Analyticity of Entropy Rate of Hidden Markov Chains with Continuous Alphabet *

Guangyue Han
The University of Hong Kong
email: ghan@hku.hk

Brian Marcus
The University of British Columbia
email: marcus@math.ubc.ca

August 29, 2014

Abstract

We first prove that under certain mild assumptions, the entropy rate of a hidden Markov chain, observed when passing a finite-state stationary Markov chain through a discrete-time continuous-output channel, is analytic with respect to the input Markov chain parameters. We then further prove, under strengthened assumptions on the channel, that the entropy rate is jointly analytic as a function of both the input Markov chain parameters and the channel parameters. In particular, the main theorems establish the analyticity of the entropy rate for two representative channels: Cauchy and Gaussian.

Index Terms: hidden Markov chain, entropy rate, analyticity, continuous alphabet, Hilbert metric.

1 Main Results and Related Work

Introduction and background. Consider a discrete-time channel with a finite input alphabet $\mathcal{Y} = \{1, 2, \cdots, l\}$ and the continuous output alphabet $\mathcal{Z} = \mathbb{R}$. We assume that the input process is a $\mathcal{Y}$-valued first-order stationary Markov chain Y with transition probability matrix $\Pi = (\pi_{ij})_{l \times l}$ and stationary vector $\pi = (\pi_i)_{1 \times l}$ (here we assume Y is first-order only for simplicity; a standard "blocking" approach can be used to reduce higher order cases to the first-order case). We assume that the channel is memoryless in the sense that at each time, the distribution of the output $z \in \mathcal{Z}$, given the input $y \in \mathcal{Y}$, is independent of the past and future inputs and outputs, and is distributed according to a probability density function $q(z|y)$.

The corresponding output process of this channel is a *hidden Markov chain*, which will be denoted by Z throughout the paper. The entropy rate $h(Z)$ is defined as

$$h(Z) = \lim_{n \to \infty} \frac{1}{n+1} h(Z_{-n}^0), \tag{1}$$

*A preliminary version of this paper has been presented in IEEE ISIT 2010 [14].

when the limit exists, where [1]

$$h(Z_{-n}^0) = -\int_{\mathcal{Z}^{n+1}} p(z_{-n}^0) \log p(z_{-n}^0) dz_{-n}^0;$$

here $z_{-n}^0 \triangleq (z_{-n}, z_{-n+1}, \cdots, z_0)$ denotes an instance of $Z_{-n}^0 \triangleq (Z_{-n}, Z_{-n+1}, \cdots, Z_0)$, and $p(z_{-n}^0)$ denotes the probability density of z_{-n}^0. It is well-known (e.g., see page 60 of [19]) that if $h(Z_{-n}^0)$ is finite for all n, then the limit in (1) exists and can be written as

$$h(Z) = \lim_{n\to\infty} h_n(Z),$$

where

$$h_n(Z) = -\int_{\mathcal{Z}^{n+1}} p(z_{-n}^0) \log p(z_0|z_{-n}^{-1}) dz_{-n}^0; \tag{2}$$

here $p(z_0|z_{-n}^{-1})$ denotes the conditional density of z_0 given z_{-n}^{-1}. Since the channels considered in this paper are memoryless and Y is stationary, we have

$$h(Z_{-n}^0|Y_{-n}^0) = (n+1)h(Z_0|Y_0),$$

where

$$h(Z_0|Y_0) = -\sum_{y\in\mathcal{Y}} \pi_y \int_{\mathcal{Z}} q(z|y) \log q(z|y) dz.$$

It then follows from

$$(n+1) = h(Z_0|Y_0)h(Z_{-n}^0|Y_{-n}^0) \leq h(Z_{-n}^0) \leq h(Y_{-n}^0) + h(Z_{-n}^0|Y_{-n}^0)$$

that if

$$\int_{z\in\mathcal{Z}} q(z|y) \log q(z|y) dz \text{ is finite for all } y, \tag{3}$$

so is $h(Z_0|Y_0)$ and then $h(Z_{-n}^0)$, which implies that $h(Z)$ as in (1) is well-defined and finite.

Unless specified otherwise, we will assume that $\Pi = \Pi^{\vec{\varepsilon}} = (\pi_{ij}^{\vec{\varepsilon}})$ is analytically parameterized by $\vec{\varepsilon} \in \Omega_1$, where Ω_1 denotes a bounded domain in $\mathbb{R}^{m_1}$ (a domain is an open and connected set), and for any $(y, z) \in \mathcal{Y} \times \mathcal{Z}$, $q^{\vec{\theta}}(z|y)$ is analytically parameterized by $\vec{\theta} \in \Omega_2$, where Ω_2 denotes a bounded domain in $\mathbb{R}^{m_2}$.

Main results. We are now ready to state our main results. Here, we remark that all our results in this paper can be straightforwardly translated to the setting where $\mathcal{Z}$ is finite or countably infinite.

Under some mild regularity condition, we prove the following theorem, which establishes the analyticity of $h(Z)$ as a function of the underlying Markov chain parameters.

Theorem 1.1. *Assume that for any $y \in \mathcal{Y}$, $q(z|y)$ is positive and continuous on $\mathcal{Z}$, and the following two integrals*

$$\int_{\mathcal{Z}} q(z|y)|\log \min_{y'} q(z|y')|dz, \qquad \int_{\mathcal{Z}} q(z|y)|\log \max_{y'} q(z|y')|dz \tag{4}$$

are finite. If Π is strictly positive at $\vec{\varepsilon}_0$, then $h(Z)$ is analytic around $\vec{\varepsilon}_0$.

[1]Throughout the paper, since all the integrands are at least continuous and most are analytic, all the integrals should be interpreted as Riemann integrals.

Remark 1.2. Theorem 1.1 has been announced in [14], where Condition (3) has been assumed. Unfortunately, we have found that, at least for the rigorous proof in this paper, Condition (3) is not strong enough. More specifically, as in the proof of Theorem 1.1, we need to use (4), a strengthened version of Condition (3), to establish the analyticity of $h_n^{\vec{\varepsilon}}(Z_0|Z_{-n}^{-1})$ on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$; here and throughout the paper, $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$ denotes the r_1-ball of $\vec{\varepsilon}_0$ in $\mathbb{C}^{m_1}$ with respect to the Euclidean metric.

Remark 1.3. When $\mathcal{Z}$ is finite, the analyticity of $h(Z)$ has been established under mild positivity assumptions in [11]. Due to the fact that for any y, $q(z|y)$ can be arbitrarily close to 0, the complexfication and uniform convergence arguments in [11] do not carry over to Theorem 1.1. As elaborated in Sections 5 and 6, the proof of Theorem 1.1 requires a critical use of the complex Hibert metric in Section 5.

We will also prove analyticity of $h(Z)$ as a function of both the Markov chain parameters and channel parameters. Simple examples show that $h(Z)$ can fail to be analytic as a function of the channel parameter alone; see Example 4.1 and Remark 4.2. Our positive results require several technical regularity conditions, which we describe as follows. These conditions involve the complexification of the channel density functions (by definition, any real analytic function, such as $q^{\vec{\theta}}(z|y)$ as above, at a given point can be uniquely extended to a complex analytic function on some complex neighborhood of the given point; we will continue to use the same notation, such as $q^{\vec{\theta}}(z|y)$, for this complex extension). We require these technical conditions, which abstract the properties of commonly used probability density functions (e.g., Cauchy and Gaussian), in order to make our proofs work. There may be more general conditions that suffice. On a first reading, the reader may want to skip directly to the statements of results below.

Before listing our regularity conditions, we remind the reader that a family of functions $\{f(\vec{\theta}, z)\}_z$ of $\vec{\theta}$ is equicontinuous if for any $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(\vec{\theta}_1, z) - f(\vec{\theta}_2, z)| \leq \varepsilon$ for all z as long as $\|\vec{\theta}_1 - \vec{\theta}_2\| \leq \delta$. Note that uniform boundedness (over all feasible z) of the derivatives of $f(\vec{\theta}, z)$ with respect to $\vec{\theta}$ will imply its equicontinuity.

The regularity conditions are as follows: For given $(\vec{\varepsilon}_0, \vec{\theta}_0) \in \Omega_1 \times \Omega_2$,

(a) Π is strictly positive at $\vec{\varepsilon}_0$;

(b) for all $y \in \mathcal{Y}$, $q^{\vec{\theta}_0}(z|y)$ is positive on $\mathcal{Z}$;

(c) there exists $r_2 > 0$ such that

  (i) for any $(y, z) \in \mathcal{Y} \times \mathcal{Z}$, $q^{\vec{\theta}}(z|y)$ is analytic on $\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

  (ii) for any $y \in \mathcal{Y}$, $q^{\vec{\theta}}(z|y)$ is jointly continuous on $\mathcal{Z} \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

  (iii) the following three integrals

$$\int \check{q}(z, r_2) dz, \quad \int \check{q}(z, r_2) \log \hat{q}(z, r_2) dz, \quad \int \check{q}(z, r_2) \log \check{q}(z, r_2) dz, \tag{5}$$

  are all finite, where

$$\check{q}(z, r_2) \triangleq \sup_{(y, \vec{\theta}) \in \mathcal{Y} \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}}(z|y)|, \qquad \hat{q}(z, r_2) \triangleq \inf_{(y, \vec{\theta}) \in \mathcal{Y} \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}}(z|y)|;$$

(d) for some $I \in \{1, 2, \dots, l\}$,

(i) there exist $r_2 > 0$ such that for all j, the family of functions $\{q^{\vec{\theta}}(z|j)/q^{\vec{\theta}}(z|I)\}_z$ on $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ is equicontinuous,

(ii) there exists $r_2 > 0$ such that for each z, the real $\log q^{\vec{\theta}}(z|I)$ can be analytically extended to $\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ and for all j,

$$\int_{\mathcal{Z}} \sup_{\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} \left| q^{\vec{\theta}}(z|j) \log q^{\vec{\theta}}(z|I) \right| dz < \infty. \tag{6}$$

It is easily seen that the most commonly used channel models, including Cauchy and Gaussian channels, satisfy all of these conditions; see Examples 1.6 and 1.7.

The following theorem is our first result on the joint analyticity of $h(Z)$.

Theorem 1.4. *For given $(\vec{\varepsilon}_0, \vec{\theta}_0) \in \Omega_1 \times \Omega_2$, assume Conditions (a), (b), (c) and (d). If, in addition, one of the following two conditions holds,*

- *there exist $C', C'' > 0$ such that for all i, j and all $z \in \mathcal{Z}$,*

$$C' \le \left| \frac{q^{\vec{\theta}_0}(z|j)}{q^{\vec{\theta}_0}(z|i)} \right| \le C''; \tag{7}$$

- *for the same I as in Condition (d), there exists $r_2 > 0$ such that for any $\varepsilon > 0$, there exists a compact subset $\Sigma \subset \mathcal{Z}$ such that for all $z \notin \Sigma$, all $j \neq I$ and all $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$*

$$\left| \frac{q^{\vec{\theta}}(z|j)}{q^{\vec{\theta}}(z|I)} \right| \le \varepsilon, \tag{8}$$

then $h(Z)$ is analytic around $(\vec{\varepsilon}_0, \vec{\theta}_0)$.

Remark 1.5. Conditions (7) and (8) are in fact mutually exclusive in the sense that if one of them holds, then the other one must fail. The reader should also note that while Theorem 1.4 requires Condition (7) to hold only at one given parameter value, $\vec{\theta}_0$, Condition (8) needs to hold for all parameter values $\vec{\theta}$ in a complex neighborhood of the given $\vec{\theta}_0$. On the other hand, if Condition (8) holds only at $\vec{\theta}_0$, then one can conclude that $h(Z)$ is analytic with respect to $\vec{\varepsilon}$ at $\vec{\varepsilon}_0$. Theorem 1.1, however, says that a condition as simple as (4) is already sufficient to imply the analyticity of $h(Z)$ with respect to $\vec{\varepsilon}$ alone.

In the following example, we show that Conditions (b)-(d) and (7) are satisfied for additive Cauchy channels.

Example 1.6. Consider an additive Cauchy channel with channel transition probability function taking the following form:

$$q^{\vec{\theta}}(z|i) = \frac{1}{\pi} \frac{\gamma_i}{(z - \mu_i)^2 + \gamma_i^2}, \quad \gamma_i > 0, \quad i = 1, 2, \dots, l, \tag{9}$$

which is parameterized by $\vec{\theta} \in \Omega_2$, where

$$\Omega_2 \triangleq \{(\gamma_1, \mu_1, \gamma_2, \mu_2, \ldots, \gamma_l, \mu_l) \in \mathbb{R}^{2l} : \gamma_i > 0, i = 1, 2, \ldots, l\}.$$

Let $\vec{\theta}_0 \in \Omega_2$.

Obviously, Condition (b) is trivially satisfied.

Now, for Condition (c), note that since $\gamma_i > 0$ for each i, each $(z - \mu_i)^2 + \gamma_i^2$ is bounded away from 0 for all $\vec{\theta} \in \mathbb{C}^{2l}_{\vec{\theta}_0}(r_2)$ if r_2 is sufficiently small. The existence of r_2 for (i) and (ii) then immediately follows. As for (iii), notice that for sufficiently small r_2, both $\check{q}(z, r_2)$ and $\hat{q}(z, r_2)$ are of order $O(1/z^2)$ for large z and are uniformly bounded below away from 0 for all z; this implies that the three integrals in (5) are all finite.

We now check Condition (d) with I set to be 1 (here I can be chosen arbitrarily). For (i), observe that for sufficiently small r_2, the derivative of $q^{\vec{\theta}}(z|j)/q^{\vec{\theta}}(z|1)$ with respect to $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ is bounded uniformly over all j and $z \in \mathcal{Z}$. As for (ii), observe that for sufficiently small r_2, the imaginary part of $q^{\vec{\theta}}(z|1)$ is dominated by the real part, which implies that $\log q^{\vec{\theta}}(z|1)$ can be analytically extended to $\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$, and the finiteness of the integral in (6) follows from the fact that for large z, each $q^{\vec{\theta}}(z|i)$ is of order $O(1/z^2)$.

For Condition (7), it can be easily checked that for all i, j,

$$\lim_{z \to \infty} \frac{q^{\vec{\theta}_0}(z|j)}{q^{\vec{\theta}_0}(z|i)} = 1,$$

which immediately implies Condition (7).

So, for an additive Cauchy channel, if Π is strictly positive at $\vec{\varepsilon}_0 \in \Omega_1$, then $h(Z)$ is analytic around $(\vec{\varepsilon}_0, (\gamma_1, \mu_1, \gamma_2, \mu_2, \ldots, \gamma_l, \mu_l))$.

