) + \varphi(x_{n+1})), \end{aligned}$$

where $E_{n+1}(\epsilon)$ ranges over all $(n+1,\epsilon)$-separated subset of X and E ranges over all ϵ-separated subset of $\mathcal{O}_n(\mathcal{C}_f)$. Since $s_n(\mathcal{C}_f,\widehat{\varphi},\epsilon)=\sup_E\sum_{\underline{x}\in E}\exp(S_n\widehat{\varphi}(\underline{x}))$, we have

$$e^{-\|\varphi\|_\infty}s_n(\mathcal{C}_f,\widehat{\varphi},\epsilon)\leqslant\sup_{E_{n+1}(\epsilon)}\sum_{x\in E_{n+1}(\epsilon)}\exp\biggl(\sum_{j=0}^{n}\varphi(f^j(x))\biggr)\leqslant e^{\|\varphi\|_\infty}s_n(\mathcal{C}_f,\widehat{\varphi},\epsilon).$$

Therefore by (B.6) we have $P(f,\phi)=\lim_{\epsilon\to0^+}\limsup_{n\to+\infty}\frac{1}{n}\log(s_n(\mathcal{C}_f,\phi,\epsilon))=\lim_{\epsilon\to0^+}s(\mathcal{C}_f,\phi,\epsilon)=P(\mathcal{C}_f,\phi)$. □

The only transition probability kernel supported by $\mathcal{C}_f$ is $\widehat{f}$, defined in Definition B.1, and what we shall consider is the Borel probability measure μ which is $\widehat{f}$-invariant, or equivalently, f-invariant, where the equivalence has been shown in Lemma B.2.

By applying Variational Principle to φ in the dynamical system (X,f), we have

$$P(f,\varphi)=\sup\biggl\{h_\mu(f)+\int_X\varphi\,\mathrm{d}\mu:\mu\text{ is }f\text{-invariant}\biggr\}.\tag{B.7}$$

Recall $\mathcal{C}_f(x_1)=\{f(x_1)\}$ and $\widehat{f}_{x_1}=\delta_{f(x_1)}$ for all $x_1\in X$. By (2.4) we have

$$\int_X\int_{\mathcal{C}_f(x_1)}\widehat{\varphi}(x_1,x_2)\,\mathrm{d}\widehat{f}_{x_1}(x_2)\,\mathrm{d}\mu(x_2)=\int_X\int_{\{f(x_1)\}}\varphi(x_1)\,\mathrm{d}\delta_{f(x_1)}(x_2)\,\mathrm{d}\mu(x_2)=\int_X\varphi\,\mathrm{d}\mu.\tag{B.8}$$

By (B.7), Proposition B.3, (B.5) and (B.8), we get

$$P(\mathcal{C}_f,\widehat{\varphi})=\sup\biggl\{h_\mu(\widehat{f})+\int_X\int_{\mathcal{C}_f(x_1)}\widehat{\varphi}(x_1,x_2)\,\mathrm{d}\widehat{f}_{x_1}(x_2)\,\mathrm{d}\mu(x_2):\mu\text{ is }\widehat{f}\text{-invariant}\biggr\}.$$

Therefore the Variational Principle holds when the correspondence is induced by a single-valued continuous map.

B.3. **Several properties for $\mathcal{C}_f$.** Let $f\colon X\to X$ be a single-valued continuous map on a compact metric space (X,d). Recall that $\mathcal{C}_f\colon X\to\mathcal{F}(X)$ is the correspondence on X that assigns each point $x\in X$ the closed subset $\{f(x)\}$ of X. In this subsection, we point out some relations between properties for the correspondence $\mathcal{C}_f$ and properties of the single-valued map f, all of which are not difficult to check by their definitions.

(i) The correspondence $\mathcal{C}_f$ is forward expansive with an expansive constant $\epsilon>0$ if and only if the single-valued map f is forward expansive with an expansive constant ϵ, i.e., for each pair of different points $x_1,\,x_2\in X$, there exists $n\in\mathbb{N}_0$ such that $d(f^n(x_1),f^n(x_2))>\epsilon$.

(ii) The correspondence $\mathcal{C}_f$ has the specification property in the sense of Definition 7.1 if and only if the single-valued map f has the specification property in the sense of Definition 7.2.

(iii) The correspondence $\mathcal{C}_f$ is distance-expanding in the sense of Definition 7.5 if and only if the single-valued map f is distance-expanding.

(iv) The correspondence $\mathcal{C}_f$ is open in the sense of Definition 7.11 if and only if the single-valued map f is open.

(v) If the correspondence $\mathcal{C}_f$ is strongly transitive in the sense of Definition 7.12, then the single-valued map f is topologically transitive.

(vi) The correspondence $\mathcal{C}_f$ is topologically exact in the sense of Definition 7.13 if and only if the single-valued map f is topologically exact.

Appendix C. Finite cases

In this section, we focus on another case when X is a finite set, and examine our notions and theorems.

C.1. **Transition probability kernels and measure-theoretic entropy.**

Definition C.1. Let $d \in \mathbb{N}$, $X = Y = (d]$, $\mathscr{B}(X) = \mathscr{B}(Y) = 2^X$, the set of all subsets of X, and $P = (p_{ij})_{1 \leqslant i,j \leqslant d}$ be a matrix satisfying $p_{ij} \geqslant 0$ for all $1 \leqslant i, j \leqslant d$ and $\sum_{j=1}^{d} p_{ij} = 1$ for all $1 \leqslant i \leqslant d$. The transition probability kernel $\widehat{P}$ induced by P is defined as

$$\widehat{P}(i, A) \coloneqq \sum_{j \in A} p_{ij}$$

for all $i \in (d]$ and $A \subseteq (d]$. In particular, for arbitrary i, $j \in X$, we have $\widehat{P}(i, \{j\}) = p_{ij}$.

Remark. In this case, we can consider the matrix P as the transition matrix of a Markov chain with the state space $X = (d]$. For each $i \in (d]$, the measure $\widehat{P}_i$ can be represented by the probability vector $(p_{i1}, \dots, p_{id})$.

Let $d \in \mathbb{N}$ and $P = (p_{ij})_{1 \leqslant i,j \leqslant d}$ be the transition matrix of a Markov chain with state space $X = (d]$. We use a column vector $v_f = (f(1), f(2), \dots, f(d))^T$ to denote a function $f \colon X \to \mathbb{R}$. Additionally, for a distribution p on X, we write $p = (p_1, p_2, \dots, p_d)$, where $p_j = p(\{j\})$ for each $j \in X$.

