Entropy Rate of Continuous-State Hidden Markov Chains

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

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

March 21, 2011

Abstract

We prove that under mild positivity assumptions, the entropy rate of a continuous-state hidden Markov chain, observed when passing a finite-state Markov chain through a discrete-time continuous-output channel, is analytic as a function of the transition probabilities of the underlying Markov chain. We further prove that the entropy rate of a continuous-state hidden Markov chain, observed when passing a mixing finite-type constrained Markov chain through a discrete-time Gaussian channel, is smooth as a function of the transition probabilities of the underlying Markov chain.

1 Main Results

Consider a discrete-time channel with a finite input alphabet $\mathcal{Y}$ and the continuous output alphabet $\mathcal{Z} = \mathbb{R}$. Assume that the input process is a $\mathcal{Y}$-valued first order stationary Markov chain Y with transition probability matrix $\Pi = (\pi_{ij})_{|\mathcal{Y}|\times|\mathcal{Y}|}$ and stationary vector $\pi = (\pi_i)_{|\mathcal{Y}|}$ (here we assume Y is first order only for simplicity; an usual "blocking" approach can be used to reduce higher order case to first order case). 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 probability density function $q(z|y)$.

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

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

when the limit exists, where

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

here $z_{-n}^0 := (z_{-n}, z_{-n+1}, \cdots, z_0)$ denotes an instance of $Z_{-n}^0 := (Z_{-n}, Z_{-n+1}, \cdots, Z_0)$, and $p(z_{-n}^0)$ denotes the probability density of z_{-n}^0. It is well-known (e.g., see page 60 of [3]) that if $H(Z_{-n}^0)$ is finite for all n, $H(Z)$ is well-defined and can be written as

$$H(Z) = \lim_{n \to \infty} H_n(Z),$$

where

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

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

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

where $H(Z_0|Y_0)$ can be computed as

$$H(Z_0|Y_0) = -\sum_{i \in \mathcal{Y}} \pi_i \int_{z \in \mathcal{Z}} q(z|i) \log q(z|i) dz.$$

It then follows from

$$H(Z_{-n}^0 | Y_{-n}^0) \le H(Z_{-n}^0) \le H(Y_{-n}^0) + H(Z_{-n}^0 | Y_{-n}^0)$$

that if

$$\int_{z \in \mathcal{Z}} q(z|i) \log q(z|i) dz$$

is finite for all i, $H(Z)$ is well-defined and finite.

The following theorem states that under positivity assumptions, $H(Z)$ is analytic as a function of Π.

Theorem 1.1. *Consider a discrete-time memoryless continuous-output channel as above. Assume that Π is analytically parameterized by $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m) \in \Omega$, where Ω denotes an open and bounded subset of $\mathbb{R}^m$, and assume that $q(z|y) > 0$ for all $(y, z) \in (\mathcal{Y}, \mathcal{Z})$, and the integral*

$$\int_{z \in \mathcal{Z}} q(z|i) \log q(z|i) dz$$

is finite for all i. If Π is strictly positive at $\vec{\varepsilon}_0$, then $H(Z)$ is analytic around $\vec{\varepsilon}_0$.

Our next result deals with a discrete-time memoryless Gaussian channel, a special type of discrete-time memoryless continuous-output channel. We shall relax the positivity assumptions in Theorem 1.1, and we assume that the input Markov chain is supported on a mixing finite-type constraint. The consideration of such channels mainly comes from practice: Gaussian channels are of great importance in a variety of scenarios in real applications, and often (particularly in magnetic recording) input sequences are required to satisfy certain constraints in order to eliminate the most damaging error events [8] and the constraints are often mixing finite-type constraints.