In the following example, we show that Conditions (b)-(d) and (8) are satisfied for additive Gaussian channels.

Example 1.7. Consider an additive Gaussian channel with channel transition probability function taking the following form:

$$q^{\vec{\theta}}(z|i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(z-\mu_i)^2/(2\sigma_i^2)}, \quad \sigma_i > 0, \ i = 1, 2, \ldots, l, \text{ and all } \sigma_i \text{ are distinct}, \qquad (10)$$

which is parameterized by $\vec{\theta} \in \Omega_2$, where

$$\Omega_2 \triangleq \{(\sigma_1, \mu_1, \sigma_2, \mu_2, \ldots, \sigma_l, \mu_l) \in \mathbb{R}^{2l} : \sigma_i > 0, i = 1, 2, \ldots, l\}.$$

Let $\vec{\theta}_0 \in \Omega_2$.

Obviously, Condition (b) is trivially satisfied.

Now, for Condition (c), note that (i) and (ii) hold for any r_2 with $0 < r_2 < \min_i\{\sigma_i\}$. And for any such fixed r_2, there exists $C_1, C_1', C_2, C_2' > 0$, which are independent of z, such that

$$C_1 e^{-C_1' z^2} \leq \check{q}(z, r_2), \quad \hat{q}(z, r_2) \leq C_2 e^{-C_2' z^2},$$

which immediately implies (iii).

We now check Condition (d) with I set to be the index i corresponding to the largest σ_i. For (i), observe that for sufficiently small r_2, the derivative of $q^{\vec{\theta}}(z|j)/q^{\vec{\theta}}(z|I)$ with respect to $\vec{\theta}$ is bounded on $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ uniformly over all j and $z \in \mathcal{Z}$, which immediately implies the equicontinuity. As for (ii), observe that

$$\log q^{\vec{\theta}}(z|I) = -\log(\sqrt{2\pi}\sigma_I) - \frac{(z-\mu_I)^2}{2\sigma_I^2},$$

which can be analytically extended to $\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ for sufficiently small r_2. And the finiteness of the integral in (6) follows from the fact that for large z, $q^{\vec{\theta}}(z|j)$ is of order $e^{-O(z^2)}$ for all j.

For Condition (8), it can be easily checked that for all i, j,

$$\lim_{z\to\infty} \frac{q^{\vec{\theta}_0}(z|j)}{q^{\vec{\theta}_0}(z|I)} = 0,$$

which immediately implies Condition (8).

So, for an additive Gaussian channel, if Π is strictly positive at $\vec{\varepsilon}_0 \in \Omega_1$, then $h(Z)$ is analytic around $(\vec{\varepsilon}_0, (\sigma_1, \mu_1, \sigma_2, \mu_2, \cdots, \sigma_l, \mu_l))$.

Our second result on joint analyticity deals with the case when for all i, the real part of $q^{\theta}(z|i)$ "dominates" the imaginary part of the complex extension.

Theorem 1.8. *For given $(\vec{\varepsilon}_0, \vec{\theta}_0) \in \Omega_1 \times \Omega_2$, assume Conditions (a), (b) and (c). If, in addition, for any $\delta > 0$, there exists $r_2 > 0$ such that for all $(y, z) \in (\mathcal{Y}, \mathcal{Z})$ and all $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$*

$$(i)\quad |\Im(q^{\vec{\theta}}(z|y))| < \delta |\Re(q^{\vec{\theta}}(z|y))|, \quad (ii)\quad \left|\log \frac{q^{\vec{\theta}}(z|y)}{q^{\vec{\theta}_0}(z|y)}\right| \le \delta, \tag{11}$$

then $h(Z)$ is analytic around $(\vec{\varepsilon}_0, \vec{\theta}_0)$.

Theorem 1.8 applies to Cauchy channels (9) roughly because small complex perturbations of the parameters will have small effect on the imaginary part of $q^{\vec{\theta}}(z|i)$ uniformly over z and i. However, it need not apply to Gaussian channels (10) because a small imaginary perturbation of a variance may overwhelm the real part of $q^{\vec{\theta}}(z|i)$ for large z. Theorem 1.8 also applies to other channels for which Conditions (7) and (8) fail. Roughly speaking, Condition (7) says that all $q^{\vec{\theta}}(z|i)$ are in some sense "comparable" with one another at given point $\vec{\theta}_0$, and Condition (8) says that one of $q^{\vec{\theta}}(z|i)$ dominates all the others. It is not surprising that one can find channels that do not satisfy both conditions, however satisfy (11). A simple yet somewhat artificial example of such channels with channel parameter $\vec{\theta} = (\mu_1, \gamma_1, \mu_2, \gamma_2)$ is defined as follows:

$$q^{\vec{\theta}}(z|i) = \frac{1}{\pi}\frac{\gamma_i}{(z-\mu_i)^2 + \gamma_i^2}, \quad \gamma_i > 0, \quad i = 1, 2,$$

and

$$q(z|3) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

It is easy to see that $q^{\vec{\theta}}(z|1)$ is comparable to $q^{\vec{\theta}}(z|2)$, however not to $q(z|3)$; and clearly, none of the three distributions dominates. On the other hand, such a channel satisfies the assumptions in Theorem 1.8, and therefore $h(Z)$ is analytic with respect to $(\mu_1, \gamma_1, \mu_2, \gamma_2)$ for the same reason as Theorem 1.8 applies to Cauchy channels in Example 1.6.

Theorems 1.4 and 1.8 can be regarded as extensions of [11, Theorem 1.1], which deals with the case where $\mathcal{Z}$ is finite. In particular, the flow of the proof of Theorem 1.4 follows closely that of this case. However, new techniques are needed to deal with the continuous case. The extension to Theorems 1.8 is less direct, where the use of a complex Hilbert metric [15] to replace the classical real Hilbert metric, is critical. This metric was also used in the proof of Theorem 1.1. It is not needed for Theorem 1.4, because it assumes stronger conditions on the channel density functions.

We remark that in [11, Theorem 1.1], zero values are allowed for some transition probabilities. It seems more difficult to handle this phenomena in the continuous-alphabet setting; this is the subject of forthcoming work.

To the best of our knowledge, the results in this paper, together with those in [14], are among the first results establishing analyticity of the entropy rate of hidden Markov chains with continuous alphabet. Nevertheless, following the approach in [11], the analyticity of the expected log-likelihood for certain hidden Markov models has been established in [40]. Our paper has some similarities with [40]. Specifically, the entropy rate of a hidden Markov chain corresponding to parameter θ is exactly the expected value of the log-likelihood when both the expectation and the log-likelihood are taken with the same parameter value. So, to prove analyticity of the entropy rate, we must let the expectation and log-likelihhod vary together. In contrast, in [40], the expectation is taken with respect to a fixed parameter value.

Related work in finite alphabet. As opposed to continuous alphabet, the entropy rate of hidden Markov chains with finite alphabet has been extensively studied. Next, we briefly review the related results in the literature. In the remainder of this section, unless specified otherwise, we assume $\mathcal{Z}$ is a finite alphabet and X is a finite-state stationary Markov chain with transition probability matrix Π and stationary vector π; then, the output process Z is a finite-state hidden Markov chain.

It is well known that $h(X)$, the entropy rate of X, has a simple analytic formula in terms of Π and π; in stark contrast, there is no simple and explicit formula of $h(Z)$ for most non-degenerate channels ever since hidden Markov chains (or, more precisely, hidden Markov models) were formulated half a century ago. Here, we remark that Blackwell [4] showed that $h(Z)$ can be written as an integral of an explicit function on a simplex with respect to the Blackwell Measure. However, the Blackwell measure seems to be rather complicated for effective computation of $h(Z)$.

Recently there has been a rebirth of interest in computing and estimating $h(Z)$ in a variety of scenarios, and many approaches have been adopted to tackle this problem: the Blackwell measure has been used to bound $h(Z)$ [27], a variation on the classical Birch bounds [3] can be found in [7] and a new numerical approximation of $h(Z)$ has been proposed in [25]. Generalizing Blackwell's idea, an integral formula for the derivatives of $h(Z)$ has been derived in [36]. In another direction, [2, 8, 18, 27, 42, 43, 26, 21, 33, 36, 29] have studied the variation of the entropy rate as parameters of the underlying Markov chain vary. Particularly in [42], for a special type of hidden Markov chain Z, the Taylor series expansion of $h(Z)$ is given, under the assumption that $h(Z)$ is analytic.

Under mild positivity assumptions, we prove [11] that, $h(Z)$ is an analytic function of Π, thereby confirming the analyticity assumption in [42]. Aside from its significance in mathematics (see, e.g., [31, 32, 5, 22]), this analyticity result is of interest to a wide range of applications (see, e.g., [42, 43, 1, 25, 40, 36]). In particular, the analyticity result and techniques employed in [11], such as complexification and exponential convergence, are rather useful in many aspects of information theory as well. Using the ideas for proving the analyticity result, we derive [16] an asymptotic formula of $h(Z)$ at so-called "weak Black Holes", rather general settings which include "Black Holes" [12] and the input-restricted channels in [13] as special cases. These results provide much insight for characterizing [13] the asymptotic behavior (as ε tends to zero) of capacity of a class of input-restricted channels and deriving [17] concavity of the mutual information rate for such channels. Here, we remark that the concavity result in [17] confirms the concavity conjecture in [30] in a special scenario; on the other hand, the conjecture in the general setting has been disproved in [24] by counter-examples constructed based on the results in [12]. It turns out the result and techniques in [11] are instrumental in terms of computing the mutual information rate and the channel capacity of finite-state channel as well. Recently, employing analyticity in a critical way, we have established refinements of the Shannon-McMillan-Breiman theorem which sheds light on effective computation of the mutual information rate of a class of finite-state channels and we further proposed [10] a randomized algorithm to compute the capacity of a class of finite-state channels, whose convergence behavior is analyzed using the analyticity result.

In this paper, we will establish the analyticity of the entropy rate of hidden Markov chains with continuous alphabet. Given the implications of the counterpart results in the discrete setting, we expect that the results in this paper will be of great interest and significance in relevant areas.

Organization of the paper. The remainder of this paper is organized as follows. In Section 2, we review the (real) Hilbert metric, outline the framework of the proofs of our theorems and highlight the differences among the proofs. We prove Theorem 1.4 in Section 3; on the other hand, we show that analyticity fails for the Gaussian channel with uniform i.i.d. inputs in Section 4. Section 5 is devoted to reviewing the complex Hilbert metric, which is of critical use to the proofs of Theorem 1.1 in Section 6 and Theorem 1.8 in Section 7.

2 The Main Idea of the Proofs

We first briefly review the classical (real) Hilbert metric. The real Hilbert metric will be used in the proof of Theorem 1.4.

Let W be the standard simplex in the l-dimensional real Euclidean space,

$$W = \{w = (w_1, w_2, \cdots, w_l) \in \mathbb{R}^l : w_i \geq 0, \sum_i w_i = 1\},$$

and let W° denote its interior, consisting of the vectors with positive coordinates. For any two vectors $v, w \in W^\circ$, the Hilbert metric [38] is defined as

$$d_H(w, v) = \max_{i,j} \log \left(\frac{w_i/w_j}{v_i/v_j} \right). \tag{12}$$

Remark 2.1. It immediately follows from the definition that for any diagonal matrix A with strictly positive diagonal entries, we have

$$d_H(w, v) = d_H(wA, vA).$$

For an $l \times l$ positive matrix $T = (t_{ij})$ (i.e., each $t_{ij} > 0$), the mapping f_T induced by T on W is defined by

$$f_T(w) = \frac{wT}{wT\mathbf{1}}, \tag{13}$$

where $\mathbf{1}$ is the all 1's column vector. The following theorem is well-known (see [38]).

Theorem 2.2. *For a positive T, f_T is a contraction mapping on the entire W° under the Hilbert metric and the contraction coefficient (often referred to as the Birkhoff coefficient), is given by*

$$\tau(T) = \sup_{v \neq w} \frac{d_H(vT, wT)}{d_H(v, w)} = \frac{1 - \sqrt{\phi(T)}}{1 + \sqrt{\phi(T)}}, \tag{14}$$

where $\phi(T) = \min_{i,j,k,l} \frac{t_{ik} t_{jl}}{t_{jk} t_{il}}$.

We will also need a complex version of W,

$$\tilde{W} = \{w = (w_1, w_2, \cdots, w_l) \in \mathbb{C}^l : \sum_i w_i = 1\}.$$

Now, for each $z \in \mathcal{Z}$, define $\Pi^{\vec{\varepsilon},\vec{\theta}}(z)$ as an $l \times l$ matrix with the entries

$$\Pi^{\vec{\varepsilon},\vec{\theta}}(z)_{ij} = \pi^{\vec{\varepsilon}}_{ij} q^{\vec{\theta}}(z|j), \text{ for all } i, j. \tag{15}$$

By (13), $\Pi^{\vec{\varepsilon},\vec{\theta}}(z)$ will induce a mapping $f_z^{\vec{\varepsilon},\vec{\theta}} \triangleq f_{\Pi^{\vec{\varepsilon},\vec{\theta}}(z)}$ from W to W. For any fixed n and z_{-n}^0, define

$$x_i^{\vec{\varepsilon},\vec{\theta}} = x_i^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^i) = p^{\vec{\varepsilon},\vec{\theta}}(y_i = \cdot\, | z_i, z_{i-1}, \cdots, z_{-n}), \tag{16}$$

(here $\cdot$ represent the states of the Markov chain Y) then similar to Blackwell [4], $\{x_i^{\vec{\varepsilon},\vec{\theta}}\}$ satisfies the random dynamical system

$$x_{i+1}^{\vec{\varepsilon},\vec{\theta}} = f_{z_{i+1}}^{\vec{\varepsilon},\vec{\theta}}(x_i^{\vec{\varepsilon},\vec{\theta}}), \tag{17}$$

starting with

$$x_{-n-1}^{\vec{\varepsilon},\vec{\theta}} = \pi^{\vec{\varepsilon}}, \tag{18}$$

where $\pi^{\vec{\varepsilon}}$ is the stationary vector for $\Pi^{\vec{\varepsilon}}$. And it can be verified that

$$p^{\vec{\varepsilon},\vec{\theta}}(z_0 | z_{-n}^{-1}) = x_{-1}^{\vec{\varepsilon},\vec{\theta}} \Pi^{\vec{\varepsilon},\vec{\theta}}(z_0)\mathbf{1}, \tag{19}$$

and

$$p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) = \pi^{\vec{\varepsilon}} \Pi^{\vec{\varepsilon},\vec{\theta}}(z_{-n}) \Pi^{\vec{\varepsilon},\vec{\theta}}(z_{-n+1}) \cdots \Pi^{\vec{\varepsilon},\vec{\theta}}(z_0)\mathbf{1}. \tag{20}$$

Evidently $x_i^{\vec{\varepsilon},\vec{\theta}}$, $p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n})$ and $p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)$ all depend on the real vector $(\vec{\varepsilon},\vec{\theta}) \in \Omega_1 \times \Omega_2$. In what follows, we shall show that they can be "complexified". For any $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$, one checks that for $r_1 > 0$ small enough, the stationary vector $\pi^{\vec{\varepsilon}}$ is unique and analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$ as a function of $\vec{\varepsilon}$ (because it is the unique solution of $\pi^{\vec{\varepsilon}}\Pi^{\vec{\varepsilon}} = \pi^{\vec{\varepsilon}}, \quad \sum_y \pi_y^{\vec{\varepsilon}} = 1$).