For a function $f \colon X \to \mathbb{R}$, we have $\widehat{P}f(i) = \int_X f(j) \,\mathrm{d}\widehat{P}_i(j) = \sum_{j=1}^{d} f(j) p_{ij}$, and thus $v_{\widehat{P}f} = P v_f$. Let $p = (p_1, p_2, \dots, p_d)$ be a distribution on X. For each $i \in X$, we have $p\widehat{P}(\{i\}) = \sum_{j=1}^{d} \widehat{P}(j, \{i\}) p(\{j\}) = \sum_{j=1}^{d} p_j p_{ji}$, so

$$p\widehat{P} = pP. \tag{C.1}$$

This leads to the following lemma.

Lemma C.2. *For a d-state Markov chain with transition matrix P, a probability distribution vector p on the state space is $\widehat{P}$-invariant if and only if $pP = p$.*

Let $d \in \mathbb{N}$, $X = (d]$, and $P_1 = \bigl(p^{(1)}_{ij}\bigr)_{1\leqslant i,j\leqslant d}$, $P_2 = \bigl(p^{(2)}_{ij}\bigr)_{1\leqslant i,j\leqslant d}$ be transition matrices on X. From Definition C.1 and (C.1), for each i, $j \in X$, we have $\bigl(\widehat{P}_2\widehat{P}_1\bigr)(i,\{j\}) = \bigl(\bigl(\widehat{P}_2\bigr)_i\widehat{P}_1\bigr)(\{j\}) = \sum_{k=1}^{d} p^{(2)}_{ik}p^{(1)}_{kj} = \widehat{P_2P_1}(i,\{j\})$, so $\widehat{P}_2\widehat{P}_1 = \widehat{P_2P_1}$.

Let $d \in \mathbb{N}$ and $P = (p_{ij})_{1\leqslant i,j\leqslant d}$ be the transition matrix of a Markov chain with the state space $X = (d]$. Then for arbitrary $n \in \mathbb{N}_0$ and i, j_0, j_1, $\dots$, $j_n \in X$, we have

$$\widehat{P}^{[n]}(i,\{(j_0,j_1,\dots,j_n)\}) = \delta_{ij_0}p_{j_0j_1}p_{j_1j_2}\dots p_{j_{n-1}j_n},$$

which can be verified by induction on n based on Definition 5.14. This property and Definition 5.6 imply that for an arbitrary probability measure μ on X, we have

$$\mu\widehat{P}^{[n]}(\{j_0,\, j_1,\, \dots,\, j_n\}) = \mu(\{j_0\})p_{j_0j_1}p_{j_1j_2}\dots p_{j_{n-1}j_n}. \tag{C.2}$$

Let μ be a $\widehat{P}$-invariant Borel probability measure on X. Let $p = (\mu(\{1\}),\dots,\mu(\{d\}))$ be the distribution vector associated with μ. Lemma C.2 ensures that $pP = p$, so (C.2) and (5.7) reveal that the measure-preserving system $\bigl(X^\omega,\mathscr{B}(X^\omega),\mu\widehat{P}^\omega,\sigma\bigr)$ is a one-sided (p,P)-Markov shift. Write $p_i = \mu(\{i\})$ for each $i \in (d]$.

The following result is shown in [Wa82, Theorem 4.27]:

$$h_{\mu\widehat{P}^\omega}(\sigma) = -\sum_{i,j=1}^{d} p_ip_{ij}\log(p_{ij}),$$

where we follow the convention that $0\log 0 = 0$. Therefore by Theorem 5.24, we have

$$h_\mu\bigl(\widehat{P}\bigr) = -\sum_{i,j=1}^{d} p_ip_{ij}\log(p_{ij}). \tag{C.3}$$

C.2. **Correspondences and topological pressure.** Let $d \in \mathbb{N}$, $X = (d]$ be a finite space with the discrete topology, and $A = (a_{ij})_{1\leqslant i,j\leqslant d}$ be a 0-1 matrix with at least one entry 1 in each row. Denote by $\mathcal{C}_A\colon X \to \mathcal{F}(X)$ the map that assigns each point $i \in X$ the subset $\{j \in X : a_{ij} = 1\}$ of X. Since there is at least one entry 1 in each row of A, we have $\mathcal{C}_A(i) \neq \emptyset$ for each $i \in X$, which ensures $\mathcal{C}_A(i) \in \mathcal{F}(X)$. Note that both X and $\mathcal{F}(X)$ are endowed with the discrete topology, so $\mathcal{C}_A$ is a correspondence on X.

This appendix is devoted to computing the topological pressure of $\mathcal{C}_A$.

Let $\phi\colon \mathcal{O}_2(\mathcal{C}_A) \to \mathbb{R}$ be a function and $A_\phi := (a_{ij}\cdot e^{\phi(i,j)})_{1\leqslant i,j\leqslant d}$ be a $d\times d$ matrix (if $(i,j) \notin \mathcal{O}_2(\mathcal{C}_A)$, then $a_{ij} = 0$, so in this case we do not need to define $\phi(i,j)$). Let $n \in \mathbb{N}$. By definition of the metric d_{n+1}, the only ϵ-spanning subset of $(\mathcal{O}_{n+1}(\mathcal{C}_A),d_{n+1})$

is $\mathcal{O}_{n+1}(\mathcal{C}_A)$ for $\epsilon > 0$ small enough. As a result, by (4.4) we get

$$\begin{aligned} P(\mathcal{C}_A, \phi) &= \limsup_{n\to\infty} \frac{1}{n} \log\Biggl(\sum_{\underline{x}\in\mathcal{O}_{n+1}(\mathcal{C}_A)} \exp(S_n\phi(\underline{x})) \Biggr) \\ &= \limsup_{n\to+\infty} \frac{1}{n} \log\Biggl(\sum_{(i_1,\dots,i_{n+1})\in X^{n+1}} \Biggl(\prod_{j=1}^{n} a_{i_j i_{j+1}} \Biggr) \cdot \exp\Biggl(\sum_{j=1}^{n} \phi(i_j, i_{j+1}) \Biggr) \Biggr) \\ &= \limsup_{n\to+\infty} \frac{1}{n} \log\Biggl(\sum_{i_1,\dots,i_{n+1}=1}^{d} \prod_{j=1}^{n-1} \Bigl(a_{i_j i_{j+1}} \cdot e^{\phi(i_j, i_{j+1})} \Bigr) \Biggr) \\ &= \limsup_{n\to+\infty} \frac{1}{n} \log\bigl(\|A_\phi^n\|_1 \bigr), \end{aligned}$$

where the norm $\|\cdot\|_1$ is given by $\|B\|_1 := \sum_{i,j=1}^{d} |b_{ij}|$ for every $d\times d$ matrix $B = (b_{ij})_{1\leqslant i,j\leqslant d}$.