Let $\mathcal{X}$ be a finite alphabet, and let $\mathcal{X}^n$ denote the set of words over $\mathcal{X}$ of length n. Let $\mathcal{X}^* = \cup_n \mathcal{X}^n$. A *finite-type* constraint $\mathcal{S}$ over $\mathcal{X}$ is a subset of $\mathcal{X}^*$ defined by a finite list $\mathcal{F}$ of forbidden words [7, 8]; in other words, $\mathcal{S}$ is the set of words over $\mathcal{X}$ that do not contain any element in $\mathcal{F}$ as a contiguous subsequence. We define $\mathcal{S}_n = \mathcal{S} \cap \mathcal{X}^n$. The constraint $\mathcal{S}$ is said to be *mixing* if there exists N such that, for any $u, v \in \mathcal{S}$ and any $n \geq N$, there is a $w \in \mathcal{S}_n$ such that $uwv \in \mathcal{S}$.

The *maximal length* of a forbidden list $\mathcal{F}$ is the length of the longest word in $\mathcal{F}$. In general, there can be many forbidden lists $\mathcal{F}$ which define the same finite type constraint $\mathcal{S}$. However, we may always choose a list with smallest maximal length. The *(topological) order* of $\mathcal{S}$ is defined to be $\hat{m} = \hat{m}(\mathcal{S})$ where $\hat{m}+1$ is the smallest maximal length of any forbidden list that defines $\mathcal{S}$ (the order of the trivial constraint $\mathcal{X}^*$ is taken to be 0). For example, one checks that the order of the (d, k)-RLL constraint [7], which is a commonly seen mixing finite-type constraint, is k when $k < \infty$, and is d when $k = \infty$.

For a stationary stochastic process X over $\mathcal{X}$, the set of *allowed* words with respect to X is defined as

$$\mathcal{A}(X) = \{w_{-n}^0 : n \geq 0, P(X_{-n}^0 = w_{-n}^0) > 0\}.$$

For any m-th order Markov process X, we say X is *supported* on a constraint $\mathcal{S}$ if $\mathcal{S} = \mathcal{A}(X)$; note that in this case, the constraint $\mathcal{S}$ is necessarily of finite-type with order $\hat{m} \leq m$. Also, X is mixing if and only if S is mixing (recall that a Markov chain is mixing if its transition probability matrix (obtained by appropriately enlarging the state space) is irreducible and aperiodic).

Now, consider a discrete-time memoryless Gaussian channel, which is a special case of the generic channel model described in the beginning of this paper. More specifically, for any input $y \in \mathcal{Y}$, the channel is characterized by the transition probability density function

$$q(z|y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(z-y)^2/(2\sigma_y^2)}, \tag{2}$$

where $\sigma_y > 0$, and $z \in \mathcal{Z}$ denotes a possible output of the channel.

The following theorem states that under certain assumptions, $H(Z)$ is smooth (infinitely differentiable) as a function of the transition probabilities of Y. More specifically, we state our second result of this paper as follows.

Theorem 1.2. *Consider a discrete-time memoryless Gaussian channel as above. Assume that* Π *is analytically parameterized by* $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m) \in \Omega$, *where* Ω *denotes an open and bounded subset of* $\mathbb{R}^m$, *and assume that at* $\vec{\varepsilon}_0 \in \Omega$, *the input Markov chain* Y *is supported on a mixing finite-type constraint* $\mathcal{S}$, *i.e.,* $\mathcal{A}(X) = \mathcal{S}$, *then* $H(Z)$ *is smooth around* $\vec{\varepsilon}_0$.

The rest of the paper is organized as follows. In Section 2, we briefly review the Hilber metric and introduce a complex Hilbert. In Section 3, using the complex Hilbert metric, we prove that for any continuous channel, under mild positivity assumptions, $H(Z)$ is analytic with respect to the input Markov parameters (Theorem 1.1). In Section 4, we prove that for a Gaussian channel, where the input Markov chain is supported on a mixing finite-type constraint, $H(Z)$ is smooth with respect to the input Markov parameters (Theorem 1.2).

2 A Complex Hilbert Metric

In this section, we briefly review the classical Hilbert metric and review a new complex Hilbert metric, which we will use to prove Theorem 1.1.