Then through (18) and (17), $x_i^{\vec{\varepsilon},\vec{\theta}}$ can be analytically extended to $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$; furthermore, through (19) and (20), $p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n})$ and $p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)$ can be analytically extended to $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$. Ultimately $h_n^{\vec{\varepsilon},\vec{\theta}}(Z)$ (which is defined by (2) with real superscripts $(\vec{\varepsilon},\vec{\theta})$ on $p(z_{-n}^0)$ and $p(z_0|z_{-n}^{-1})$) can be analytically extended to $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$ as well.

The framework for the proofs of the main theorems can be outlined as follows:

(I) If necessary, we consecutively re-block the Z process to a $\hat{Z}$ process such that $\hat{Z}_i$ is of the form $Z_{j(i)}^{k(i)}$.

(II) We then show that there exists a complex neighborhood of a subset of W such that each complexified $f_{\hat{z}}^{\vec{\varepsilon},\vec{\theta}}$ is a contraction mapping, with respect to some metric, and moreover the complexified $\hat{x}_i^{\vec{\varepsilon},\vec{\theta}}(\hat{z}_{-\hat{n}}^i)$ stays within the neighborhood.

(III) It then follows that the complexified $x_i^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^i)$ and thus the complexified $p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n})$ exponentially forget their initial conditions.

(IV) This, together with bounding arguments, will further imply that the complexified $h_n^{\vec{\varepsilon},\vec{\theta}}(Z)$ uniformly converges to a complex analytic function, which is necessarily the complexified $h^{\vec{\varepsilon},\vec{\theta}}(Z)$, on a complex domain, and therefore $h^{\vec{\varepsilon},\vec{\theta}}(Z)$ is analytic.

The proofs of Theorem 1.1 and 1.8 are somewhat parallel, while the latter requiring extra attention to address the technical issues arising in proving joint analyticity. On the other hand, despite the fact that the proofs of Theorems 1.4 and Theorems 1.8 all fit in the same above-mentioned framework, there does not seem to be a natural way to unify them. Among numerous differences, the most essential one is the way we establish (II):

- To establish (II), we will use the fact that each real $f_z^{\vec{\varepsilon},\vec{\theta}}$ is a contraction on W°, with respect to the Hilbert metric. Then we use the "equivalence" between the Euclidean metric and the Hilbert metric, and equicontinuity in Condition (d(i)) to establish the contractiveness (with respect to the Euclidean metric) of the complexified $f_{\hat{z}}^{\vec{\varepsilon},\vec{\theta}}$ on a complex neighborhood of W.

- For Theorem 1.8, to establish (II), the complex Hilbert metric in Section 5 is employed to directly show the contractiveness, with respect to the complex Hilbert metric, of the complexified $f_z^{\vec{\varepsilon},\vec{\theta}}$ on a complex neighborhood of W°.

3 Proof of Theorem 1.4

The following lemma says that $f_z^{\vec{\varepsilon},\vec{\theta}}$ does not change much under a small complex perturbation of $(\vec{\varepsilon},\vec{\theta})$.

Lemma 3.1. *For any $\delta > 0$, there exist $r_1, r_2 > 0$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$, any $z \in \mathcal{Z}$ and any $x \in W$, we have*

$$\|f_z^{\vec{\varepsilon},\vec{\theta}}(x) - f_z^{\vec{\varepsilon}_0,\vec{\theta}_0}(x)\| \leq \delta$$

where $\|\cdot\|$ means the Euclidean norm.

Proof. We first note that

$$f_z^{\vec{\varepsilon},\vec{\theta}}(x) = \frac{x\Pi^{\vec{\varepsilon},\vec{\theta}}(z)}{x\Pi^{\vec{\varepsilon},\vec{\theta}}(z)\mathbf{1}} = \frac{x(\Pi^{\vec{\varepsilon},\vec{\theta}}(z)/q^{\vec{\theta}}(z|I))}{x(\Pi^{\vec{\varepsilon},\vec{\theta}}(z)/q^{\vec{\theta}}(z|I))\mathbf{1}}.$$

Assuming (7): The lemma immediately follows from (7) and Condition (d(i)).

Assuming (8): It follows from (8) that for sufficiently small $r_1, r_2 > 0$, there exists a compact subset $\Sigma \subset \mathcal{Z}$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$, any $z \notin \Sigma$ and any $x \in W$, $x(\Pi^{\vec{\varepsilon},\vec{\theta}}(z)/q^{\vec{\theta}}(z|I))\mathbf{1}$ is bounded away from 0. On the other hand, by the compactness of Σ, we deduce that there exist $r_1, r_2 > 0$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$, any $z \in \mathcal{Z}$ and any $x \in W$, $x(\Pi^{\vec{\varepsilon},\vec{\theta}}(z)/q^{\vec{\theta}}(z|I))\mathbf{1}$ is bounded away from 0. The lemma then follows from Condition (d(i)). □

Now, define

$$\tilde{W}_W(\delta) = \{\tilde{w} \in \tilde{W} : \|\tilde{w} - w\| < \delta \text{ for some } w \in W\}.$$

We need the following lemma.

Lemma 3.2. *Given any $(\vec{\varepsilon}_0, \vec{\theta}_0) \in \Omega_1 \times \Omega_2$, there exist $r_1, r_2, \delta > 0$, $0 < \rho_1 < 1$ and a positive integer n_0 such that, for all z_i^j with $j \geq i + n_0$ and all $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$, $f_{z_i^j}^{\vec{\varepsilon},\vec{\theta}}$ is a ρ_1-contraction mapping on $\tilde{W}_W(\delta)$ under the Euclidean metric.*

Proof. For any $z \in \mathcal{Z}$ and sufficiently small $r_1, r_2 > 0$, it follows from Remark 2.1 that for any $u, v \in W$, we have

$$d_H(u\Pi^{\vec{\varepsilon}_0,\vec{\theta}_0}(z), v\Pi^{\vec{\varepsilon}_0,\vec{\theta}_0}(z)) = d_H(u\Pi^{\vec{\varepsilon}_0}, v\Pi^{\vec{\varepsilon}_0}). \tag{21}$$

It then follows that for any z_i^j,

$$d_H(f_{z_i^j}^{\vec{\varepsilon}_0,\vec{\theta}_0}(u), f_{z_i^j}^{\vec{\varepsilon}_0,\vec{\theta}_0}(v)) \leq \tau(\Pi^{\vec{\varepsilon}_0})^{j-i} d_H(u\Pi^{\vec{\varepsilon}_0}, v\Pi^{\vec{\varepsilon}_0}),$$

where $\tau(\Pi^{\vec{\varepsilon}_0})$, the Birkhoff coefficient of $\Pi^{\vec{\varepsilon}_0}$ as defined in (14), is strictly less than 1. It then follows from Lemma 3.4 in [23] (the mixing assumption is satisfied due to (15) and Condition (a)) that there exists $C > 0$ such that for any z_i^j and any $u, v \in W$,

$$\|f_{z_i^j}^{\vec{\varepsilon}_0,\vec{\theta}_0}(u) - f_{z_i^j}^{\vec{\varepsilon}_0,\vec{\theta}_0}(v)\| \leq C\tau(\Pi^{\vec{\varepsilon}_0})^{j-i}\|u - v\|.$$

For n_0 sufficiently large, $\rho_0 := C\tau(\Pi^{\vec{\varepsilon}_0})^{n_0} < 1$. Thus, for $j \geq i + n_0$,

$$\|\nabla f_{z_i^j}^{\vec{\varepsilon}_0,\vec{\theta}_0}(w)\| \leq \rho_0 < 1$$

for all $w \in W$. From this and Condition (d(i)), it follows that there exists $r_1, r_2, \delta > 0$, $0 < \rho_1 < 1$ such that

$$\|\nabla f_{z_i^j}^{\vec{\varepsilon},\vec{\theta}}(w)\| \le \rho_1 < 1$$

for all $w \in \tilde{W}_W(\delta)$ and $(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$. The lemma then follows. □

The following lemma essentially follows from the framework in the proof of Theorem 1.1 in [11]. We briefly outline the proof for completeness.

We first introduce some notation. Let

$$\mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) \triangleq p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})/q^{\vec{\theta}}(z_0|I),$$

where I is the same as in Condition (d).

Lemma 3.3. *1. For any $\delta > 0$, there exist $r_1, r_2 > 0$ such that for any $(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$ and for all $z_{-n}^0 \in \mathcal{Z}^{n+1}$ and $-n-1 \le i \le -1$,*

$$x_i^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^i) \in \tilde{W}_W(\delta). \tag{22}$$

2. *There exist $r_1, r_2 > 0$ such that for all $z_{-n}^0 \in \mathcal{Z}^{n+1}$, $p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$.*

3. *There exist $r_1, r_2 > 0$, $0 < \rho_1 < 1$ and $L_1 > 0$ such that for any two $\mathcal{Z}$-valued sequences $\{a_{-n_1}^0\}$ and $\{b_{-n_2}^0\}$ with $a_{-n}^0 = b_{-n}^0$ and for all $(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$, we have*

$$|\mathring{p}^{\vec{\varepsilon},\vec{\theta}}(a_0|a_{-n_1}^{-1}) - \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(b_0|b_{-n_2}^{-1})| \le L_1 \rho_1^n. \tag{23}$$

Proof. 1. For a fixed $n_0 > 0$, we will consecutively reblock z_{-n}^{-1} into $\hat{z}_{-\hat{n}}^{-1}$ such that each $\hat{z}_i$ is of the form $z_{j(i)}^{k(i)}$, where $k(i) - j(i) + 1 = n_0$ (n_0 is determined below).

By (17), for any z_{-n}^0 and i, $x_i^{\vec{\varepsilon}_0,\vec{\theta}_0}$ (and thus $\hat{x}_i^{\vec{\varepsilon}_0,\vec{\theta}_0}$) belongs to W°. By Lemma 3.2, we can choose $r_1, r_2, \delta > 0$ sufficiently small, n_0 sufficiently large and $0 < \rho_1 < 1$ such that for all $(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$, $f_{\hat{z}}^{\vec{\varepsilon},\vec{\theta}}$ is a ρ_1-contraction on $\tilde{W}_W(\delta)$ under the Euclidean metric.

To prove (22), it is enough to prove the version of (22) with $x_i^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^i)$ replaced by $\hat{x}_i^{\vec{\varepsilon},\vec{\theta}}(\hat{z}_{-n}^i)$ (with perhaps smaller r_1, r_2).

To see this, note that by Lemma 3.1, for sufficiently small $r_1, r_2 > 0$, for all $\hat{z}$, $x \in W$, and $(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$,

$$\|f_{\hat{z}}^{\vec{\varepsilon},\vec{\theta}}(x) - f_{\hat{z}}^{\vec{\varepsilon}_0,\vec{\theta}_0}(x)\| \le \delta(1-\rho_1), \tag{24}$$

and for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$

$$\|\pi^{\vec{\varepsilon}} - \pi(\vec{\varepsilon}_0)\| \le \delta(1-\rho_1). \tag{25}$$

Thus,

$$\begin{aligned} \|\hat{x}_i^{\vec{\varepsilon},\vec{\theta}} - \hat{x}_i^{\vec{\varepsilon}_0,\vec{\theta}_0}\| &= \|f_{\hat{z}_i}^{\vec{\varepsilon},\vec{\theta}}(\hat{x}_{i-1}^{\vec{\varepsilon},\vec{\theta}}) - f_{\hat{z}_i}^{\vec{\varepsilon}_0,\vec{\theta}_0}(\hat{x}_{i-1}^{\vec{\varepsilon}_0,\vec{\theta}_0})\| \\ &\le \|f_{\hat{z}_i}^{\vec{\varepsilon},\vec{\theta}}(\hat{x}_{i-1}^{\vec{\varepsilon},\vec{\theta}}) - f_{\hat{z}_i}^{\vec{\varepsilon},\vec{\theta}}(\hat{x}_{i-1}^{\vec{\varepsilon}_0,\vec{\theta}_0})\| + \|f_{\hat{z}_i}^{\vec{\varepsilon},\vec{\theta}}(\hat{x}_{i-1}^{\vec{\varepsilon}_0,\vec{\theta}_0}) - f_{\hat{z}_i}^{\vec{\varepsilon}_0,\vec{\theta}_0}(\hat{x}_{i-1}^{\vec{\varepsilon}_0,\vec{\theta}_0})\|. \end{aligned} \tag{26}$$

Then by (24) and (25), and (26), we have

$$\|\hat{x}_i^{\vec{\varepsilon},\vec{\theta}} - \hat{x}_i^{\vec{\varepsilon}_0,\vec{\theta}_0}\| \leq \rho_1 \|\hat{x}_{i-1}^{\vec{\varepsilon},\vec{\theta}} - \hat{x}_{i-1}^{\vec{\varepsilon}_0,\vec{\theta}_0}\| + \delta(1-\rho_1).$$

So, for all i,

$$\|\hat{x}_i^{\vec{\varepsilon},\vec{\theta}} - \hat{x}_i^{\vec{\varepsilon}_0,\vec{\theta}_0}\| \leq \delta,$$

and thus for all i, we have $\hat{x}_i^{\vec{\varepsilon},\vec{\theta}} \in \tilde{W}_W(\delta)$, as desired.