By the Gelfand's formula, we have

$$P(\mathcal{C}_A, \phi) = \limsup_{n\to+\infty} \frac{1}{n} \log\bigl(\bigl\|A_\phi^n\bigr\|_1 \bigr) = \log\Bigl(\limsup_{n\to+\infty} \bigl\|A_\phi^n\bigr\|_1^{\frac{1}{n}} \Bigr) = \log(\rho(A_\phi)), \tag{C.4}$$

where $\rho(A_\phi)$ is the spectral radius of A_ϕ. Moreover, by (4.5), we get

$$\rho(A_\phi) = \exp(P(\mathcal{C}_A, \phi)) \geqslant \exp(-\|\phi\|_\infty). \tag{C.5}$$

Notice that $(\mathcal{O}_\omega(\mathcal{C}_A), \sigma)$ is the one-sided subshift of finite type defined by A. Theorem 4.9 and (C.4) imply that the topological pressure of the one-sided subshift of finite type $(\mathcal{O}_\omega(\mathcal{C}_A), \sigma)$ defined by A with respect to the potential $\widetilde{\phi}$ is $\log(\rho(A_\phi))$. Taking $\phi \equiv 0$ we get the topological entropy of $(\mathcal{O}_\omega(\mathcal{C}_A), \sigma)$ is $\log(\rho(A))$, which has been proved by W. Parry in [Pa64, Theorem 7].

C.3. **Construction of an equilibrium state.** By the discreteness of the finite space, all correspondences on $X = (d]$ are forward expansive, so by Theorem B, the Variational Principle always holds and equilibrium states always exist in this case. This appendix is devoted to constructing an equilibrium state explicitly. The equilibrium state may be not unique because the 0-1 matrix A may be not irreducible, so we do not discuss thermodynamic formalism beyond the existence of equilibrium states here.

Let $d \in \mathbb{N}$, $X = (d]$ be a compact metric space with the discrete topology, $A = (a_{ij})_{1\leqslant i,j\leqslant d}$ be a 0-1 matrix with at least one entry 1 in each row, and $\phi\colon \mathcal{O}_2(\mathcal{C}_A) \to \mathbb{R}$ be a function. We focus on the correspondence $\mathcal{C}_A$. Recall from (C.4) that the topological for ϕ is $\log(\rho(A_\phi))$. We have shown $\rho(A_\phi) > 0$ in (C.5).

A transition probability kernel on X supported by $\mathcal{C}_A$ is of the form $\widehat{P}$, where $P = (p_{ij})_{1\leqslant i,j\leqslant d}$ is a transition matrix satisfying $p_{ij} = 0$ for all $i, j \in (d]$ with $a_{ij} = 0$. Let μ be a $\widehat{P}$-invariant Borel probability measure on X with the associated distribution vector $p = (p_1, \dots, p_d)$, where $p_i = \mu(\{i\})$ for each $i \in (d]$. Recall

$h_\mu\bigl(\widehat{P}\bigr) = -\sum_{i,j=1}^{d} p_i p_{ij} \log(p_{ij})$ from (C.3). Thereby, that the pair $\bigl(\mu, \widehat{P}\bigr)$ is an equilibrium state for $\mathcal{C}_A$ and ϕ is equivalent to the following equality:

$$\log(\rho(A_\phi)) = -\sum_{i,j=1}^{d} p_i p_{ij} \log(p_{ij}) + \sum_{i,j=1}^{d} p_i p_{ij} \phi(i,j). \tag{C.6}$$

Now we construct P and p such that (C.6) holds. In [Ki98, Section 6.2], there is a construction for irreducible A and $\phi \equiv 0$.

By the Perron–Frobenius Theorem, $\lambda \coloneqq \rho(A_\phi) > 0$ is an eigenvalue of A_ϕ, and we can choose an associated non-zero right eigenvector $q = (q_1, \dots, q_d)^T$ with $q_i \geqslant 0$ for all $i \in (d]$ such that $A_\phi q = \lambda q$. Write $L \coloneqq \{i \in (d] : q_i \neq 0\}$. Since q is non-zero, we have $L \neq \emptyset$. Define $P = (p_{ij})_{1 \leqslant i,j \leqslant d}$ as follows:

$$p_{ij} \coloneqq \begin{cases} \frac{q_j}{\lambda q_i} a_{ij} e^{\phi(i,j)}, & \text{if } i \in L, \\ \frac{a_{ij}}{\#\mathcal{C}_A(i)}, & \text{otherwise} \end{cases}$$

for all i, $j \in (d]$, where $\#\mathcal{C}_A(i)$ denotes the cardinality of $\mathcal{C}_A(i)$.

Clearly, for all i, $j \in (d]$, we have $p_{ij} \geqslant 0$, and, moreover, if $a_{ij} = 0$, then $p_{ij} = 0$. To show that P is a transition matrix, we rewrite the equality $A_\phi q = \lambda q$ as $\sum_{j=1}^{n} q_j a_{ij} e^{\phi(i,j)} = \lambda q_i$ for all $i \in (d]$, which implies $\sum_{j=1}^{d} p_{ij} = \sum_{j=1}^{d} \frac{q_j}{\lambda q_i} a_{ij} e^{\phi(i,j)} = 1$ for all $i \in L$. Moreover, for each $i \in (d] \setminus L$, we have $\sum_{j=1}^{d} p_{ij} = \sum_{j=1}^{d} \frac{a_{ij}}{\#\mathcal{C}_A(i)} = \frac{\#\{j \in (d] : a_{ij} = 1\}}{\#\{j \in (d] : a_{ij} = 1\}} = 1$. Hence, P is a transition matrix.

Note that for each $i \in L$ and each $j \in (d] \setminus L$, we have $p_{ij} = \frac{q_j}{\lambda q_i} a_{ij} e^{\phi(i,j)} = 0$, so for each $i \in L$, we have

$$\sum_{j \in L} p_{ij} = \sum_{j=1}^{d} p_{ij} = 1. \tag{C.7}$$

This equality implies that the submatrix $P_L \coloneqq (p_{ij})_{i,j \in L}$ is a $l \times l$ transition matrix, whose spectral radius is 1, where $l \coloneqq \#L$. By the Perron–Frobenius Theorem, we can choose a distribution vector $p_L = (p_i)_{i \in L}$ such that $p_L P_L = p_L$, i.e., for each $j \in L$,

$$\sum_{i \in L} p_i p_{ij} = p_j. \tag{C.8}$$

We set $p_i = 0$ for all $i \in (d] \setminus L$ and get a distribution vector $p = (p_1, \dots, p_d)$. Now we check $pP = p$. Firstly, for each $j \in L$, we have $\sum\limits_{i=1}^{d} p_i p_{ij} = \sum_{i \in L} p_i p_{ij} = p_j$. Secondly, since $p_{ij} = 0$ for all $i \in L$ and $j \in (d] \setminus L$, we have $\sum_{i=1}^{d} p_i p_{ij} = \sum_{i \in L} p_i \cdot 0 + \sum_{i \notin L} 0 \cdot p_{ij} = 0 = p_j$ for all $j \in (d] \setminus L$. Hence, $pP = p$, and thus μ is $\widehat{P}$-invariant.