Let W be the standard simplex in $|\mathcal{Y}|$-dimensional real Euclidean space,

$$W = \{w = (w_1, w_2, \cdots, w_{|\mathcal{Y}|}) \in \mathbb{R}^{|\mathcal{Y}|} : w_i \geq 0, \sum_i w_i = 1\},$$

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 [9] is defined as

$$d_H(w, v) = \max_{i,j} \log \left(\frac{w_i/w_j}{v_i/v_j} \right). \tag{3}$$

For a $|\mathcal{Y}| \times |\mathcal{Y}|$ strictly positive matrix $T = (t_{ij})$, the mapping f_T induced by T on W is defined by

$$f_T(w) = \frac{wT}{(wT\mathbf{1})}, \tag{4}$$

where $\mathbf{1}$ is the all 1 column vector. It is well known that f_T is a contraction mapping under the Hilbert metric [9]. The contraction coefficient of T, which is also called the Birkhoff coefficient, is given by

$$\tau(T) = \sup_{v \neq w} \frac{d_H(vT, wT)}{d_H(v, w)} = \frac{1 - \sqrt{\phi(T)}}{1 + \sqrt{\phi(T)}}, \tag{5}$$

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

Let $\hat{W}$ denote the complex version of W,

$$\hat{W} = \{w = (w_1, w_2, \cdots, w_{|\mathcal{Y}|}) \in \mathbb{C}^{|\mathcal{Y}|} : \sum_i w_i = 1\}.$$

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

$$\hat{d}_H(v, w) = \max_{i,j} \left| \log \left(\frac{w_i/w_j}{v_i/v_j} \right) \right|, \tag{6}$$

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, $\hat{d}_H$ is a metric on $\hat{W}^+$, which we call a *complex Hilbert metric*.

Let M denote the set of all stochastic matrices with dimension $|\mathcal{Y}| \times |\mathcal{Y}|$, i.e.,

$$M = \{\Pi = (\pi_{ij}) \in \mathbb{R}^{|\mathcal{Y}| \times |\mathcal{Y}|} : \pi_{ij} \geq 0, \sum_{j=1}^{|\mathcal{Y}|} \pi_{ij} = 1\}.$$

Let $\hat{M}$ denote the complex version of M, defined as

$$\hat{M} = \{\Pi = (\pi_{ij}) \in \mathbb{C}^{|\mathcal{Y}| \times |\mathcal{Y}|} : \sum_{j=1}^{|\mathcal{Y}|} \pi_{ij} = 1\}.$$

For a given positive Π and a small $\delta_1 > 0$, let $\hat{M}_\Pi(\delta_1)$ denote the δ_1-neighborhood around Π within $\hat{M}$. For an element $\hat{\Pi} \in \hat{M}_\Pi(\delta_1)$, similar to (4), $\hat{\Pi}$ will induce a mapping $f_{\hat{\Pi}}$ on $\hat{W}$. For a small $\delta_2 > 0$, let $\hat{W}^\circ_H(\delta_2)$ denote the δ_2-neighborhood of W° within $\hat{W}^+$ under the complex Hilbert metric, i.e.,

$$\hat{W}^\circ_H(\delta_2) = \{v = (v_1, v_2, \cdots, v_{|\mathcal{Y}|}) \in \hat{W}^+ : \exists u \in W^\circ, \hat{d}_H(v, u) \leq \delta_2\}.$$

The main theorem in [5] says:

Theorem 2.1. *For sufficiently small $\delta_1, \delta_2 > 0$, there exists $0 < \rho_1 < 0$ such that for any $\hat{\Pi} \in \hat{M}_\Pi(\delta_1)$, $f_{\hat{\Pi}}$ is a ρ_1-contraction mapping on $\hat{W}^\circ_H(\delta_2)$ under the complex Hilbert metric in (6).*

3 Proof of Theorem 1.1

In this section, we consider a discrete-time memoryless continuous-output channel as in Theorem 1.1, which was described in the beginning of Section 1.