2. It follows from Condition (d(i)) that for sufficiently small $r_1, r_2, \delta > 0$ and any $z \in \mathcal{Z}$, $f_z^{\vec{\varepsilon},\vec{\theta}}(x)$ is analytic with respect to $(\vec{\varepsilon}, \vec{\theta}, x) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2) \times \tilde{W}_W(\delta)$. It then follows from this fact and the iterative nature of $x_i^{\vec{\varepsilon},\vec{\theta}}$ (see (17)) and Part 1 that for sufficiently small $r_1, r_2 > 0$, each $x_i^{\vec{\varepsilon},\vec{\theta}}$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$. Part 2 then immediately follows from (19).

3. Applying the same reblocking as in Part 1, we write

$$\hat{x}_{i,\hat{a}}^{\vec{\varepsilon},\vec{\theta}} = \hat{x}_i^{\vec{\varepsilon},\vec{\theta}}(\hat{a}_{-\hat{n}_1}^i) = p^{\vec{\varepsilon},\vec{\theta}}(y_{k(i)} = \cdot\, |\hat{a}_{-\hat{n}_1}^i),$$

$$\hat{x}_{i,\hat{b}}^{\vec{\varepsilon},\vec{\theta}} = \hat{x}_i^{\vec{\varepsilon},\vec{\theta}}(\hat{b}_{-\hat{n}_2}^i) = p^{\vec{\varepsilon},\vec{\theta}}(y_{k(i)} = \cdot\, |\hat{b}_{-\hat{n}_2}^i).$$

Evidently we have

$$\hat{x}_{i+1,\hat{a}}^{\vec{\varepsilon},\vec{\theta}} = f_{\hat{a}_{i+1}}^{\vec{\varepsilon},\vec{\theta}}(\hat{x}_{i,\hat{a}}^{\vec{\varepsilon},\vec{\theta}}), \qquad \hat{x}_{i+1,\hat{b}}^{\vec{\varepsilon},\vec{\theta}} = f_{\hat{b}_{i+1}}^{\vec{\varepsilon},\vec{\theta}}(\hat{x}_{i,\hat{b}}^{\vec{\varepsilon},\vec{\theta}}).$$

Note that there exists a positive constant L_1' such that for all $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$,

$$\|\hat{x}_{-\hat{n},\hat{a}}^{\vec{\varepsilon},\vec{\theta}} - \hat{x}_{-\hat{n},\hat{b}}^{\vec{\varepsilon},\vec{\theta}}\| \leq L_1',$$

for all $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$, where $r_1, r_2 > 0$ are chosen sufficiently small. Since $f_{\hat{z}}^{\vec{\varepsilon},\vec{\theta}}$ is a ρ_1-contraction on $\tilde{W}_W(\delta)$, we have, by Part 1,

$$\|\hat{x}_{-1,\hat{a}}^{\vec{\varepsilon},\vec{\theta}} - \hat{x}_{-1,\hat{b}}^{\vec{\varepsilon},\vec{\theta}}\| \leq L_1' \rho_1^{\hat{n}-1}.$$

This implies that there exists $L_1 > 0$, independent of n_1, n_2, such that for all $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$,

$$\|\mathring{p}^{\vec{\varepsilon},\vec{\theta}}(a_0|a_{-n_1}^{-1}) - \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(b_0|b_{-n_2}^{-1})\| \leq L_1 \rho_1^n. \tag{27}$$

□

The proof of the following lemma is straightforward. Nonetheless, we sketch the proof for completeness.

Lemma 3.4. *Let n_1 and n_2 be positive integers and D a compact domain in $\mathbb{C}^{n_1}$. Let $f(\theta, z)$ be a jointly continuous function on $D \times \mathbb{R}^{n_2}$. Assume that*

$$\int_{\mathbb{R}^{n_2}} \sup_{\theta \in D} |f(\theta, z)| dz < \infty. \tag{28}$$

Then

1. $\int_{\mathbb{R}^{n_2}} f(\theta, z)dz$ *is continuous on* D.

2. If, for each $z \in \mathcal{Z}$, f *is analytic on* D, *then* $\int_{\mathbb{R}^{n_2}} f(\theta, z)dz$ *is analytic on* D.

Proof. Let Σ be a compact domain in $\mathbb{R}^{n_2}$. Let $\delta_i, i = 1, 2, \dots$ be a sequence of positive numbers converging to 0. Consider a sequence of partitions of Σ:

$$\Sigma = \cup_{i=1}^{m_n} \Delta_{n,i},$$

where $diam(\Delta_{n,i}) \leq \delta_n$ for all $1 \leq i \leq m_n$. Evidently, the corresponding Riemann sum

$$R_n = \sum_{i=1}^{m_n} f(\theta, z_i) vol(\Delta_{n,i})$$

(here $z_i \in \Delta_{n,i}$) is continuous in θ. Then

$$\int_\Sigma f(\theta, z)dz - R_n = \sum_{i=1}^{m_n} \int_{\Delta_{n,i}} (f(\theta, z) - f(\theta, z_i))dz.$$

By the compactness of D and Σ, we deduce that for any $\varepsilon_0 > 0$, there exists N_0 such that for all $n \geq N_0$ and all $i = 1, 2, \dots, m_n$

$$|f(\theta, z) - f(\theta, z_i)| \leq \varepsilon_0, \text{ for any } \theta \in D \text{ and any } z \in \Delta_{n,i},$$

which implies that for any $\varepsilon > 0$, there exists N_1 such that for all $n \geq N_1$ and all $\theta \in D$,

$$\left| \int_\Sigma f(\theta, z)dz - R_n \right| \leq \varepsilon.$$

In other words, R_n uniformly (in $\theta \in D$) converges to

$$\int_\Sigma f(\theta, z)dz,$$

and so $\int_\Sigma f(\theta, z)dz$ is continuous in $\theta \in D$.

Now, take any increasing sequence of compact sets Σ_i whose union is $\mathbb{R}^{n_2}$. By (28), $\int_{\Sigma_i} f(\theta, z)dz$ converges uniformly, in $\theta \in D$, to $\int_{\mathbb{R}^{n_2}} f(\theta, z)dz$, which is therefore continuous on D. This gives Part 1.

Part 2 follows in the same way with analyticity replacing continuity. □

We are now ready for the proof of Theorem 1.4.

Proof of Theorem 1.4. We first show that there exist $r_1, r_2 > 0$ such that for any n, $h_n^{\vec{\varepsilon}, \vec{\theta}}(Z)$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$.

For a fixed n, recall that

$$h_n^{\vec{\varepsilon}, \vec{\theta}}(Z) = -\int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon}, \vec{\theta}}(z_{-n}^0) \log p^{\vec{\varepsilon}, \vec{\theta}}(z_0 | z_{-n}^{-1}) dz_{-n}^0,$$

where

$$p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) = \sum_{y_{-n}^0} p^{\vec{\varepsilon}}(y_{-n}^0) \prod_{i=-n}^{0} q^{\vec{\theta}}(z_i|y_i) \tag{29}$$

and

$$p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) = x_{-1}^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^{-1})\Pi^{\vec{\varepsilon},\vec{\theta}}(z_0)\mathbf{1}.$$

Now, it follows from (29) and the fact that $\sum_{y_{-n}^0} p^{\vec{\varepsilon}}(y_{-n}^0)$ is continuous and therefore bounded as a function of $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$ that there is a constant $K > 0$ such that

$$\sup_{(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)| \leq K \prod_{i=-n}^{0} \sup_{(y,\vec{\theta}) \in \mathcal{Y} \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |q^{\vec{\theta}}(z_i|y)|. \tag{30}$$

It also follows from Part 1 of Lemma 3.3 and (Condition (7) or (8)) that for sufficiently small $r_1, r_2 > 0$, there exist $C_1, C_2 > 0$ such that for any z_{-n}^0,

$$C_1 \leq |\mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})| \leq C_2, \tag{31}$$

which implies that for some $C_3 > 0$,

$$|\log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})| \leq C_3. \tag{32}$$

We then deduce that for any $(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$,

$$\begin{aligned}
&\int_{\mathcal{Z}^{n+1}} \sup_{(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} \left| p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) \right| dz_{-n}^0 \\
&\quad = \int_{\mathcal{Z}^{n+1}} \sup_{(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} \left| p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log q^{\vec{\theta}}(z_0|I) + p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) \right| dz_{-n}^0 \\
&\quad \leq \int_{\mathcal{Z}^{n+1}} \sup_{\vec{\theta} \in \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} K \left| \prod_{i=-n}^{0} q^{\vec{\theta}}(z_i|y_i) \right| \left| \log q^{\vec{\theta}}(z_0|I) \right| dz_{-n}^0 \\
&\quad + \int_{\mathcal{Z}^{n+1}} \sup_{(\vec{\varepsilon},\vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)| \quad |\log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})| dz_{-n}^0 < \infty.
\end{aligned} \tag{33}$$

(for the first term we have used (5(i)) and (6); for the second term, we have used (5(i)), (30) and (32)). By lemma 3.4 (Part 2),

$$h_n^{\vec{\varepsilon},\vec{\theta}}(Z_0|Z_{-n}^{-1}) = \int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) dz_{-n}^0,$$

is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$.

Now, to prove the theorem, we only need to prove that there exist $r_1, r_2 > 0$ such that $h_n^{\vec{\varepsilon},\vec{\theta}}(Z)$ uniformly converges on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ as $n \to \infty$. First, we observe that

$$\begin{aligned}
|h_{n+1}^{\vec{\varepsilon},\vec{\theta}}(Z) - h_n^{\vec{\varepsilon},\vec{\theta}}(Z)| &= \left| \int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0) \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) dz_{-n-1}^0 - \int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) dz_{-n}^0 \right| \\
&= \left| \int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0) \log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) dz_{-n-1}^0 - \int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) dz_{-n}^0 \right| \\
&= \left| \int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0) (\log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})) dz_{-n-1}^0 \right|.
\end{aligned}$$

Fix $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$. Then, by (23), (31), (32), we have, for some $0 < \rho_1 < 1$, $L_1', L_1 > 0$,

$$\begin{aligned}
|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))| &\le L_1' \left| p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})) \right| \\
&\le L_1' |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)| L_1 \rho_1^n.
\end{aligned}$$

Notice that for any given $\delta > 0$, there exist $r_1, r_2 > 0$ such that for all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$,

$$\|\pi_{y_{-n}}^{\vec{\varepsilon}}\| \le (1+\delta)\pi_{y_{-n}}^{\vec{\varepsilon}_0}, \qquad \|\pi_{y_i y_{i+1}}^{\vec{\varepsilon}}\| \le (1+\delta)\pi_{y_i y_{i+1}}^{\vec{\varepsilon}_0},$$

and for any $y \in \mathcal{Y}$ and all $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

$$\int_{\mathcal{Z}} |q^{\vec{\theta}}(z|y)| dz \le (1+\delta) \int_{\mathcal{Z}} q^{\vec{\theta}_0}(z|y) dz = 1 + \delta,$$

(here we have used the fact that $\int_{\mathcal{Z}} |q^{\vec{\theta}}(z|y)| dz$ is a continuous function of $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$; this follows from Lemma 3.4 (Part 1)). It then follows from (30) that

$$\int_{\mathcal{Z}^{n+1}} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^{-1})| dz_{-n-1}^{-1} \le (1+\delta)^{2(n+2)}.$$

By choosing $\delta > 0$ sufficiently small, we can combine all the relevant inequalities above to obtain some $L > 0$ and some $0 < \rho < 1$ such that for all $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

$$|h_{n+1}^{\vec{\varepsilon},\vec{\theta}}(Z) - h_n^{\vec{\varepsilon},\vec{\theta}}(Z)| \le \int_{\mathcal{Z}^{n+2}} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log \mathring{p}^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))| dz_{-n-1}^0 \le L\rho^n,$$

which implies the uniform convergence of $h_n^{\vec{\varepsilon},\vec{\theta}}(Z)$ on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ as n tends to infinity, and thus the analyticity of $h^{\vec{\varepsilon},\vec{\theta}}(Z)$ around $(\vec{\varepsilon}_0, \vec{\theta}_0)$. □

4 Failure of Analyticity

With the following example, we show that for Gaussian channels, $h(Z)$ need not be analytic even as a function of the channel parameters alone, when the largest σ_i is not unique.

Example 4.1. Consider an additive Gaussian channel parameterized as in (10) with the binary input alphabet $\mathcal{Y} = \{1, 2\}$. Assume that the input Y is an i.i.d. process with

$$P(Y_1 = 1) = P(Y_1 = 2) = 1/2;$$

and assume that

$$q(z|1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(z+1)^2/\sigma_1^2}, \quad q(z|2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(z-1)^2/\sigma_2^2}.$$

We then have

$$p(z) = P(Y=1)q(z|1) + P(Y=2)q(z|2) = \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma_1} e^{-(z+1)^2/\sigma_1^2} + \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma_2} e^{-(z-1)^2/\sigma_2^2}.$$

We claim that for any fixed $\sigma > 0$, analyticity of $h(Z)$ as a function of (σ_1, σ_2) fails at $(\sigma_1, \sigma_2) = (\sigma, \sigma)$. To see this, we fix $\sigma_1 = \sigma$, and we show that $h(Z)$ is not analytic with respect to σ_2 at $\sigma_2 = \sigma$. Note that for any real σ_2,

$$\begin{aligned} h(Z) &= -\int_{-\infty}^{\infty} p(z) \log p(z) dz \\ &= -\int_0^{\infty} p(z) \left(\log \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma_2} e^{-(z-1)^2/\sigma_2^2} \right) dz - \int_{-\infty}^0 p(z) \left(\log \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma} e^{-(z-1)^2/\sigma^2} \right) dz \quad (34) \\ &\quad - \int_0^{\infty} p(z) \log\left(1 + \Phi_z(\sigma_2)\right) dz - \int_{-\infty}^0 p(z) \log\left(1 + \Phi_z^{-1}(\sigma_2)\right) dz, \quad (35) \end{aligned}$$

where

$$\Phi_z(\sigma_2) \triangleq \frac{\sigma_2}{\sigma} e^{(z-1)^2/\sigma_2^2 - (z+1)^2/\sigma^2} \text{ and } \Phi_z^{-1}(\sigma_2) \triangleq \frac{1}{\Phi_z(\sigma_2)}. \quad (36)$$

We note that (34) is analytic as a function of σ_2 at $\sigma_2 = \sigma$. To see this, observe that the first term of (34) can be further computed as

$$-\log(2\sqrt{2\pi}\sigma_2) \int_0^{\infty} p(z) dz - \int_0^{\infty} \frac{1}{2\sqrt{2\pi}\sigma\sigma_2^2} (z-1)^2 e^{-(z+1)^2/\sigma^2} dz - \int_0^{\infty} \frac{1}{2\sqrt{2\pi}\sigma_2^3} (z-1)^2 e^{-(z-1)^2/\sigma_2^2} dz,$$

which is analytic at $\sigma_2 = \sigma$ since each of the three terms above is analytic at $\sigma_2 = \sigma$ (for the second or third term, regard σ_2 as a complex variable and use the exponentially-decaying tail of the integrand). With a similar argument applied to the second term of (34), we can then establish the analyticity of (34). So, to prove $h(Z)$ is not analytic at $\sigma_2 = \sigma$, it suffices to show that (35) is not analytic at $\sigma_2 = \sigma$.