Now we verify (C.6) for $\bigl(\mu, \widehat{P}\bigr)$.

$$-\sum_{i,j=1}^{d} p_i p_{ij} \log(p_{ij}) = -\sum_{i \in L} \sum_{j=1}^{d} p_i p_{ij} \log(p_{ij}) \qquad \text{(since } p_i = 0 \text{ for each } i \notin L\text{)}$$

$$
\begin{aligned}
&= -\sum_{i,j\in L} p_i p_{ij} \log(p_{ij}) && \text{(since } p_{ij}=0 \text{ for all } i\in L,\ j\notin L\text{)}\\
&= -\sum_{i,j\in L} p_i p_{ij} \log \frac{q_j a_{ij} e^{\phi(i,j)}}{\lambda q_i}\\
&= -\sum_{i,j\in L} p_i p_{ij} \log \frac{q_j e^{\phi(i,j)}}{\lambda q_i}. && \text{(if } a_{ij}\neq 1 \text{ then } p_{ij}=0\text{)}
\end{aligned}
$$

Thus it follows from (C.8) and (C.7) that

$$
\begin{aligned}
&-\sum_{i,j=1}^{d} p_i p_{ij} \log(p_{ij})\\
&\quad= -\sum_{i,j\in L} p_i p_{ij} (\log(q_j) + \phi(i,j) - \log\lambda - \log(q_i))\\
&\quad= \sum_{i\in L}\Big(\sum_{j\in L} p_{ij}\Big) p_i(\log(q_i)+\log\lambda) - \sum_{j\in L}\log(q_j)\sum_{i\in L} p_i p_{ij} - \sum_{i,j\in L} p_i p_{ij}\phi(i,j)\\
&\quad= \sum_{i\in L} p_i(\log(q_i)+\log\lambda) - \sum_{j\in L} p_j\log(q_j) - \sum_{i,j\in L} p_i p_{ij}\phi(i,j)\\
&\quad= \log\lambda - \sum_{i,j=1}^{d} p_i p_{ij}\phi(i,j).
\end{aligned}
$$

Therefore (C.6) holds and $\big(\mu, \widehat{P}\big)$ is an equilibrium state for $\mathcal{C}_A$ and ϕ.

Appendix D. Correspondences and corresponding shift maps

We assume that T is a correspondence on a compact metric space (X,d) and $\sigma\colon \mathcal{O}_\omega(T)\to\mathcal{O}_\omega(T)$ is the corresponding shift map on the orbit space $\mathcal{O}_\omega(T)$ throughout this appendix. This appendix is devoted to establishing some relations between some properties of T and the corresponding properties of σ. These relations are mainly used in Section 7.

Forward expansiveness.

Proposition D.1. *If T is forward expansive with an expansive constant ϵ, then σ is forward expansive with an expansive constant $\frac{\epsilon}{2(1+\epsilon)}$.*

Proof. For each pair of distinct orbits $(x_1,x_2,\dots)$, $(y_1,y_2,\dots)\in\mathcal{O}_\omega(T)$, by the forward expansiveness of T, we can choose a positive integer n such that $d(x_n,y_n)>\epsilon$. This implies that

$$
\begin{aligned}
&d_\omega\big(\sigma^{n-1}(x_1,\dots,x_n,\dots),\sigma^{n-1}(y_1,\dots,y_n,\dots)\big)\\
&\qquad= d_\omega((x_n,x_{n+1},\dots),(y_n,y_{n+1},\dots)) \geqslant \frac{d(x_n,y_n)}{2(1+d(x_n,y_n))} > \frac{\epsilon}{2(1+\epsilon)}.
\end{aligned}
$$

Therefore the shift map σ is forward expansive with an expansive constant $\frac{\epsilon}{2(1+\epsilon)}$. $\square$

Proposition D.2. *If $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ is forward expansive with an expansive constant $\epsilon < 1$, then T is forward expansive with an expansive constant $\frac{\epsilon}{1-\epsilon}$.*

Proof. We argue by contradiction and assume that there is a pair of distinct orbits $(x_1, x_2, \dots)$, $(y_1, y_2, \dots) \in \mathcal{O}_\omega(T)$ such that $d(x_k, y_k) \leqslant \frac{\epsilon}{1-\epsilon}$ for all $k \in \mathbb{N}$. Then for all $n \in \mathbb{N}$, we have

$$\begin{aligned}
&d_\omega\bigl(\sigma^{n-1}(x_1, \dots, x_n, \dots), \sigma^{n-1}(y_1, \dots, y_n, \dots)\bigr)\\
&\qquad = d_\omega((x_n, x_{n+1}, \dots), (y_n, y_{n+1}, \dots))\\
&= \sum_{k=0}^{+\infty} \frac{d(x_{k+n}, y_{k+n})}{2^{k+1}(1 + d(x_{k+n}, y_{k+n}))}\\
&\leqslant \sum_{k=0}^{+\infty} \frac{\epsilon}{2^{k+1}}\\
&= \epsilon,
\end{aligned}$$

which contradicts the fact that $\sigma\colon \mathcal{O}_\omega(T) \to \mathcal{O}_\omega(T)$ is forward expansive with an expansive constant ϵ. Therefore, T is forward expansive with an expansive constant $\frac{\epsilon}{1-\epsilon}$. $\square$

Specification property.

Proposition D.3. *If T has the specification property in the sense of Definition 7.1, then σ has the specification property in the sense of Definition 7.2.*

Proof. Fix an arbitrary number $\epsilon > 0$. Choose $K \in \mathbb{N}$ such that $1/2^K < \epsilon/2$.