For each $z \in \mathcal{Z}$, define $\Pi(z)$ as a $|\mathcal{Y}| \times |\mathcal{Y}|$ matrix with the entries

$$\Pi(z)_{ij} = \pi_{ij}(\vec{\varepsilon})q(z|j), \text{ for all } i, j, \tag{7}$$

here we suppressed the dependence of $\Pi(z)$ on $\vec{\varepsilon}$ for notational simplicity. By (4), $\Pi(z)$ will induce a mapping $f_z^{\vec{\varepsilon}} := f_{\Pi(z)}$ from W to W. For any fixed n and z_{-n}^0, define

$$x_i^{\vec{\varepsilon}} = x_i^{\vec{\varepsilon}}(z_{-n}^i) = p(y_i = \cdot\, |z_i, z_{i-1}, \cdots, z_{-n}), \tag{8}$$

(here $\cdot$ represent the states of the Markov chain Y,) then similar to Blackwell [1], $\{x_i^{\vec{\varepsilon}}\}$ satisfies the random dynamical system

$$x_{i+1}^{\vec{\varepsilon}} = f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}}), \tag{9}$$

starting with

$$x_{-n-1}^{\vec{\varepsilon}} = \pi(\vec{\varepsilon}). \tag{10}$$

And obviously we have

$$p^{\vec{\varepsilon}}(z_0|z_{-n}) = x_{-1}^{\vec{\varepsilon}}\Pi(z_0)\mathbf{1}, \tag{11}$$

and

$$p^{\vec{\varepsilon}}(z_{-n}^0) = \pi(\vec{\varepsilon})\Pi(z_{-n})\Pi(z_{-n+1})\cdots\Pi(z_0)\mathbf{1}. \tag{12}$$

Apparently $x_i^{\vec{\varepsilon}}$, $p^{\vec{\varepsilon}}(z_0|z_{-n})$ and $p^{\vec{\varepsilon}}(z_{-n}^0)$ all depend on the real vector $\vec{\varepsilon} \in \Omega$. In what follows, we shall show that they can be "complexified". For $r > 0$, let $\mathbb{C}^m_{\vec{\varepsilon}_0}(r)$ denote a r-ball around $\vec{\varepsilon}_0$ in $\mathbb{C}^m$. For any $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r)$, one checks that for r small enough, the following system of equations with respect to $\pi(\vec{\varepsilon})$

$$\pi(\vec{\varepsilon})\Pi = \pi(\vec{\varepsilon}), \qquad \sum_y \pi(\vec{\varepsilon})_y = 1$$

has a unique solution $\pi(\vec{\varepsilon})$, which is analytic on $\mathbb{C}^m_{\vec{\varepsilon}_0}(r)$ as a function of $\vec{\varepsilon}$. Then through (10) and (9), $x_i^{\vec{\varepsilon}}$ can be analytically extended to $\mathbb{C}^m_{\vec{\varepsilon}_0}(r)$; furthermore, through (11) and (12), $p^{\vec{\varepsilon}}(z_0|z_{-n})$ and $p^{\vec{\varepsilon}}(z_{-n}^0)$ can be analytically extended to $\mathbb{C}^m_{\vec{\varepsilon}_0}(r)$. Eventually, $H_n^{\vec{\varepsilon}}(Z)$ can be analytically extended to $\mathbb{C}^m_{\vec{\varepsilon}_0}(r)$ as well.

For any $z \in \mathcal{Z}$, by the definition of $\Pi(z)$, one checks that for any $u, v \in \hat{W}$, we have

$$\hat{d}_H(u\Pi(z), v\Pi(z)) = \hat{d}_H(u\Pi, v\Pi). \tag{13}$$

Then immediately by Theorem 2.1, we have the following lemma, which, roughly speaking, says that if we perturb $\vec{\varepsilon}_0$ "a bit" to $\vec{\varepsilon}$, $f_z^{\vec{\varepsilon}}$ is a contraction mapping on a complex neighborhood of W°, and the contraction coefficient is uniform over all the values of z.