Suppose, by way of contradiction, that (35) is analytic at σ, or equivalently, the following function of ω

$$\begin{aligned} &\int_0^{\infty} \left(\frac{1}{2}\frac{\sigma^{-1}}{\sqrt{2\pi}} e^{-(z+1)^2\sigma^{-2}} + \frac{1}{2}\frac{\omega}{\sqrt{2\pi}} e^{-(z-1)^2\omega^2} \right) \log\left(1 + \Phi_z(1/\omega)\right) dz \\ &+ \int_{-\infty}^0 \left(\frac{1}{2}\frac{\sigma^{-1}}{\sqrt{2\pi}} e^{-(z+1)^2\sigma^{-2}} + \frac{1}{2}\frac{\omega}{\sqrt{2\pi}} e^{-(z-1)^2\omega^2} \right) \log\left(1 + \Phi_z^{-1}(1/\omega)\right) dz \quad (37) \end{aligned}$$

is analytic at $\sigma^{-1} \in \overline{\mathbb{C}_{\sigma^{-1}}(r)}$ (the closure of the r-neighborhood of σ^{-1} in $\mathbb{C}$) for some $r > 0$, where, recalling from (36),

$$\Phi_z(1/\omega) = \frac{\sigma^{-1}}{\omega} e^{(z-1)^2\omega^2-(z+1)^2\sigma^{-2}}.$$

Then, by uniqueness, the analytic extension of (37) to $\overline{\mathbb{C}_{\sigma^{-1}}(r)}$ would agree with any analytic extension along the circle $\{\sigma^{-1} + re^{i\alpha} : \alpha \in [-\pi/2, 3\pi/2]\}$ (from $\alpha = -\pi/2$ to $\alpha = 3\pi/2$). Such an analytic extension is obtained by regarding ω as a complex variable on the circle (this is a valid analytic extension by virtue of the exponentially-decaying tails of the integrands in (37)). Here, we remark that for any $r > 0$ and α, there are at most two "singular" z (note that the following inequality boils down to a system of two quadratic equations in z) such that

$$\Phi_z(1/(\sigma^{-1} + re^{i\alpha})) = -1,$$

which means $\log(1 + \Phi_z(1/(\sigma^{-1} + re^{i\alpha})))$ or $\log(1 + \Phi_z^{-1}(1/(\sigma^{-1} + re^{i\alpha})))$ would "blow up" at such z. However, an easy bounding argument (roughly speaking, the two "blowing up" terms will only do so "slowly") yields that during the analytic extension, (37) is still well-defined with the presence of such singular z, and so the above analytic extension is indeed valid.

Next, we will find a contradiction by showing that the analytic extension of (37) disagrees at $\alpha = -\pi/2$ and $\alpha = 3\pi/2$. Setting $\omega = \sigma^{-1} + re^{i\alpha}$, we then have

$$\frac{\sigma^{-1}}{w} = \frac{\sigma^{-1}}{\sigma^{-1} + r\cos\alpha + ir\sin\alpha} = \frac{\sigma^{-2} + \sigma^{-1}r\cos\alpha - i\sigma^{-1}r\sin\alpha}{\sigma^{-2} + 2\sigma^{-1}r\cos\alpha + r^2} \triangleq e^{a(r,\alpha)+ib(r,\alpha)},$$

where one can easily check that

$$a(r, \alpha) = O(r), \quad b(r, \alpha) = O(r), \quad \frac{\partial b(r, \alpha)}{\partial \alpha} = O(r).$$

Then, some straightforward computations yield that

$$\Phi_z(1/\omega) = e^{A(z,r,\alpha)}e^{iB(z,r,\alpha)},$$

where

$$A(z, r, \alpha) \triangleq 2(z - 1)^2\sigma^{-1}r\cos\alpha + (z - 1)^2r^2\cos 2\alpha - 4z\sigma^{-2} + a(r, \alpha),$$

and

$$B(z, r, \alpha) \triangleq 2(z - 1)^2\sigma^{-1}r\sin\alpha + (z - 1)^2r^2\sin 2\alpha + b(r, \alpha).$$

Now, for some small yet fixed $\varepsilon > 0$, choose $N > 0$ large enough and then $r > 0$ small enough such that

(I) for all $0 \le z \le N$ and all $\alpha \in [-\pi/2, 3\pi/2]$, $B(z, r, \alpha) \in (-\pi, \pi)$;

(II) for all $z \ge N$ and all $\alpha \in [-\pi/2 + \varepsilon, \pi/2 - \varepsilon] \cup [\pi/2 + \varepsilon, 3\pi/2 - \varepsilon]$,

$$4z\sigma^{-2} \gg |a(r, \alpha)|, \quad |2(z - 1)^2\sigma^{-1}r\cos\alpha| \gg |(z - 1)^2r^2\cos 2\alpha + a(r, \alpha)|$$

and

$$\left|\frac{\partial}{\partial\alpha}(2(z - 1)^2\sigma^{-1}r\sin\alpha)\right| \gg \left|\frac{\partial}{\partial\alpha}((z - 1)^2r^2\sin 2\alpha + b(r, \alpha))\right|.$$

Note that for all $0 \leq z \leq N$, by (I), $\Phi_z(1/\omega)$ will not go around -1 (in any direction) for one complete round as α increases from $-\pi/2$ to $3\pi/2$. Next, we consider the case when $z \geq N$. Notice that, by (II), for any fixed $z \geq N$, as α increases from $-\pi/2 + \varepsilon$ to $\pi/2 - \varepsilon$, $B(z, r, \alpha)$ increases as well. If, for some $z \geq N$ and $\alpha_0 \in [-\pi/2, -\pi/2 + \varepsilon] \cup [\pi/2 - \varepsilon, \pi/2]$,

$$A(z, r, \alpha_0) > 0,$$

it then follows from (II) that there exists $\ell = \ell(\varepsilon) > 0$ such that $\ell \to \infty$ as $\varepsilon \to 0$ and for the same z and any $\alpha \in [-\ell\varepsilon, \ell\varepsilon]$,

$$A(z, r, \alpha) > 0.$$

On the other hand, it follows from (II) that for any $z \geq N$ and for any $\alpha \in [\pi/2+\varepsilon, 3\pi/2-\varepsilon]$,

$$A(z, r, \alpha) < 0;$$

straightforward computations also yield that for any $z \geq N$ and for any $\alpha \in [\pi/2, \pi/2+\varepsilon] \cup [3\pi/2 - \varepsilon, 3\pi/2]$,

$$A(z, r, \alpha) < 0.$$

It then follows that $\Phi_z(1/\omega)$ will not go around -1 (in any direction) for one complete round as α increases from $\pi/2$ to $3\pi/2$.

We are now ready to conclude that as α increases from $-\pi/2$ to $3\pi/2$, for any $z \geq N$ with $\Phi_z(1/(\sigma^{-1} + re^{i\alpha})) \neq -1$, $\Phi_z(1/\omega)$ will go around -1 anti-clockwise $k(z)$ times, where $k(z)$ is a non-negative integer; meanwhile, one checks that when z is large enough, $k(z)$ is strictly positive. The idea can be roughly described as follows. Consider the "trajectory" of $\Phi_z(1/\omega)$ as α increases from $-\pi/2$ to $3\pi/2$. Obviously, $A(z, r, \alpha) > 0$ means the magnitude of the corresponding "location" is strictly bigger than 1; $B(z, r, \alpha) > 0$ means at the corresponding "location", $\Phi_z(1/\omega)$ is going anti-clockwise. The above argument shows that given sufficiently small ε (and thus ℓ sufficiently large), for all the time when the "location" is at least 1 away from the origin, "more often" $\Phi_z(1/\omega)$ goes around -1 anti-clockwise (for any $\alpha \in [-\pi/2, -\pi/2+\varepsilon] \cup [\pi/2-\varepsilon, \pi/2]$, $\Phi_z(1/\omega)$ may go around -1 clockwise, whereas for all $\alpha \in [-\ell\varepsilon, \ell\varepsilon]$, $\Phi_z(1/\omega)$ must go around -1 anti-clockwise).

So, for any analytic extension along the circle $\{\sigma^{-1} + re^{i\alpha} : \alpha \in [-\pi/2, 3\pi/2]\}$ (from $\alpha = -\pi/2$ to $\alpha = 3\pi/2$), we have proven that for any $z \geq 0$ with $\Phi_z(1/(\sigma^{-1} + re^{i\alpha})) \neq -1$,

$$\Im\left(\lim_{\alpha \to (3\pi/2)-} \log\left(1 + \Phi_z(1/(\sigma^{-1} + re^{i\alpha}))\right)\right) = \Im\left(\lim_{\alpha \to (-\pi/2)+} \log\left(1 + \Phi_z(1/(\sigma^{-1} + re^{i\alpha}))\right)\right) + 2k(z)\pi i,$$

where $k(z)$ is a non-negative integer for all z and a strictly positive integer for all sufficiently large z. Using a similar argument, we can also prove that for any $z \leq 0$ with $\Phi_z^{-1}(1/(\sigma^{-1} + re^{i\alpha})) \neq -1$,

$$\Im\left(\lim_{\alpha \to (3\pi/2)-} \log\left(1 + \Phi_z^{-1}(1/(\sigma^{-1} + re^{i\alpha}))\right)\right) = \Im\left(\lim_{\alpha \to (-\pi/2)+} \log\left(1 + \Phi_z^{-1}(1/(\sigma^{-1} + re^{i\alpha}))\right)\right) + 2k(z)\pi i,$$

where $k(z)$ is a non-negative integer for all z and a strictly positive integer for all sufficiently large $|z|$. This, however, implies that for the above analytic extension, (37) disagrees at $\alpha = -\pi/2$ and $\alpha = 3\pi/2$, which is a contradiction.

Remark 4.2. The previous example shows that for fixed $\sigma_1 = \sigma$, $h(Z)$ is not analytic with respect to σ_2 at $\sigma_2 = \sigma$. On the other hand, when σ_1, σ_2 are both equal to $\sigma > 0$ and vary together keeping $\sigma_1 = \sigma_2$, $h(Z)$ is in fact analytic with respect to σ. To see this, note that when $\sigma_1 = \sigma_2 = \sigma > 0$, we have

$$p(z) = \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\left(e^{-(z+1)^2/\sigma^2} + e^{-(z-1)^2/\sigma^2}\right),$$

and furthermore,

$$\begin{aligned} h(Z) &= -\int_{-\infty}^{\infty} p(z)\log p(z) \\ &= -\log(2\sqrt{2\pi}\sigma) + \int_{-\infty}^{\infty} p(z)(z+1)^2/\sigma^2 dz - \int_{-\infty}^{\infty} p(z)\log(1+e^{-4z/\sigma^2})dz \\ &= -\log(2\sqrt{2\pi}\sigma) + \int_{-\infty}^{\infty} p(z)(z+1)^2/\sigma^2 dz - \sigma^2\int_{-\infty}^{\infty} p(\sigma^2 z)\log(1+e^{-4z})dz. \end{aligned} \tag{38}$$

Note that it is easy to check that all the three terms in (38) are analytic with respect to σ, which immediately implies that $h(Z)$ is analytic with respect to σ.

5 A Complex Hilbert Metric

In this section, we briefly review the complex Hilbert metric that will be of critical use in the proofs of Theorem 1.1 and 1.8. Several useful properties of this metric will be stated/proved as well.

Recall that $\tilde{W}$ denote the complex version of W,

$$\tilde{W} = \{w = (w_1, w_2, \cdots, w_l) \in \mathbb{C}^l : \sum_i w_i = 1\}.$$

Let

$$\tilde{W}^+ = \{v \in \tilde{W} : \Re(v_i/v_j) > 0 \text{ for all } i, j\}.$$

For $v, w \in \tilde{W}^+$, define

$$\tilde{d}_H(v, w) = \max_{i,j}\left|\log\left(\frac{w_i/w_j}{v_i/v_j}\right)\right|, \tag{39}$$

where log is taken as the principal branch of the complex $\log(\cdot)$ function (i.e., the branch whose branch cut is the negative real axis). Since the principal branch of log is additive on the right-half plane, $\tilde{d}_H$ is a metric on $\tilde{W}^+$, which we call a *complex Hilbert metric* (for alternative complex Hilbert metrics, see [37] and [6]).

Let M denote the set of all $l \times l$ stochastic matrices, i.e.,

$$M = \{\Pi = (\pi_{ij}) \in \mathbb{R}^{l\times l} : \pi_{ij} \geq 0, \quad \sum_{j=1}^{l} \pi_{ij} = 1\},$$

and let $\tilde{M}$ denote the complex version of M, defined as

$$\tilde{M} = \{\Pi = (\pi_{ij}) \in \mathbb{C}^{l\times l} : \sum_{j=1}^{l} \pi_{ij} = 1, \text{ for all } i\}.$$

For a given positive $\Pi \in M$ and a small $\delta_1 > 0$, let $\tilde{M}_\Pi(\delta_1)$ denote the δ_1-neighborhood, under the Euclidean metric, around Π within $\tilde{M}$. For an element $\tilde{\Pi} \in \tilde{M}_\Pi(\delta_1)$, similar to (13), $\tilde{\Pi}$ will induce a mapping $f_{\tilde{\Pi}}$ on $\tilde{W}$. For a small $\delta_2 > 0$, let $\tilde{W}_{W^\circ,H}(\delta_2)$ denote the δ_2-neighborhood of W° within $\tilde{W}^+$ under the complex Hilbert metric, i.e.,

$$\tilde{W}_{W^\circ,H}(\delta_2) = \{v = (v_1, v_2, \cdots, v_l) \in \tilde{W}^+ : \tilde{d}_H(v, u) \leq \delta_2, \text{ for some } u \in W^\circ\}.$$

The main result of [15] states:

Theorem 5.1. *Let Π be a positive matrix in M. For sufficiently small $\delta_1, \delta_2 > 0$, there exists $0 < \rho_1 < 0$ such that for any $\tilde{\Pi} \in \tilde{M}_\Pi(\delta_1)$, $f_{\tilde{\Pi}}$ is a ρ_1-contraction mapping on $\tilde{W}_{W^\circ,H}(\delta_2)$ under the complex Hilbert metric in (39).*

For $\delta > 0$, let $\mathbb{C}_{\mathbb{R}^+}[\delta]$ denote the "δ-cone" of $\mathbb{R}^+$ within $\mathbb{C}$, i.e.,

$$\mathbb{C}_{\mathbb{R}^+}[\delta] = \{x + yi \in \mathbb{C} : x > 0, -\delta x \leq y \leq \delta x\}.$$

The following Lemma can be easily checked.