By the specification property of T, suppose that $M \in \mathbb{N}$ satisfies the following property:

For arbitrary $n \in \mathbb{N}$, $x_0^1, \dots, x_0^n \in X$, $m_1, \dots, m_n$, $p_1, \dots, p_n$ with $p_j > M$ for every $j \in (n]$, and orbits $\bigl(x_0^j, x_1^j, \dots, x_{m_j-1}^j\bigr) \in \mathcal{O}_{m_j}(T)$ for every $j \in (n]$, there exists an orbit $z = (z_0, z_1, \dots) \in \mathcal{O}_\omega(T)$ such that $d\bigl(z_{m(j-1)+i}, x_i^j\bigr) < \epsilon/2$ for all $j \in (n]$ and $i \in [m_j - 1]$, where $m(j) \coloneqq \sum_{k=1}^{j} (m_k + p_k)$.

Now fix arbitrary $n \in \mathbb{N}$, orbits $x^j = \bigl(x_0^j, x_1^j, \dots\bigr) \in \mathcal{O}_\omega(T)$, $j \in (n]$, and $m_1, \dots, m_n$, $p_1, \dots, p_n \in \mathbb{N}$ with $p_j > M + K$. Since $p_j - K > M$, we can choose an orbit $z = (z_0, z_1, \dots) \in \mathcal{O}_\omega(T)$ such that $d\bigl(z_{m(j-1)+i}, x_i^j\bigr) < \frac{\epsilon}{2}$ for all $j \in (n]$ and $i \in [m_j + K - 1]$, where $m(j) = \sum_{k=1}^{j} (m_k + K + p_k - K) = \sum_{k=1}^{j} (m_k + p_k)$.

Then for all $j \in (n]$ and $i \in [m_j - 1]$, we have

$$\begin{aligned} d_\omega\big(\sigma^{m(j-1)+i}(z), \sigma^i(x^j)\big) &= d_\omega\big((z_{m(j-1)+i}, z_{m(j-1)+i+1}, \dots), \big(x^j_i, x^j_{i+1}, \dots\big)\big) \\ &= \sum_{r=0}^{+\infty} \frac{1}{2^{r+1}} \frac{d\big(z_{m(j-1)+i+r}, x^j_{i+r}\big)}{1 + d\big(z_{m(j-1)+i+r}, x^j_{i+r}\big)} \\ &\leqslant \sum_{r=0}^{K-1} \frac{1}{2^{r+1}} d\big(z_{m(j-1)+i+r}, x^j_{i+r}\big) + \sum_{r=K}^{+\infty} \frac{1}{2^{r+1}} \\ &\leqslant \sum_{r=1}^{K} \frac{1}{2^r}\frac{\epsilon}{2} + \frac{1}{2^K} \\ &= \epsilon. \end{aligned}$$

Therefore we conclude that the shift map σ has the specification property. $\square$

Openness.

Proposition D.4. *If T is open, then σ is an open map.*

Proof. Fix an arbitrary open set $\underline{U} \subseteq \mathcal{O}_\omega(T)$. We show that $\sigma(\underline{U})$ is an open subset of $\mathcal{O}_\omega(T)$. Fix an arbitrary orbit $\underline{y} \in \sigma(\underline{U})$.

Assume that $\underline{x} \in \underline{U}$ and $\sigma(\underline{x}) = \underline{y}$. Since $\underline{U}$ is an open subset of $cO_\omega(T)$, we can choose $n \in \mathbb{N}$ and open subsets $V_1, \dots, V_n$ of X such that $\underline{x} \in V_1 \times \cdots \times V_n \times X^\omega \cap \mathcal{O}_\omega(T) \subseteq \underline{U}$. So $\underline{y} = \sigma(\underline{x}) \in \sigma(V_1 \times \cdots \times V_n \times X^\omega \cap \mathcal{O}_\omega(T)) \subseteq \sigma(\underline{U})$. Since T is open in the sense of Definition 7.11, we have $T(V_1)$ is an open subset of X, and thus $\sigma(V_1 \times \cdots \times V_n \times X^\omega \cap \mathcal{O}_\omega(T)) = (T(V_1) \cap V_2) \times V_3 \times \cdots \times V_n \times X^\omega \cap \mathcal{O}_\omega(T)$ is an open subset of $\mathcal{O}_\omega(T)$. Because $\underline{y} \in \sigma(\underline{U})$ is chosen arbitrarily, we know that $\sigma(\underline{U})$ is an open subset of $\mathcal{O}_\omega(T)$. Therefore, the shift map σ is an open map. $\square$

Proposition D.5. *If σ is an open map, then T is open.*

Proof. Assume that σ is an open map. Then choose an arbitrary open subset U of X. We have that $\sigma(U \times X^\omega \cap \mathcal{O}_\omega(T)) = T(U) \times X^\omega \cap \mathcal{O}_\omega(T)$ is an open subset of $cO_\omega(T)$. Since the projection map is an open map, we have $T(U)$, as the projection of $T(U) \times X^\omega \cap \mathcal{O}_\omega(T)$ on the first coordinate, is an open subset of X. Therefore, the correspondence T is open. $\square$

Strong transitivity.

Proposition D.6. *If T is open and strongly transitive, then σ is topologically transitive.*

Proof. To show that σ is transitive, we choose two arbitrary non-empty open subsets $\underline{U}_1$, $\underline{U}_2$ of $\mathcal{O}_\omega(T)$ and show that there exists $n \in \mathbb{N}$ such that $\sigma^n(\underline{U}_1) \cap \underline{U}_2 \neq \emptyset$. Assume $\underline{U}_1 = V_1 \times \cdots \times V_m \times X^\omega \cap \mathcal{O}_\omega(T)$, where $m \in \mathbb{N}$ and $V_1, \dots, V_m$ are open subsets of X.

Note that

$$\begin{aligned}\sigma^{m-1}(\underline{U}_1) &= \sigma^{m-1}(V_1 \times \cdots \times V_m \times X^\omega \cap \mathcal{O}_\omega(T)) \\ &= \Big(\bigcap_{k=1}^m T^{m-k}(V_k)\Big) \times X^\omega \cap \mathcal{O}_\omega(T).\end{aligned}$$

Since the correspondence T is open in the sense of Definition 7.11 and $V_1, \dots, V_m$ are open subsets of X, we have $\bigcap_{k=1}^m T^{m-k}(V_k)$ is an open subset of X. Moreover, $\bigcap_{k=1}^m T^{m-k}(V_k)$ is non-empty because $\sigma^{m-1}(\underline{U}_1)$ is non-empty due to $\underline{U}_1 \neq \emptyset$. So $\bigcap_{k=1}^m T^{m-k}(V_k)$ is a non-empty open subset of X.