Lemma 3.1. *For sufficiently small $r, \delta > 0$, there exists $0 < \rho_1 < 1$ such that for any $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r)$ and any $z \in \mathcal{Z}$, $f_z^{\vec{\varepsilon}}$ is a ρ_1-contraction mapping on $\hat{W}_H^\circ(\delta)$ under the complex Hilbert metric in (6).*

The following lemma, roughly speaking, says that if we perturb $\vec{\varepsilon}_0$ "a bit" to $\vec{\varepsilon}$, the image of any point in W under $f_z^{\vec{\varepsilon}}$, for any $z \in \mathcal{Z}$, does not change much.

Lemma 3.2. *Consider any $\vec{\varepsilon}_0 \in \Omega$ with $\pi_{ij}(\vec{\varepsilon}_0) > 0$ for all i, j. For any $\delta > 0$, there exists $r > 0$ such that for any $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r)$, any $z \in \mathcal{Z}$ and any $x \in W$, we have*

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

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

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

Now for any $x = (x_1, x_2, \cdots, x_{|\mathcal{Y}|}) \in W$ and for any j and for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r)$, we have

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

Thus, for δ_1 small enough, we have

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

Notice that

$$\begin{aligned} \hat{d}_H(f_z^{\vec{\varepsilon}}(x), f_z^{\vec{\varepsilon}_0}(x)) &= \max_{j,k} \left(\log \frac{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ij}(\vec{\varepsilon}) q(z|j)}{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ij}(\vec{\varepsilon}_0) q(z|j)} - \log \frac{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ik}(\vec{\varepsilon}) q(z|k)}{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ik}(\vec{\varepsilon}_0) q(z|k)} \right) \\ &= \max_{j,k} \left(\log \frac{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ij}(\vec{\varepsilon})}{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ij}(\vec{\varepsilon}_0)} - \log \frac{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ik}(\vec{\varepsilon})}{\sum_{i=1}^{|\mathcal{Y}|} x_i \pi_{ik}(\vec{\varepsilon}_0)} \right). \end{aligned}$$

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

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

$\square$

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

$$\mathbb{C}_{\mathbb{R}^+}[\delta] = \{z \in \mathbb{C} : |z - x| \leq \delta x, \text{ for some } x > 0\}.$$

The following lemma, which is implied by the proof of Lemma 1.3 in [5], allows us to connect the complex Hilbert metric and the Euclidean metric. We give a proof for completeness.

Lemma 3.3. *1. For any $\delta > 0$, there exists $\xi > 0$ such that for any $\hat{x} \in \hat{W}^+$, $x \in W^\circ$ with $\hat{d}_H(\hat{x}, x) \leq \xi$, we have $\hat{x}_i \in \mathbb{C}_{\mathbb{R}^+}[\delta]$ for all i.*

2. For any $\zeta > 0$ and any $\delta > 0$, there exists $\xi > 0$ such that for any $\hat{x}, \hat{y} \in \hat{W}^+$ with $|\hat{x} - x|, |\hat{y} - y| \leq \zeta$ for some $x, y \in W^\circ$, and $\hat{d}_H(\hat{x}, \hat{y}) \leq \xi$, we have $|\hat{x} - \hat{y}| \leq \delta$.

Proof. **1.** Fix certain $\xi > 0$ and assume that $\hat{d}(\hat{x}, x) \leq \xi$. Then we have for all i, j,

$$\left|\log\left(\frac{\hat{x}_i/x_i}{\hat{x}_j/x_j}\right)\right| < \xi.$$

It follows that for some $L > 0$ and for all i, j, $\left|\frac{\hat{x}_i/x_i}{\hat{x}_j/x_j} - 1\right| < L\xi$. Let $\alpha_j = \hat{x}_j/x_j$. Then for all i, j,

$$|\hat{x}_i - \alpha_j x_i| \leq L\xi|\alpha_j| x_i,$$

and so

$$|1 - \alpha_j| = \left|\sum_{i=1}^n (\hat{x}_i - \alpha_j x_i)\right| \leq \sum_{i=1}^n |\hat{x}_i - \alpha_j x_i| \leq L\xi|\alpha_j| \sum_{i=1}^n x_i = L\xi|\alpha_j|.$$