Lemma 5.2. *For sufficiently small $\delta_1 > 0$, there exists a positive constant L_1 such that for any $\alpha, \beta \in \mathbb{C}_{\mathbb{R}^+}[\delta_1]$*

$$|\log \alpha - \log \beta| \leq L_1 \max\left(\frac{|\alpha - \beta|}{|\alpha|}, \frac{|\alpha - \beta|}{|\beta|}\right).$$

The following lemma essentially follows from the proof of Part 2 of Lemma 2.3 in [15] (in particular, its Part 1 is just a rephrased version of Part 2 of that lemma), allows us to connect the complex Hilbert metric and the Euclidean metric. We give a proof for completeness.

Lemma 5.3. *1. For any $\delta > 0$, there exists $\xi > 0$ such that for any $\tilde{x} \in \tilde{W}^+$, $x \in W^\circ$ with $\tilde{d}_H(\tilde{x}, x) \leq \xi$, we have $\tilde{x}_i \in \tilde{W}_{W^\circ,H}(\delta)$ for all i.*

2. For any $\zeta > 0$, there exists a constant $C > 0$ such that for any $\tilde{x}, \tilde{y} \in \tilde{W}^+$ with $\|\tilde{x} - x\|, \|\tilde{y} - y\| \leq \zeta$ for some $x, y \in W^\circ$, we have

$$\|\tilde{x} - \tilde{y}\| \leq C\tilde{d}_H(\tilde{x}, \tilde{y}).$$

Proof. We only prove Part 2. Let $\xi = \tilde{d}_H(\tilde{x}, \tilde{y})$. Then we have for all i, j,

$$\left|\log\left(\frac{\tilde{x}_i/\tilde{y}_i}{\tilde{x}_j/\tilde{y}_j}\right)\right| \leq \xi.$$

There exists $C_1 > 0$ such that for ξ sufficiently small, and for all i, j, $\left|\frac{\tilde{x}_i/\tilde{y}_i}{\tilde{x}_j/\tilde{y}_j} - 1\right| \leq C_1\xi$. Let $\alpha_j = \tilde{x}_j/\tilde{y}_j$. Then for all i, j,

$$|\tilde{x}_i - \alpha_j \tilde{y}_i| \leq C_1\xi|\alpha_j||\tilde{y}_i|,$$

and so

$$|1-\alpha_j| = \left|\sum_{i=1}^{n}(\tilde{x}_i - \alpha_j \tilde{y}_i)\right| \leq \sum_{i=1}^{n} |\tilde{x}_i - \alpha_j \tilde{y}_i| \leq C_1 \xi |\alpha_j| \sum_{i=1}^{n} |\tilde{y}_i| = C_1(1+B\zeta)\xi|\alpha_j|.$$

It follows that $|\tilde{x}_j - \tilde{y}_j| \leq C_1(1+B\zeta)\xi|\tilde{x}_j| \leq C_1(1+B\zeta)\xi(x_j+\zeta)$, which implies Part 2, if ξ is sufficiently small. □

6 Proof of Theorem 1.1

The following lemma is an analog of Lemma 3.1.

Lemma 6.1. *For any $\delta > 0$, there exists $r_1 > 0$ such that for any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$, any $z \in \mathcal{Z}$ and any $x \in W$, we have*

$$\tilde{d}_H(f_z^{\vec{\varepsilon}}(x), f_z^{\vec{\varepsilon}_0}(x)) \leq \delta.$$

Proof. Since all $\pi_{ij}(\vec{\varepsilon}_0)$ are strictly positive, for any $\delta_1 > 0$, there exists $r_1 > 0$ such that for all i, j and all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$,

$$\frac{|\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0)|}{\pi_{ij}(\vec{\varepsilon}_0)} \leq \delta_1.$$

Now, for any $x = (x_1, x_2, \cdots, x_l) \in W$, any j and any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$, we have

$$\left|\frac{\sum_{i=1}^{l} x_i(\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0))}{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0)}\right| = \left|\frac{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0)(\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0))/\pi_{ij}(\vec{\varepsilon}_0)}{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0)}\right| \leq \delta_1.$$

Thus, for any $\delta_2 > 0$, choosing δ_1 sufficiently small, we have

$$\left|\log \frac{\sum_{i=1}^{l} x_i \pi_{ij}^{\vec{\varepsilon}}}{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0)}\right| = \left|\log\left(1 + \frac{\sum_{i=1}^{l} x_i(\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0))}{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0)}\right)\right| \leq \delta_2.$$

Notice that

$$\tilde{d}_H(f_z^{\vec{\varepsilon}}(x), f_z^{\vec{\varepsilon}_0}(x)) = \max_{j,k}\left|\log \frac{\sum_{i=1}^{l} x_i \pi_{ij}^{\vec{\varepsilon}} q(z|j)}{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0) q(z|j)} - \log \frac{\sum_{i=1}^{l} x_i \pi_{ik}^{\vec{\varepsilon}} q(z|k)}{\sum_{i=1}^{l} x_i \pi_{ik}(\vec{\varepsilon}_0) q(z|k)}\right|$$

$$= \max_{j,k}\left|\log \frac{\sum_{i=1}^{l} x_i \pi_{ij}^{\vec{\varepsilon}}}{\sum_{i=1}^{l} x_i \pi_{ij}(\vec{\varepsilon}_0)} - \log \frac{\sum_{i=1}^{l} x_i \pi_{ik}^{\vec{\varepsilon}}}{\sum_{i=1}^{l} x_i \pi_{ik}(\vec{\varepsilon}_0)}\right|.$$

It then follows that for any $\delta > 0$, there exists $r_1 > 0$ such that for any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$ and any $x \in W$, we have

$$\tilde{d}_H(f_z^{\vec{\varepsilon}}(x), f_z^{\vec{\varepsilon}_0}(x)) \leq \delta.$$

□

The following lemma, roughly speaking, says that when we perturb $\vec{\varepsilon}_0$ "a bit" to $\vec{\varepsilon}$, $f_z^{\vec{\varepsilon}}$ is still a contraction mapping on a complex neighborhood of W°, and the contraction coefficient is uniform over all the values of z. More precisely, recalling $\tilde{W}_{W^\circ,H}(\delta)$ denote the δ-neighborhood of W° of $\tilde{W}$ under the complex Hilbert metric, we have

Lemma 6.2. *For sufficiently small $r_1, \delta > 0$, there exists $0 < \rho_1 < 1$ such that for any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$ and any $z \in \mathcal{Z}$, $f_z^{\vec{\varepsilon}}$ is a ρ_1-contraction mapping on $\tilde{W}_{W^\circ,H}(\delta)$ under the complex Hilbert metric in (39).*

Proof. It can be easily checked that there exist some $r_1, \delta > 0$ such that for any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$, any $u, v \in \tilde{W}_{W^\circ,H}(\delta)$ and any $z \in \mathcal{Z}$,

$$\tilde{d}_H(u\Pi^{\vec{\varepsilon}}(z), v\Pi^{\vec{\varepsilon}}(z)) = \tilde{d}_H(u\Pi^{\vec{\varepsilon}}, v\Pi^{\vec{\varepsilon}}). \tag{40}$$

The lemma then immediately follows from Theorem 5.1. □

The following lemma is also needed.

Lemma 6.3. *1. For any $\delta > 0$, there exist $r_1 > 0$ such that for any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$ and for any $z_{-n}^0 \in \mathcal{Z}^{n+1}$ and $-n-1 \le i \le -1$,*

$$x_i^{\vec{\varepsilon}}(z_{-n}^i) \in \tilde{W}_{W^\circ,H}(\delta), \tag{41}$$

and

$$p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) \in \mathbb{C}_{\mathbb{R}^+}[\delta]. \tag{42}$$

2. There exist $r_1 > 0$ such that for all $z_{-n}^0 \in \mathcal{Z}^{n+1}$, $p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})$ is analytic on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$.

3. For sufficiently small $r_1 > 0$, there exist $0 < \rho_1 < 1$ and a positive constant L_1 such that for any two $\mathcal{Z}$-valued sequences $\{a_{-n_1}^0\}$ and $\{b_{-n_2}^0\}$ with $a_{-n}^0 = b_{-n}^0$ and for all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$, we have

$$|p^{\vec{\varepsilon}}(a_0|a_{-n_1}^{-1}) - p^{\vec{\varepsilon}}(b_0|b_{-n_2}^{-1})| \le L_1\rho_1^n \max_{y' \in \mathcal{Y}} q(a_0|y').$$

Proof. 1. By Lemma 6.2, we can choose $r_1, \delta > 0$ sufficiently small such that there exists $0 < \rho_1 < 1$ such that for all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$, $f_z^{\vec{\varepsilon}}$ is a ρ_1-contraction mapping on $\tilde{W}_{W^\circ,H}(\delta)$ under the complex Hilbert metric.

Now, choose $r_1 > 0$ so small (the existence of r_1 is guaranteed by Lemma 6.1) such that for any $z \in \mathcal{Z}$, for all $x \in W$, all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$

$$\tilde{d}_H(f_z^{\vec{\varepsilon}}(x), f_z^{\vec{\varepsilon}_0}(x)) \le \delta(1-\rho_1), \tag{43}$$

and for all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$,

$$\tilde{d}_H(\pi^{\vec{\varepsilon}}, \pi(\vec{\varepsilon}_0)) \le \delta(1-\rho_1). \tag{44}$$

We then deduce that

$$\begin{aligned} \tilde{d}_H(x_{i+1}^{\vec{\varepsilon}}, x_{i+1}^{\vec{\varepsilon}_0}) &= \tilde{d}_H(f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}}), f_{z_{i+1}}^{\vec{\varepsilon}_0}(x_i^{\vec{\varepsilon}_0})) \\ &\le \tilde{d}_H(f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}}), f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}_0})) + \tilde{d}_H(f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}_0}), f_{z_{i+1}}^{\vec{\varepsilon}_0}(x_i^{\vec{\varepsilon}_0})). \end{aligned} \tag{45}$$

Then, by (43), (44) and (45), for $i > -n-1$, we have

$$\tilde{d}_H(x_{i+1}^{\vec{\varepsilon}}, x_{i+1}^{\vec{\varepsilon}_0}) \leq \rho \tilde{d}_H(x_i^{\vec{\varepsilon}}, x_i^{\vec{\varepsilon}_0}) + \delta(1-\rho_1).$$

So, for all i,

$$\tilde{d}_H(x_{i+1}^{\vec{\varepsilon}}, x_{i+1}^{\vec{\varepsilon}_0}) \leq \delta,$$

and thus for all i, we have $x_{i+1}^{\vec{\varepsilon}} \in \tilde{W}_{W^\circ,H}(\delta)$, as desired. This, together with (19) and Lemma 5.3 (Part 2), implies (42).

2. It follows from Part 1 of Lemma 5.3 that for sufficiently small $r_1, \delta > 0$ and any $z \in \mathcal{Z}$, $f_z^{\vec{\varepsilon}}(x)$ is analytic with respect to $(\vec{\varepsilon}, x) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \tilde{W}_{W^\circ,H}(\delta)$. It follows from this fact, the iterative nature of $x_i^{\vec{\varepsilon},\vec{\theta}}$ (see (17)) and Part 1 that for sufficiently small $r_1 > 0$, each $x_i^{\vec{\varepsilon}}$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$. Part 2 then immediately follows from (19).

3. For all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$, we write

$$x_{i,a}^{\vec{\varepsilon}} = x_i^{\vec{\varepsilon}}(a_{-n_1}^i) = p^{\vec{\varepsilon}}(y_i = \cdot\, |a_{-n_1}^i),$$

$$x_{i,b}^{\vec{\varepsilon}} = x_i^{\vec{\varepsilon}}(b_{-n_2}^i) = p^{\vec{\varepsilon}}(y_i = \cdot\, |b_{-n_2}^i).$$

Apparently we have

$$x_{i+1,a}^{\vec{\varepsilon}} = f_{a_{i+1}}^{\vec{\varepsilon}}(x_{i,a}^{\vec{\varepsilon}}), \qquad x_{i+1,b}^{\vec{\varepsilon}} = f_{b_{i+1}}^{\vec{\varepsilon}}(x_{i,b}^{\vec{\varepsilon}}).$$

Note that there exists a positive constant L_1' such that

$$\tilde{d}_H(x_{-n,a}^{\vec{\varepsilon}}, x_{-n,b}^{\vec{\varepsilon}}) \leq L_1',$$

for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$, where $r_1 > 0$ are chosen sufficiently small. Then from (52), we have

$$\tilde{d}_H(x_{-1,a}^{\vec{\varepsilon}}, x_{-1,b}^{\vec{\varepsilon}}) \leq L_1' \rho_1^{n-1}.$$

Therefore, by Lemma 5.3, there exists a positive constant L_1'' independent of n_1, n_2 such that for any $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$, we have

$$\|x_{-1,a}^{\vec{\varepsilon}} - x_{-1,b}^{\vec{\varepsilon}}\| \leq L_1'' \rho_1^n. \tag{46}$$

Now, using (19) and the fact that

$$p^{\vec{\varepsilon}}(a_0) = \sum_y \pi_y^{\vec{\varepsilon}} q(a_0|y),$$

we conclude that there is a positive constant L_1, independent of n_1, n_2 such that

$$|p^{\vec{\varepsilon}}(a_0|a_{-n_1}^{-1}) - p^{\vec{\varepsilon}}(b_0|b_{-n_2}^{-1})| \leq L_1 \rho_1^n \max_{y' \in \mathcal{Y}} q(a_0|y'). \tag{47}$$

We then have finished the proof. □

We are now ready for the proof of Theorem 1.1.

Proof of Theorem 1.1. We first prove that there exist $r_1 > 0$ such that for any n, $h_n^{\vec{\varepsilon}}(Z)$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$.