As $\underline{U}_2$ is non-empty, we can choose an orbit $(x_1, x_2, \dots) \in \underline{U}_2$. We have $\bigcup_{n=1}^{+\infty} T^{-n}(x_1)$ is dense in X, ensured by the strong transitivity of T. So there exists $n \in \mathbb{N}$ such that $T^{-n}(x_1) \cap \big(\bigcap_{k=1}^m T^{m-k}(V_k)\big) \neq \emptyset$, i.e., there exists $(y_0, \dots, y_n) \in \mathcal{O}_n(T)$ with $y_0 \in \bigcap_{k=1}^m T^{m-k}(V_k)$ and $y_n = x_1$. Consider the orbit $\underline{x}_0 := (y_0, \dots, y_{n-1}, x_1, x_2, \dots) \in \mathcal{O}_\omega(T)$. Since $y_0 \in \bigcap_{k=1}^m T^{m-k}(V_k)$, we have

$$y_0 \in \Big(\bigcap_{k=1}^m T^{m-k}(V_k)\Big) \times X^\omega \cap \mathcal{O}_\omega(T) = \sigma^{m-1}(\underline{U}_1).$$

Moreover, $\sigma^n(\underline{x}_0) = (x_1, x_2, \dots) \in \underline{U}_2$. Therefore $\sigma^{m+n-1}(\underline{U}_1) \cap \underline{U}_2 \neq \emptyset$. □

Topological exactness.

Proposition D.7. *If T is open and topologically exact, then σ is topologically exact.*

Proof. To show that σ is topologically exact, we fix an arbitrary non-empty open subset $\underline{U}$ of $\mathcal{O}_\omega(T)$ and show that there exists $m \in \mathbb{N}$ such that $\sigma^m(\underline{U}) = \mathcal{O}_\omega(T)$. Choose $\underline{x} \in \underline{U}$. Suppose $\underline{x} \in U_1 \times \cdots U_n \times X^\omega \cap \mathcal{O}_\omega(T) \subseteq \underline{U}$, where $U_1, \dots, U_n$ are open subsets of X. We have $\sigma^{n-1}(\underline{x}) \in \sigma^{n-1}(U_1 \times \cdots U_n \times X^\omega \cap \mathcal{O}_\omega(T)) \subseteq \sigma^{n-1}(\underline{U})$.

Note that $\sigma^{n-1}(\underline{x}) \in \sigma^{n-1}(U_1 \times \cdots U_n \times X^\omega \cap \mathcal{O}_\omega(T)) = \big(\bigcap_{k=1}^n T^{n-k}(U_k)\big) \times X^\omega \cap \mathcal{O}_\omega(T)$, so $\bigcap_{k=1}^n T^{n-k}(U_k)$ is non-empty. Since the correspondence T is open in the sense of Definition 7.11 and $U_1, \dots, U_n$ are open subsets of X, we have $\bigcap_{k=1}^n T^{n-k}(U_k)$ is an non-empty open subset of X. Thereby, the topologically exact property of T indicates that there exists $N \in \mathbb{N}$ such that $T^N\big(\bigcap_{k=1}^n T^{n-k}(U_k)\big) = X$, and thus

$$\begin{aligned}\sigma^{n+N-1}(U_1 \times \cdots U_n \times X^\omega \cap \mathcal{O}_\omega(T)) &= \sigma^N\bigg(\Big(\bigcap_{k=1}^n T^{n-k}(U_k)\Big) \times X^\omega \cap \mathcal{O}_\omega(T)\bigg) \\ = T^N\Big(\bigcap_{k=1}^n T^{n-k}(U_k)\Big) \times X^\omega \cap \mathcal{O}_\omega(T) &= X \times X^\omega \cap \mathcal{O}_\omega(T) = \mathcal{O}_\omega(T).\end{aligned}$$

Therefore, $\sigma^{n+N-1}(\underline{U}) = \mathcal{O}_\omega(T)$. □

Distance-expanding property.

Proposition D.8. *Let T be a distance-expanding correspondence on a compact metric space X. Suppose that $\lambda > 1$, $\eta > 0$, and $n \in \mathbb{N}$ satisfy* $\inf\{d(x', y') : x' \in T^n(x),\, y' \in T^n(y)\} \geqslant \lambda d(x, y)$ *for all $x, y \in X$ with $d(x, y) \leqslant \eta$. Then for an arbitrary $\lambda' \in (1, \lambda)$, there exists $\eta' > 0$ and $k \in \mathbb{N}$ with the property that:*

For each $(x_1, x_2, \dots)$, $(y_1, y_2, \dots) \in \mathcal{O}_\omega(T)$, *if* $d_\omega((x_1, x_2, \dots), (y_1, y_2, \dots)) < \eta'$, *then* $d_\omega(\sigma^{kn}(x_1, x_2, \dots), \sigma^{kn}(y_1, y_2, \dots)) \geqslant \lambda' d_\omega((x_1, x_2, \dots), (y_1, y_2, \dots)$.

In short, if T *is distance-expanding, then* σ *is distance-expanding.*

Proof. Choose $k \in \mathbb{N}$ with $2^{kn} \cdot \frac{\lambda - \lambda'}{2\lambda} \geqslant \lambda'$ and set $\eta' := \frac{1}{2^{2kn}} \min\bigl\{\frac{\eta}{1+\eta}, \frac{\lambda-\lambda'}{2\lambda'(\lambda-1)}\bigr\}$. Fix arbitrary $\underline{x} = (x_1, x_2, \dots)$, $\underline{y} = (y_1, y_2, \dots) \in \mathcal{O}_\omega(T)$ with $d_\omega(\underline{x}, \underline{y}) < \eta'$. We aim to prove $d_\omega(\sigma^{kn}(\underline{x}), \sigma^{kn}(\underline{y})) \geqslant \lambda' d_\omega(\underline{x}, \underline{y})$.