It follows that $|\hat{x}_j - x_j| \leq L\xi|\hat{x}_j|$, and so $|\hat{x}_j| \leq \frac{x_j}{1-L\xi}$, and so $|\hat{x}_j - x_j| \leq \frac{L\xi}{1-L\xi} x_j \leq 2L\xi x_j$, which implies **1.**, if ξ is sufficiently small.

2. Fix certain $\xi > 0$ and assume that $\hat{d}(\hat{x}, \hat{y}) \leq \xi$. Then we have for all i, j,

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

It follows that for some $L > 0$ and for all i, j, $\left|\frac{\hat{x}_i/\hat{y}_i}{\hat{x}_j/\hat{y}_j} - 1\right| < L\xi$. Let $\alpha_j = \hat{x}_j/\hat{y}_j$. Then for all i, j,

$$|\hat{x}_i - \alpha_j \hat{y}_i| \leq L\xi|\alpha_j||\hat{y}_i|,$$

and so

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

It follows that $|\hat{x}_j - \hat{y}_j| \leq L(1 + B\zeta)\xi|\hat{x}_j| \leq L(1 + B\zeta)\xi(x_j + \zeta)$, which implies **2.**, if ξ is sufficiently small. $\square$

We are now ready for the following lemma, whose proof can be roughly described as follows. As before, we complexify the real random dynamical system corresponding to (9). Lemma 3.1 and Lemma 3.2 can guarantee the complex orbit will be exponentially close to the original real orbit under the complex Hilbert metric, thus implying the complex orbit will be close to W under the Euclidan metric and further, with (11), establishing part 1). For part 2), again by Lemma 3.1, we can show that the complex orbits, starting from possibly different initial points, get exponentially close under the complex Hilbert metric, then with (11) and Lemma 3.3, we can establish part 2).

Lemma 3.4. *1) For any $\delta > 0$, there exists $r > 0$ such that for any $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r)$ and for all $z_{-n}^0 \in \mathcal{Z}^{n+1}$,*

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

2) For sufficiently small $r > 0$, 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 $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r)$, we have

$$|p^{\vec{\varepsilon}}(a_0|a_{-n_1}^{-1}) - p^{\vec{\varepsilon}}(b_0|b_{-n_2}^{-1})| \leq L_1 \rho_1^n p^{\vec{\varepsilon}_0}(a_0).$$

Proof. By Lemma 3.1, we can choose r and δ sufficiently small such that

$$\text{for some } 0 < \rho_1 < 1, f_z^{\vec{\varepsilon}} \text{ is analytic on } \mathbb{C}_{\vec{\varepsilon}_0}(r) \text{ and is a } \rho_1-\text{contraction on } \hat{W}_H^\circ(\delta)$$
$$\text{under the complex Hilbert metric.} \tag{14}$$

Further, we claim that by choosing r smaller, if necessary, such that

$$\text{for all } \vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r), \text{ all } i, n \text{ and all choices of } z_{-n}^i, x_i^{\hat{\varepsilon}} \in \hat{W}_H^\circ(\delta), \tag{15}$$

To see this, fixing ρ_1 and δ, choose $r > 0$ so small (the existence of r is guaranteed by Lemma 3.2) such that

$$\hat{d}_H(f_z^{\vec{\varepsilon}}(x), f_z^{\vec{\varepsilon}_0}(x)) \leq \delta(1-\rho_1), \text{ for any } z \in \mathcal{Z}, \text{ for all } x \in W, \text{ all } \vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r) \tag{16}$$

and

$$\hat{d}_H(\pi(\vec{\varepsilon}), \pi(\vec{\varepsilon}_0)) \leq \delta(1-\rho_1), \text{ for all } \vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r). \tag{17}$$