For a fixed n, recall that

$$h_n^{\vec{\varepsilon}}(Z) = -\int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon}}(z_{-n}^0) \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) dz_{-n}^0,$$

where

$$p^{\vec{\varepsilon}}(z_{-n}^0) = \sum_{y_{-n}^0} p^{\vec{\varepsilon}}(y_{-n}^0) \prod_{i=-n}^{0} q(z_i|y_i)$$

and

$$p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) = x_{-1}^{\vec{\varepsilon}}(z_{-n}^{-1})\Pi^{\vec{\varepsilon}}(z_0)\mathbf{1}.$$

Now, for any $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$, we have

$$|p^{\vec{\varepsilon}}(z_{-n}^0)| \le \sum_{y_{-n}^0} |p^{\vec{\varepsilon}}(y_{-n}^0)| \prod_{i=-n}^{0} q(z_i|y_i) \le \sum_{y_{-n}^0} |p^{\vec{\varepsilon}}(y_{-n}^0)| \prod_{i=-n}^{0} \max_{y' \in \mathcal{Y}} q(z_i|y'). \tag{48}$$

And, by (41), for sufficiently small $r_1 > 0$, there exist $C_1, C_2 > 0$ such that for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$

$$C_1 \min_{y' \in \mathcal{Y}} q(z_0|y') \le |p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})| \le C_2 \max_{y' \in \mathcal{Y}} q(z_0|y'), \tag{49}$$

which, together with (42), implies that for some $C_3 > 0$,

$$|\log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})| \le C_3 + \max\{|\log \max_{y' \in \mathcal{Y}} q(z_0|y')|, |\log \min_{y' \in \mathcal{Y}} q(z_0|y')|\}.$$

This, together with (4), implies that on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$,

$$\int_{\mathcal{Z}^{n+1}} \sup_{\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)} \left| p^{\vec{\varepsilon}}(z_{-n}^0) \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) \right| dz_{-n}^0 < \infty \tag{50}$$

By Lemma 3.4 (Part 2), $h_n^{\vec{\varepsilon}}(Z_0|Z_{-n}^{-1})$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$.

Now, to prove the theorem, we only need to prove that there exist $r_1 > 0$ such that the $h_n^{\vec{\varepsilon}}(Z)$, as $n \to \infty$, uniformly converges on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$. Note that

$$\begin{aligned} |h_{n+1}^{\vec{\varepsilon}}(Z) - h_n^{\vec{\varepsilon}}(Z)| &= \left| \int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon}}(z_{-n-1}^0) \log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) dz_{-n-1}^0 - \int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon}}(z_{-n}^0) \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) dz_{-n}^0 \right| \\ &= \left| \int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})) dz_{-n-1}^0 \right|. \end{aligned}$$

Fix $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1)$. Then, by Lemmas 5.2 and 6.3, either we have, for some $0 < \rho_1 < 1$, $L_1' > 0$ and some δ_1 with $(1 + \delta_1)\rho_1 < 1$

$$|p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}))| \le L_1' \left| p^{\vec{\varepsilon}}(z_{-n-1}^0) \frac{p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})}{p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1})} \right|$$

$$\leq L_1' |p^{\vec{\varepsilon}}(z_{-n-1}^{-1})| L_1 \rho_1^n \max_{y' \in \mathcal{Y}} q(z_0|y'),$$

or we have, for some $0 < \rho_1 < 1$, $L_1' > 0$ and some δ_1 with $(1+\delta_1)\rho_1 < 1$,

$$|p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}))| \leq L_1' \left| p^{\vec{\varepsilon}}(z_{-n-1}^0) \frac{p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}))}{p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})} \right|$$

$$\leq L_1' |p^{\vec{\varepsilon}}(z_{-n}^{-1})||p^{\vec{\varepsilon}}(z_{-n-1}|z_{-n}^0)| L_1 \rho_1^n \max_{y' \in \mathcal{Y}} q(z_0|y').$$

Notice that for any given $\delta > 0$, there exist $r_1, r_2 > 0$ such that for all $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r_1)$,

$$|\pi^{\vec{\varepsilon}}_{y_{-n}}| \leq (1+\delta)\pi^{\vec{\varepsilon}_0}_{y_{-n}}, \qquad |\pi^{\vec{\varepsilon}}_{y_i y_{i+1}}| \leq (1+\delta)\pi^{\vec{\varepsilon}_0}_{y_i y_{i+1}}.$$

It then follows from the first inequality sign of (48) that

$$\int_{\mathcal{Z}^n} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^{-1})| dz_{-n}^{-1} \leq (1+\delta)^n, \qquad \int_{\mathcal{Z}^{n+1}} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^{-1})| dz_{-n-1}^{-1} \leq (1+\delta)^{n+1}.$$

Moreover, similar to (49), we have for some $C_4, C_5 > 0$,

$$C_4 \min_{y' \in \mathcal{Y}} q^{\vec{\theta}'}(z_{-n-1}|y') \leq |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}|z_{-n}^0)| \leq C_5 \max_{y' \in \mathcal{Y}} q(z_{-n-1}|y').$$

By choosing $\delta > 0$ sufficiently small, we can combine all the relevant inequalities above to obtain some $L > 0$ and some $0 < \rho < 1$ such that for all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$,

$$|h^{\vec{\varepsilon}}_{n+1}(Z) - h^{\vec{\varepsilon}}_n(Z)| \leq \int_{\mathcal{Z}^{n+2}} |p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}))| dz_{-n-1}^0 \leq L\rho^n,$$

which implies the analyticity of $h^{\vec{\varepsilon}}(Z)$ around $\vec{\varepsilon}_0$.

□

7 Proof of Theorem 1.8

The proof of Theorem 1.8 follows from a parallel flow as in that of Theorem 1.1. We will only give a sketch of the proof, while highlighting the parts requiring the extra conditions (b), (c) and (11).

The following lemma is an analog of Lemma 6.1.

Lemma 7.1. *For any $\delta > 0$, there exist $r_1, r_2 > 0$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$, any $z \in \mathcal{Z}$ and any $x \in W$, we have*

$$\tilde{d}_H(f_z^{\vec{\varepsilon},\vec{\theta}}(x), f_z^{\vec{\varepsilon}_0,\vec{\theta}_0}(x)) \leq \delta.$$

Proof. Since all $\pi_{ij}(\vec{\varepsilon}_0)$ are strictly positive, for any $\delta_1 > 0$, there exists $r_1 > 0$ such that for all i, j and all $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$,

$$\frac{|\pi^{\vec{\varepsilon}}_{ij} - \pi_{ij}(\vec{\varepsilon}_0)|}{\pi_{ij}(\vec{\varepsilon}_0)} \leq \delta_1.$$

Now, for any $x = (x_1, x_2, \cdots, x_l) \in W$, any j and any $\vec{\varepsilon} \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)$, we have

$$\left| \frac{\sum_{i=1}^l x_i(\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0))}{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0)} \right| = \left| \frac{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0)(\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0))/\pi_{ij}(\vec{\varepsilon}_0)}{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0)} \right| \le \delta_1.$$

Thus, for any $\delta_2 > 0$, choosing δ_1 sufficiently small, we have

$$\left| \log \frac{\sum_{i=1}^l x_i \pi_{ij}^{\vec{\varepsilon}}}{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0)} \right| = \left| \log \left(1 + \frac{\sum_{i=1}^l x_i(\pi_{ij}^{\vec{\varepsilon}} - \pi_{ij}(\vec{\varepsilon}_0))}{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0)} \right) \right| \le \delta_2.$$

Notice that

$$\tilde{d}_H(f_z^{\vec{\varepsilon},\vec{\theta}}(x), f_z^{\vec{\varepsilon}_0,\vec{\theta}_0}(x)) = \max_{j,k} \left| \log \frac{\sum_{i=1}^l x_i \pi_{ij}^{\vec{\varepsilon}} q^{\vec{\theta}}(z|j)}{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0) q^{\vec{\theta}_0}(z|j)} - \log \frac{\sum_{i=1}^l x_i \pi_{ik}^{\vec{\varepsilon}} q^{\vec{\theta}}(z|k)}{\sum_{i=1}^l x_i \pi_{ik}(\vec{\varepsilon}_0) q^{\vec{\theta}_0}(z|k)} \right|$$

$$= \max_{j,k} \left| \log \frac{\sum_{i=1}^l x_i \pi_{ij}^{\vec{\varepsilon}}}{\sum_{i=1}^l x_i \pi_{ij}(\vec{\varepsilon}_0)} + \log \frac{q^{\vec{\theta}}(z|j)}{q^{\vec{\theta}_0}(z|j)} - \log \frac{\sum_{i=1}^l x_i \pi_{ik}^{\vec{\varepsilon}}}{\sum_{i=1}^l x_i \pi_{ik}(\vec{\varepsilon}_0)} - \log \frac{q^{\vec{\theta}}(z|k)}{q^{\vec{\theta}_0}(z|k)} \right|.$$

It then follows from the second inequality of (11) that for any $\delta > 0$, there exist $r_1, r_2 > 0$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ and any $x \in W$, we have

$$\tilde{d}_H(f_z^{\vec{\varepsilon},\vec{\theta}}(x), f_z^{\vec{\varepsilon}_0,\vec{\theta}_0}(x)) \le \delta.$$

□

The following lemma is an analog of Theorem 6.2.

Lemma 7.2. *For sufficiently small $r_1, r_2, \delta > 0$, there exists $0 < \rho_1 < 1$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ and any $z \in \mathcal{Z}$, $f_z^{\vec{\varepsilon},\vec{\theta}}$ is a ρ_1-contraction mapping on $\tilde{W}_{W^\circ,H}(\delta)$ under the complex Hilbert metric in (39).*

Proof. By (11), we can choose $r_1, r_2, \delta > 0$ sufficiently small such that for any $z \in \mathcal{Z}$, any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ and any $u, v \in \tilde{W}_{W^\circ,H}(\delta)$,

$$\tilde{d}_H(u\Pi^{\vec{\varepsilon},\vec{\theta}}(z), v\Pi^{\vec{\varepsilon},\vec{\theta}}(z))$$

is well-defined. Moreover, it can be easily checked that

$$\tilde{d}_H(u\Pi^{\vec{\varepsilon},\vec{\theta}}(z), v\Pi^{\vec{\varepsilon},\vec{\theta}}(z)) = \tilde{d}_H(u\Pi^{\vec{\varepsilon}}, v\Pi^{\vec{\varepsilon}}). \tag{51}$$

The lemma then immediately follows from Theorem 5.1. □

The following lemma is an analog of Theorem 6.3.

Lemma 7.3. *1. For any $\delta > 0$, there exist $r_1, r_2 > 0$ such that for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$ and for any $z_{-n}^0 \in \mathcal{Z}^{n+1}$ and $-n-1 \le i \le -1$,*

$$x_i^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^i) \in \tilde{W}_{W^\circ,H}(\delta), \tag{52}$$

and

$$p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) \in \mathbb{C}_{\mathbb{R}^+}[\delta]. \tag{53}$$

2. *There exist* $r_1, r_2 > 0$ *such that for all* $z_{-n}^0 \in \mathcal{Z}^{n+1}$, $p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})$ *is analytic on* $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$.

3. *For sufficiently small* $r_1, r_2 > 0$, *there exist* $0 < \rho_1 < 1$ *and a positive constant* L_1 *such that for any two* $\mathcal{Z}$*-valued sequences* $\{a_{-n_1}^0\}$ *and* $\{b_{-n_2}^0\}$ *with* $a_{-n}^0 = b_{-n}^0$ *and for all* $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$, *we have*

$$|p^{\vec{\varepsilon},\vec{\theta}}(a_0|a_{-n_1}^{-1}) - p^{\vec{\varepsilon},\vec{\theta}}(b_0|b_{-n_2}^{-1})| \leq L_1 \rho_1^n \sup_{(y',\vec{\theta}') \in \mathcal{Y} \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |q^{\vec{\theta}'}(a_0|y')|.$$

Proof. 1. It follows from a parallel argument as in the proof of Part 1 of Theorem 6.3.

2. It follows from (11(i)) and Part 1 of Lemma 5.3 that for sufficiently small $r_1, r_2, \delta > 0$ and any $z \in \mathcal{Z}$, $f_z^{\vec{\varepsilon},\vec{\theta}}(x)$ is analytic with respect to $(\vec{\varepsilon}, \vec{\theta}, x) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2) \times \tilde{W}_{W^\circ,H}(\delta)$. It follows from this fact, the iterative nature of $x_i^{\vec{\varepsilon},\vec{\theta}}$ (see (17)) and Part 1 that for sufficiently small $r_1, r_2 > 0$, each $x_i^{\vec{\varepsilon},\vec{\theta}}$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$. Part 2 then immediately follows from (19).

3. It follows from a parallel argument as in the proof of Part 3 of Theorem 6.3. □

We are now ready for the proof of Theorem 1.8.

Proof of Theorem 1.8. We first prove that there exist $r_1, r_2 > 0$ such that for any n, $h_n^{\vec{\varepsilon},\vec{\theta}}(Z)$ is analytic on $\mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$.

For a fixed n, recall that

$$h_n^{\vec{\varepsilon},\vec{\theta}}(Z) = -\int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) dz_{-n}^0,$$

where

$$p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0) = \sum_{y_{-n}^0} p^{\vec{\varepsilon}}(y_{-n}^0) \prod_{i=-n}^{0} q^{\vec{\theta}}(z_i|y_i)$$

and

$$p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}) = x_{-1}^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^{-1}) \Pi^{\vec{\varepsilon},\vec{\theta}}(z_0) \mathbf{1}.$$

Now, for any $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}_{\vec{\varepsilon}_0}^{m_1}(r_1) \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)$, we have

$$|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)| \leq \sum_{y_{-n}^0} |p^{\vec{\varepsilon}}(y_{-n}^0)| \prod_{i=-n}^{0} |q^{\vec{\theta}}(z_i|y_i)| \leq \sum_{y_{-n}^0} |p^{\vec{\varepsilon}}(y_{-n}^0)| \prod_{i=-n}^{0} \sup_{(y',\vec{\theta}') \in \mathcal{Y} \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |q^{\vec{\theta}'}(z_i|y')|. \quad (54)$$

And, by (52), for sufficiently small $r_1, r_2 > 0$, there exist $C_1, C_2 > 0$ such that

$$C_1 \inf_{(y',\vec{\theta}') \in \mathcal{Y} \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |q^{\vec{\theta}'}(z_0|y')| \leq |p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})| \leq C_2 \sup_{(y',\vec{\theta}') \in \mathcal{Y} \times \mathbb{C}_{\vec{\theta}_0}^{m_2}(r_2)} |q^{\vec{\theta}'}(z_0|y')|, \quad (55)$$

which, together with (53), implies that for some $C_3 > 0$,

$$|\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})| \leq C_3 + \max\{|\log \sup_{(y',\vec{\theta}')\in\mathcal{Y}\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}'}(z_0|y')|, \log \inf_{(y',\vec{\theta}')\in\mathcal{Y}\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}'}(z_0|y')|\}.$$

This, together with (5), implies that on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

$$\int_{\mathcal{Z}^{n+1}} \sup_{(\vec{\varepsilon},\vec{\theta})\in\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} \left|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})\right| dz_{-n}^0 < \infty \tag{56}$$

By Lemma 3.4 (Part 2), $h_n^{\vec{\varepsilon},\vec{\theta}}(Z_0|Z_{-n}^{-1})$ is analytic on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$.