For each $j \in (2kn]$, since

$$\begin{aligned} \frac{1}{2^{2kn}} \min\biggl\{\frac{\eta}{1+\eta}, \frac{\lambda - \lambda'}{2\lambda'(\lambda-1)}\biggr\} = \eta' &> d_\omega((x_1, x_2, \dots), (y_1, y_2, \dots)) \\ &\geqslant \frac{1}{2^j} \frac{d(x_j, y_j)}{1 + d(x_j, y_j)} \\ &\geqslant \frac{1}{2^{2kn}} \frac{d(x_j, y_j)}{1 + d(x_j, y_j)}, \end{aligned}$$

we have $d(x_j, y_j) < \min\bigl\{\eta, \frac{\lambda-\lambda'}{2\lambda\lambda' - \lambda - \lambda'}\bigr\} \leqslant \eta$. This implies $d(x_{j+n}, y_{j+n}) \geqslant \lambda d(x_j, y_j)$ for all $j \in (2kn]$ since $x_{j+n} \in T^n(x_j)$ and $y_{j+n} \in T^n(y_j)$. As a result, for each $j \in (kn]$, we have $d(x_{j+kn}, y_{j+kn}) \geqslant \lambda^k d(x_j, y_j) \geqslant \lambda d(x_j, y_j)$. In addition, $d(x_j, y_j) < \frac{\lambda - \lambda'}{2\lambda\lambda' - \lambda - \lambda'}$ implies $\frac{\lambda+\lambda'}{2\lambda} \frac{\lambda d(x_j, y_j)}{1 + \lambda d(x_j, y_j)} \geqslant \lambda' \frac{d(x_j, y_j)}{1 + d(x_j, y_j)}$, which holds for all $j \in (kn]$. Recall $2^{kn} \cdot \frac{\lambda - \lambda'}{2\lambda} \geqslant \lambda'$. From the arguments above, for every $j \in (kn]$ we have

$$\begin{aligned} 2^{kn} \frac{d(x_{j+kn}, y_{j+kn})}{1 + d(x_{j+kn}, y_{j+kn})} &\geqslant 2^{kn} \frac{\lambda + \lambda'}{2\lambda} \frac{\lambda d(x_j, y_j)}{1 + \lambda d(x_j, y_j)} + 2^{kn} \frac{\lambda - \lambda'}{2\lambda} \frac{d(x_{j+kn}, y_{j+kn})}{1 + d(x_{j+kn}, y_{j+kn})} \\ &\geqslant 2^{kn} \lambda' \frac{d(x_j, y_j)}{1 + d(x_j, y_j)} + \lambda' \frac{d(x_{j+kn}, y_{j+kn})}{1 + d(x_{j+kn}, y_{j+kn})} \end{aligned}$$

Dividing both sides of the inequality above by 2^{kn+j} and then summing over j from 1 to kn, we get

$$\sum_{j=1}^{kn} \frac{1}{2^j} \frac{d(x_{j+kn}, y_{j+kn})}{1 + d(x_{j+kn}, y_{j+kn})} \geqslant \lambda' \sum_{j=1}^{2kn} \frac{1}{2^j} \frac{d(x_j, y_j)}{1 + d(x_j, y_j)} \tag{D.1}$$

Additionally, since $\lambda' \leqslant 2^{kn} \cdot \frac{\lambda - \lambda'}{2\lambda} \leqslant 2^{kn}$, we have

$$\sum_{j=2kn+1}^{+\infty} \frac{1}{2^{j-kn}} \frac{d(x_j, y_j)}{1 + d(x_j, y_j)} \geqslant \lambda' \sum_{j=2kn+1}^{+\infty} \frac{1}{2^j} \frac{d(x_j, y_j)}{1 + d(x_j, y_j)} \tag{D.2}$$

Adding both sides of inequalities (D.1) and (D.2) together, we obtain

$$d_\omega(\sigma^{kn}(\underline{x}), \sigma^{kn}(\underline{y})) \geqslant \lambda' d_\omega(\underline{x}, \underline{y}).$$

The proposition is now established. $\square$

References

[AF09] Aubin, J. and Frankowska, H., *Set-valued analysis*, Springer, Boston, 2009.

[AFL91] Aubin, J., Frankowska, H., and Lasota, A., Poincaré's recurrence theorem for set-valued dynamical systems. *Ann. Polon. Math.* 54 (1991), 85–91.

[BM10] Bonk, M. and Meyer, D., Expanding Thurston maps. Preprint, (arXiv:1009.3647v1), 2010.

[BM17] Bonk, M. and Meyer, D., *Expanding Thurston maps*, volume 225 of *Math. Surveys Monogr.*, Amer. Math. Soc., Providence, RI, 2017.

[Bow75] Bowen, R., *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Math.*, Springer, Berlin, 1975.

[Bu00] Bullett, S., A combination theorem for covering correspondences and an application to mating polynomial maps with Kleinian groups. *Conform. Geom. Dyn.* 4 (2000), 75–96.

[BF05] Bullett, S. and Freiberger, M., Holomorphic correspondences mating Chebyshev-like maps with Hecke groups. *Ergodic Theory Dynam. Systems* 25 (2005), 1057–1090.

[BH07] Bullett, S. and Haïssinsky, P., Pinching holomorphic correspondences. *Conform. Geom. Dyn.* 11 (2007), 65–89.

[BL20] Bullett, S. and Lomonaco, L., Mating quadratic maps with the modular group II. *Invent. Math.* 220 (2020), 185–210.

[BL22] Bullett, S. and Lomonaco, L., Dynamics of modular matings. *Adv. Math.* 410 (2022), 108758.

[BL23] Bullett, S. and Lomonaco, L., Mating quadratic maps with the modular group III: The modular Mandelbrot set. Preprint, (arXiv:2010.04273), 2023.

[BP94] Bullett, S. and Penrose, C., Mating quadratic maps with the modular group. *Invent. Math.* 115 (1994), 483–511.

[BP01] Bullett, S. and Penrose, C., Regular and limit sets for holomorphic correspondences. *Fund. Math.* 167 (2001), 111–171.

[CPMP08] Chinchuluun, A., Pardalos, P., Migdalas, A., and Pitsoulis, L., *Pareto, Optimality, Game Theory and Equilibria*, Springer, New York, 2008.

[CP16] Cordeiro, W. and Pacífico, M., Continuum-wise expansiveness and specification for set-valued functions and topological entropy. *Proc. Amer. Math. Soc.* 144 (2016), 4261–4271.

[DKW20] Dinh, T., Kaufmann, L., and Wu, H., Dynamics of holomorphic correspondences on Riemann surfaces. *Internat. J. Math.* 31(05) (2020), 2050036.

[Do68] Dobrušin, R., Description of a random field by means of conditional probabilities and conditions for its regularity. *Teor. Veroyatn. Primen.* 13 (1968), 201–229.

[Fa29] Fatou, P., Notice sur les travaux scientifiques de MP Fatou, 1929.

[Fol99] Folland, G.B., *Real analysis: modern techniques and their applications*, 2nd ed., Wiley, New York, 1999.

[For88] Forbus, K., Chapter 7 – *Qualitative Physics: Past, Present, and Future*, Exploring Artificial Intelligence Survey Talks from the National Conferences on Artificial Intelligence 1988, Pages 239–296.

[Ga16] Gall, J., *Brownian motion, martingales, and stochastic calculus*, Springer ,Switzerland, 2016.

[GS74] Gihman, I. and Skorohod, A., *The Theory of Stochastic Process I*, Springer, Berlin, 1974.