Now consider the Hilbert distance

$$\hat{d}_H(x_{i+1}^{\vec{\varepsilon}}, x_{i+1}^{\vec{\varepsilon}_0}) = \hat{d}_H(f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}}), f_{z_{i+1}}^{\vec{\varepsilon}_0}(x_i^{\vec{\varepsilon}_0})) \leq \hat{d}_H(f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}}), f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}_0})) + \hat{d}_H(f_{z_{i+1}}^{\vec{\varepsilon}}(x_i^{\vec{\varepsilon}_0}), f_{z_{i+1}}^{\vec{\varepsilon}_0}(x_i^{\vec{\varepsilon}_0})). \tag{18}$$

Then by (14), (16) and (17), and (18), for $i > -n-1$, we have

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

So, for all i,

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

and thus for all i, we have $x_{i+1}^{\vec{\varepsilon}} \in \hat{W}_H^\circ(\delta)$, yielding (15). Each $x_i^{\vec{\varepsilon}}$ is the composition of analytic functions on $\mathbb{C}_{\vec{\varepsilon}_0}(r)$ and so is complex analytic on $\mathbb{C}_{\vec{\varepsilon}_0}(r)$.

Through analytic continuation as before, we have for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r)$,

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

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

Apparently we still have

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

First note that there exists a positive constant L_1' such that

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

for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r)$, where r is chosen sufficiently small as above. Then from (14) and (15), we have

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

Then by Lemma 3.3, there exists a positive constant L_1'' independent of n_1, n_2 such that for any $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}(r)$, we have

$$|x_{-1,a}^{\vec{\varepsilon}} - x_{-1,b}^{\vec{\varepsilon}}| \leq L_1''\rho_1^n, \tag{19}$$

Naturally for any sequence z_{-n}^0, we have

$$p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) = x_{-1}^{\vec{\varepsilon}}(z_{-n}^{-1})\Pi(z_0)\mathbf{1}. \tag{20}$$

Together with the fact that

$$p^{\vec{\varepsilon}_0}(a_0) = \sum_i \pi(\vec{\varepsilon}_0)q(a_0|i),$$

here $\pi(\vec{\varepsilon}_0), q(a_0|i)$ are all strictly positive and bounded from above, we conclude that that there is a positive constant L_1, independent of n_1, n_2, such that

$$|p^{\vec{\varepsilon}}(a_0|a_{-n_1}^{-1}) - p^{\vec{\varepsilon}}(b_0|b_{-n_2}^{-1})| \leq L_1\rho_1^n p^{\vec{\varepsilon}_0}(a_0). \tag{21}$$

$\square$

We will need the following lemma for the proof of Theorem 1.1 as well, which can be easily proved.

Lemma 3.5. *For any $\delta > 0$, there exists $r > 0$ such that for all z_{-n}^0 and for all $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^m(r)$, we have*

$$|p^{\vec{\varepsilon}}(z_{-n}^0)| \leq (1+\delta)^n p^{\vec{\varepsilon}_0}(z_{-n}^0).$$

Proof. Note that

$$p^{\vec{\varepsilon}_0}(z_{-n}^0) = \sum_{y_{-n}^0} \pi_{y_{-n}}(\vec{\varepsilon}_0) \prod_{i=-n}^{-1} \pi_{y_i y_{i+1}}(\vec{\varepsilon}_0) \prod_{i=-n}^{0} q(z_i|y_i),$$

and

$$p^{\vec{\varepsilon}}(z_{-n}^0) = \sum_{y_{-n}^0} \pi_{y_{-n}}(\vec{\varepsilon}) \prod_{i=-n}^{-1} \pi_{y_i y_{i+1}}(\vec{\varepsilon}) \prod_{i=-n}^{0} q(z_i|y_i).$$

Notice that for any given δ, there exists r such that $|\pi_{y_{-n}}(\vec{\varepsilon})| \leq (1+\delta)\pi_{y_{-n}}(\vec{\varepsilon}_0)$ and $|\pi_{y_i y_{i+1}}(\vec{\varepsilon})| \leq (1+\delta)\pi_{y_i y_{i+1}}(\vec{\varepsilon}_0)$. The lemma then immediately follows.