Now, to prove the theorem, we only need to prove that there exist $r_1, r_2 > 0$ such that the $h_n^{\vec{\varepsilon},\vec{\theta}}(Z)$, as $n\to\infty$, uniformly converges on $\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$. Note that

$$\begin{aligned}|h_{n+1}^{\vec{\varepsilon},\vec{\theta}}(Z) - h_n^{\vec{\varepsilon},\vec{\theta}}(Z)| &= \left|\int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1})dz_{-n-1}^0 - \int_{\mathcal{Z}^{n+1}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^0)\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})dz_{-n}^0\right| \\ &= \left|\int_{\mathcal{Z}^{n+2}} p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))dz_{-n-1}^0\right|.\end{aligned}$$

Fix $(\vec{\varepsilon},\vec{\theta})\in\mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1)\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$. Then, by Lemmas 5.2 and 7.3, either we have, for some $0<\rho_1<1$, $L_1'>0$ and some δ_1 with $(1+\delta_1)\rho_1<1$

$$\begin{aligned}|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))| &\leq L_1'\left|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)\frac{p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})}{p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1})}\right| \\ &\leq L_1'|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^{-1})|L_1\rho_1^n \sup_{(y',\vec{\theta}')\in\mathcal{Y}\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}'}(z_0|y')|,\end{aligned}$$

or we have, for some $0<\rho_1<1$, $L_1'>0$ and some δ_1 with $(1+\delta_1)\rho_1<1$,

$$\begin{aligned}|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))| &\leq L_1'\left|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)\frac{p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))}{p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1})}\right| \\ &\leq L_1'|p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^{-1})||p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}|z_{-n}^0)|L_1\rho_1^n \sup_{(y',\vec{\theta}')\in\mathcal{Y}\times\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}'}(z_0|y')|.\end{aligned}$$

Notice that for any given $\delta>0$, there exist $r_1, r_2 > 0$ such that for all $\vec{\varepsilon}\in\mathbb{C}^{m}_{\vec{\varepsilon}_0}(r_1)$,

$$|\pi^{\vec{\varepsilon}}_{y_{-n}}| \leq (1+\delta)\pi^{\vec{\varepsilon}_0}_{y_{-n}}, \qquad |\pi^{\vec{\varepsilon}}_{y_iy_{i+1}}| \leq (1+\delta)\pi^{\vec{\varepsilon}_0}_{y_iy_{i+1}},$$

and for any $y\in\mathcal{Y}$ and all $\vec{\theta}\in\mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

$$\int_{\mathcal{Z}} |q^{\vec{\theta}}(z|y)|dz \leq (1+\delta)\int_{\mathcal{Z}} q^{\vec{\theta}_0}(z|y)dz = 1+\delta,$$

(here we have used the fact that $\int_{\mathcal{Z}} |q^{\vec{\theta}}(z|y)|dz$ is a continuous function of $\vec{\theta} \in \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$; this follows from Lemma 3.4 (Part 1)). It then follows from the first inequality sign of (54) that

$$\int_{\mathcal{Z}^n} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n}^{-1})| dz_{-n}^{-1} \leq (1+\delta)^{2n}, \qquad \int_{\mathcal{Z}^{n+1}} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^{-1})| dz_{-n-1}^{-1} \leq (1+\delta)^{2(n+1)}.$$

Moreover, similar to (55), we have for some $C_4, C_5 > 0$,

$$C_4 \inf_{(y',\vec{\theta}') \in \mathcal{Y} \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}'}(z_{-n-1}|y')| \leq |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}|z_{-n}^0)| \leq C_5 \sup_{(y',\vec{\theta}') \in \mathcal{Y} \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)} |q^{\vec{\theta}'}(z_{-n-1}|y')|.$$

By choosing $\delta > 0$ sufficiently small, we can combine all the relevant inequalities above to obtain some $L > 0$ and some $0 < \rho < 1$ such that for all $(\vec{\varepsilon}, \vec{\theta}) \in \mathbb{C}^{m_1}_{\vec{\varepsilon}_0}(r_1) \times \mathbb{C}^{m_2}_{\vec{\theta}_0}(r_2)$,

$$|h_{n+1}^{\vec{\varepsilon},\vec{\theta}}(Z) - h_n^{\vec{\varepsilon},\vec{\theta}}(Z)| \leq \int_{\mathcal{Z}^{n+2}} |p^{\vec{\varepsilon},\vec{\theta}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon},\vec{\theta}}(z_0|z_{-n}^{-1}))| dz_{-n-1}^0 \leq L\rho^n,$$

which implies the analyticity of $h^{\vec{\varepsilon},\vec{\theta}}(Z)$ around $(\vec{\varepsilon}_0, \vec{\theta}_0)$.

□

8 Concluding Remarks

Under certain mild assumptions, employing a complex Hilbert metric in a critical way, we show that the entropy rate of a class of hidden Markov chains with continuous alphabet is analytic with respect to the input Markov chain parameters. Joint analyticity results with respect to both the input Markov chain parameters and the channel parameters, which apply to additive Cauchy or Gaussian channels, are further obtained under strengthened assumptions on the channel. Given the implications of the analyticity results in the discrete setting, we expect that the results in this paper will be of great interest and significance in the continuous setting. Further work include the applications of the analyticity results in this paper to computations of entropy rate of hidden Markov chains and capacity of finite-state channels with continuous output alphabet.

References

[1] A. Allahverdyan. Entropy of Hidden Markov Processes via Cycle Expansion. *J. Stat. Phys.*, vol. 133, 2008, pp. 535-564.

[2] L. Arnold, V. M. Gundlach and L. Demetrius. Evolutionary formalism for products of positive random matrices. *Annals of Applied Probability*, vol. 4, 1994, pp. 859-901.

[3] J. J. Birch. Approximations for the entropy for functions of Markov chains. *Ann. Math. Statist.*, vol. 33, 1962, pp. 930-938.

[4] D. Blackwell. The entropy of functions of finite-state Markov chains. *Trans. First Prague Conf. Information Thoery, Statistical Decision Functions, Random Processes*, 1957, pp. 13-20.

[5] M. Boyle and K. Petersen. Hidden Markov processes in the context of symbolic dynamics. *Entropy of Hidden Markov Processes and Connections to Dynamical Systems, London Mathematical Society Lecture Note Series*, vol. 385, edited by B. Marcus, K. Petersen and T. Weissman, 2011, pp. 5-71.

[6] L. Dubois. Projective metrics and contraction principles for complex cones. *Journal of the London Mathematical Society*, vol. 79, no. 3, 2009, pp. 719-737.

[7] S. Egner, V. Balakirsky, L. Tolhuizen, S. Baggen and H. Hollmann. On the entropy rate of a hidden Markov model. *IEEE ISIT*, 2004, pp. 12.

[8] R. Gharavi and V. Anantharam. An upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices. *Theoretical Computer Science*, vol. 332, no. 1-3, 2005, pp. 543-557.

[9] G. Han. Limit theorems in hidden Markov models. *IEEE Trans. Info. Theory*, vol. 59, no. 3, 2013, pp. 1311-1328.

[10] G. Han. A randomized approach to the capacity of finite-state channels. *IEEE ISIT*, 2013, pp. 2109-2113.

[11] G. Han and B. Marcus. Analyticity of entropy rate of hidden Markov chains. *IEEE Trans. Info. Theory*, vol. 52, no. 12, 2006, pp. 5251-5266.

[12] G. Han and B. Marcus. Derivatives of entropy rate in special families of hidden Markov chains. *IEEE Trans. Info. Theory*, vol. 53, no. 7, 2007, pp. 2642-2652.

[13] G. Han and B. Marcus. Asymptotics of input-constrained binary symmetric channel capacity. *Annals of Applied Probability*, vol. 19, no. 3, 2009, pp. 1063-1091.

[14] G. Han and B. Marcus. Entropy rate of continuous-state hidden Markov chains. *IEEE ISIT*, 2010, pp. 1468-1472.

[15] G. Han and B. Marcus and Y. Peres. A complex Hilbert metric and applications to domain of analyticity for entropy rate of hidden Markov processes. *Entropy of Hidden Markov Processes and Connections to Dynamical Systems, London Mathematical Society Lecture Note Series*, vol. 385, edited by B. Marcus, K. Petersen and T. Weissman, 2011, pp. 98-116.

[16] G. Han and B. Marcus. Asymptotics of entropy rate in special families of hidden Markov chains. *IEEE Trans. Info. Theory*, vol. 56, no. 3, 2010, pp. 1287-1295.

[17] G. Han and B. Marcus. Concavity of mutual information rate for input-restricted finite-state memoryless channels at high SNR. *IEEE Trans. Info. Theory*, vol. 58, no. 3, 2012, pp. 1534-1548.

[18] T. Holliday, A. Goldsmith, and P. Glynn. Capacity of finite state channels based on Lyapunov exponents of random matrices. *IEEE Trans. Info. Theory*, vol. 52, no. 8, 2006, pp. 3509-3532.

[19] S. Ihara. *Information Theory for Continuous Systems,* World Scientific, 1993.

[20] P. Jacquet, G. Seroussi, and W. Szpankowski. Noisy constrained capacity. *IEEE ISIT*, 2007, pp. 986-990.

[21] P. Jacquet, G. Seroussi, and W. Szpankowski. On the entropy of a hidden Markov process. *Theoretical Computer Science*, vol. 395, 2008, pp. 203-219.

[22] F. Ledrappier. Analyticity of the entropy for some random walks. Preprint. Available at http://arxiv.org/abs/1009.5354.

[23] F. Le Gland and N.Oudjane. Stability and Uniform Approximation of Nonlinear Filters Using the Hilbert Metric and Application to Particle Filters. *The Annals of Applied Probability*, vol. 14, no. 1, 2004, pp. 144-187.

[24] Y. Li and G. Han. Concavity of mutual information rate of finite-state channels. *IEEE ISIT*, 2013, pp. 2114-2118.

[25] J. Luo and D. Guo. On the entropy rate of hidden Markov processes observed through arbitrary memoryless channels. *IEEE Trans. Info. Theory*, vol. 55, no. 4, 2009, pp. 1460-1467.

[26] C. Nair, E. Ordentlich and T. Weissman. Asymptotic filtering and entropy rate of a hidden Markov process in the rare transitions regime. *IEEE ISIT*, 2005, pp. 1838-1842.

[27] E. Ordentlich and T. Weissman. On the optimality of symbol by symbol filtering and denoising. *IEEE Trans. Info. Theory*, vol. 52, no. 1, 2006, pp. 19-40.

[28] E. Ordentlich and T. Weissman. New bounds on the entropy rate of hidden Markov process. *IEEE ITW*, 2004, pp. 117-122.

[29] E. Ordentlich and T. Weissman. Bounds on the entropy rate of biunary hidden Markov processes *Entropy of Hidden Markov Processes and Connections to Dynamical Systems, London Mathematical Society Lecture Note Series*, vol. 385, edited by B. Marcus, K. Petersen and T. Weissman, 2011, pp. 117-171.

[30] P. Vontobel, A. Kavcic, D. Arnold and Hans-Andrea Loeliger. A generalization of the Blahut-Arimoto algorithm to finite-state channels. *IEEE Trans. Info. Theory*, vol. 54, no. 5, 2008, pp. 1887-1918.

[31] Y. Peres. Analytic dependence of Lyapunov exponents on transition probabilities. *Lecture Notes in Mathematics*, Springer Verlag, vol. 1486, 1990.

[32] Y. Peres. Domains of analytic continuation for the top Lyapunov exponent. *Ann. Inst. H. Poincare Prob. Statist*, vol. 28, no. 1, 1991, pp. 131-148.

[33] Y. Peres and A. Quas. Entropy rate for hidden Markov chains with rare transitions. *Entropy of Hidden Markov Processes and Connections to Dynamical Systems, London Mathematical Society Lecture Note Series*, vol. 385, edited by B. Marcus, K. Petersen and T. Weissman, 2011, pp. 172-178.

[34] H. Pfister, J. Soriaga and P. Siegel. The achievable information rates of finite-state ISI channels. *IEEE GLOBECOM*, 2011, pp. 2992-2996.

[35] H. Pfister. On the capacity of finite state channels and the analysis of convolutional accumulate-m codes. Ph.D. thesis, University of California at San Diego, USA, 2003.

[36] H. Pfister. The capacity of finite-state channels in the high-noise regime. *Entropy of Hidden Markov Processes and Connections to Dynamical Systems, London Mathematical Society Lecture Note Series*, vol. 385, edited by B. Marcus, K. Petersen and T. Weissman, 2011, pp. 179-222.

[37] H. Rugh. Cones and gauges in complex spaces: Spectral gaps and complex Perron-Frobenius theory. *Annals of Mathematics*, vol. 171, no. 3, 2010.

[38] E. Seneta. *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York Heidelberg Berlin, 1980.

[39] V. Sharma and S. Singh. Entropy and channel capacity in the regenerative setup with applications to Markov channels. *IEEE ISIT*, 2001, pp. 283.

[40] V. Tadic. Analyticity, Convergence, and Convergence Rate of Recursive Maximum-Likelihood Estimation in Hidden Markov Models. *IEEE Trans. Info. Theory*, vol. 56, no, 12, 2010, pp. 6406-6432.

[41] E. Zehavi and J. Wolf. On runlength codes. *IEEE Trans. Info. Theory*, vol. 34, 1988, pp. 45-54.

[42] O. Zuk, I. Kanter and E. Domany. The entropy of a binary hidden Markov process. *J. Stat. Phys.*, vol. 121, no. 3-4, 2005, pp. 343-360.

[43] O. Zuk, E. Domany, I. Kanter and M. Aizenman. From finite-system entropy to entropy rate for a hidden Markov process. *IEEE Signal Processing Letters*, vol. 13, no. 9, 2006, pp. 517-520.