[Ha02] Hadamard, J., Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin* 13 (1902), 49–52.

[HP09] Haïssinsky, P. and Pilgrim, K.M., Coarse expanding conformal dynamics. *Astérisque* 325 (2009).

[IM06] Ingram, W. and Mahavier, W., Inverse limits of upper semi-continuous set valued functions (English summary). *Houston J. Math.* 32 (2006), no. 1, 119–130.

[KT17] Kelly, J. and Tennant, T., Topological entropy of set-valued functions. *Houston J. Math.* 43(1) (2017), 263–282.

[Ki98] Kitchens, B.P., *Symbolic dynamics: one-sided, two-sided, and countable state Markov shifts*, Springer, Berlin, 1998.

[Ku66] Kuratowski, K., *topology I*, Elsevier Inc, Academic Pr, 1966.

[LLMM21] Lee, S., Lyubich, M.Yu., Makarov, N.G., and Mukherjee, S., Schwarz reflections and anti-holomorphic correspondences. *Adv. Math.* 385 (2021), 107766.

[LMM23] Lyubich, M.Yu., Mazor, J., and Mukherjee, S., Antiholomorphic correspondences and mating I: realization theorems. Preprint, (arXiv:2303.02459), 2023.

[LM97] Lyubich, M.Yu. and Minsky, Y., Laminations in holomorphic dynamics. *J. Differential Geom.* 47 (1997), 17–94.

[Mc95] McMullen, C.T., The classification of conformal dynamical systems. In *Current developments in mathematics*, S.T. Yau et al. (ed.), Cambridge, MA 1995, pp. 323–360.

[Mc96] McMullen, C.T., *Renormalization and 3-manifolds which fiber over the circle*, Princeton Univ. Press, 1996.

[Mic56a] Michael, E., Continuous selections I. *Ann. of Math. (2)* 63 (1956), 361–382.

[Mic56b] Michael, E., Continuous selections II. *Ann. of Math. (2)* 64 (1956), 562–580.

[Mic57] Michael, E., Continuous selections III. *Ann. of Math. (2)* 65 (1957), 375–390.

[Mil95] Miller, W., Frobenius–Perron operators and approximation of invariant measures for set-valued dynamical systems. *Set-Valued Var. Anal.* 3 (1995), 181–194.

[MA99] Miller, W. and Akin, E., Invariant measures for set-valued dynamical systems. *Trans. Amer. Math. Soc.* 351 (1999), 1203–1225.

[MM23] Mj, M. and Mukherjee, S., Combining rational maps and Kleinian groups via orbit equivalence. *Proc. Lond. Math. Soc. (3)* 126 (2023), 1740–1809.

[MT12] Meyn, S. and Tweedie, R., *Markov chains and stochastic stability*, Springer, London, 2012.

[Par23a] Parra, V.M., Equidistribution for matings of quadratic maps with the modular group. *Ergodic Theory Dynam. Systems* (2023), 1–29.

[Par23b] Parra, V.M., Entropy of compositions of covering correspondences. Preprint, (arXiv:2310.14330), 2023.

[Pa64] Parry, W., Intrinsic Markov chains. *Trans. Amer. Math. Soc.* 112 (1964), 55–66.

[Pe93] Petrosyan, L., *Differential Games Of Pursuit*, World Scientific. 1993.

[Pil03] Pilgrim, K.M., *Combinations of Complex Dynamical Systems*, Lecture Notes in Mathematics. Springer, Berlin, Heidelberg, 2003.

[PV17] Pacifico, M.J. and Vieitez, J., Expansiveness, Lyapunov exponents and entropy for set valued maps. Preprint, (arXiv:1709.05739), 2017.

[Po21] Pommaret, J., Differential Correspondences and Control Theory. *Advances in Pure Math.* 11,(2021), 835–882.

[PU10] Przytycki, F. and Urbański, M., *Conformal fractals: ergodic theory methods*, Cambridge Univ. Press, Cambridge, 2010.

[RT18] Raines, E. and Tennant, T., The specification property on a set-valued map and its inverse limit. *Houston J. Math.* 44 (2018), no. 2, 665–677.

[Re84] Revuz, D., *Markov chains*, North-Holland, 1984.

[Ru78] Ruelle, D., *Thermodynamic formalism*, Addison-Wesley, Reading, MA, 1978.

[Ru92] Ruelle, D., Thermodynamic formalism of maps satisfying positive expansiveness and specification. *Nonlinearity* 5 (1992), 1223–1236.

[Sin72] Sinai, Ya.G., Gibbs measures in ergodic theory. *Russian Math. Surveys* 27 (1972), 21–69.

[Siq15] Siqueira, C., Dynamics of holomorphic correspondences. PhD thesis, University of São Paulo, 2015.

[Siq22] Siqueira, C., Dynamics of hyperbolic correspondences. *Ergodic Theory Dynam. Systems* 42 (2022), 2661–2692.

[Siq23] Siqueira, C., Hausdorff dimension of Julia sets of unicritical correspondences. *Proc. Amer. Math. Soc.* 151 (2023), 633–645.

[SS17] Siqueira, C. and Smania, D., Holomorphic motions for unicritical correspondences. *Nonlinearity* 30 (2017), 3104–3125.

[Su85] Sullivan, D., Quasiconformal homeomorphisms and dynamics I: Solution of the Fatou-Julia problem on wandering domains. *Ann. of Math. (2)* 122 (1985), 401–418.

[VS22] Vivas, K. and Sirvent, V., Metric entropy for set-valued maps. *Discrete Contin. Dyn. Syst. Ser. B* 27 (2022), 6589—6604.

[Wa76] Walters, P., A variational principle for the pressure of continuous transformations. *Amer. J. Math.* 17 (1976), 937–971.

[Wa82] Walters, P., *An introduction to ergodic theory*, Springer, New York, 1982.

[Wi70] Williams, R., Some results on expansive mappings. *Proc. Amer. Math. Soc.* 26 (1970), 655–663.

[Wu20] Wu, H., Dynamics of Holomorphic Maps and Correspondences. PhD thesis, National University of Singapore, 2020.

Index

Xiaoran Li, School of Mathematical Sciences, Peking University, Beijing 100871, China

Email address: 2000010758@stu.pku.edu.cn

Zhiqiang Li, School of Mathematical Sciences & Beijing International Center for Mathematical Research, Peking University, Beijing 100871, China

Email address: zli@math.pku.edu.cn

Yiwei Zhang, School of Mathematics and Statistics, Center for Mathematical Sciences, Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Science and Technology, Wuhan 430074, China

Email address: yiweizhang@hust.edu.cn