□

We are now ready for the proof of Theorem 1.1.

Proof of Theorem 1.1. We only need to prove that there is a $r > 0$ such that the $H_n^{\vec{\varepsilon}}(Z)$, as $n \to \infty$, uniformly converges on $\mathbb{C}_{\vec{\varepsilon}_0}^m(r)$. Note that

$$|H_{n+1}^{\vec{\varepsilon}}(Z) - H_n^{\vec{\varepsilon}}(Z)| = \left| \int_{\mathcal{Z}_{-n-1}^0} p^{\vec{\varepsilon}}(z_{-n-1}^0) \log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) dz_{-n-1}^0 \right.$$

$$\left. - \int_{\mathcal{Z}_{-n}^0} p^{\vec{\varepsilon}}(z_{-n}^0) \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}) dz_{-n}^0 \right|$$

$$= \left| \int_{\mathcal{Z}_{-n-1}^0} p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log f^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1})) dz_{-n-1}^0 \right|.$$

Note that for sufficiently small $\delta_1' > 0$, by the mean value theorem, there exists a positive constant L_1' such that for any $\alpha, \beta \in \mathbb{C}_{\mathbb{R}^+}[\delta_1']$

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

Now fix $\vec{\varepsilon} \in \mathbb{C}_{\vec{\varepsilon}_0}^m(r)$, then by Lemma 3.4, either we have, for some $0 < \rho_1 < 1$ and some δ_1 with $(1+\delta_1)\rho_1 < 1$,

$$|p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}))|$$

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

$$\leq L_1' |p^{\vec{\varepsilon}}(z_{-n-1}^{-1})| L_1 \rho_1^n p^{\vec{\varepsilon}_0}(z_0) \leq L_1' L_1 \rho_1^n (1+\delta_1)^n p^{\vec{\varepsilon}_0}(z_{-n-1}^{-1}) p^{\vec{\varepsilon}_0}(z_0),$$

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

$$|p^{\vec{\varepsilon}}(z_{-n-1}^0)(\log p^{\vec{\varepsilon}}(z_0|z_{-n-1}^{-1}) - \log p^{\vec{\varepsilon}}(z_0|z_{-n}^{-1}))|$$

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

$$\leq L_1' |p^{\vec{\varepsilon}}(z_{-n}^{-1}) p^{\vec{\varepsilon}}(z_{-n-1}|z_{-n}^0)| L_1 \rho_1^n p^{\vec{\varepsilon}_0}(z_0)$$

$$\leq L_1' L_1 \rho_1^n (1+\delta_1)^n p^{\vec{\varepsilon}_0}(z_0) p^{\vec{\varepsilon}_0}(z_{-n-1}) p^{\vec{\varepsilon}_0}(z_{-n}^{-1}).$$

Combining all the inequalities above gives us some $L > 0$ and some $0 < \rho < 1$ such that for all $\vec{\varepsilon} \in \mathbb{C}^m_{\vec{\varepsilon}_0}(r)$,

$$|H^{\vec{\varepsilon}}_{n+1}(Z) - H^{\vec{\varepsilon}}_n(Z)| \le \int_{\mathcal{Z}^0_{-n-1}} |p^{\vec{\varepsilon}}(z^0_{-n-1})$$

$$(\log p^{\vec{\varepsilon}}(z_0|z^{-1}_{-n-1}) - \log p^{\vec{\varepsilon}}(z_0|z^{-1}_{-n}))|dz^0_{-n-1} \le L\rho^n,$$

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

□

Remark 3.6. Consider a discrete-time memoryless discrete-output (with a possibly infinite output alphabet) channel with channel transition probability $q(z|y)$. With essentially the same proof, we can show that if $q(z|y) > 0$ for all $(y, z) \in (\mathcal{Y}, \mathcal{Z})$, and

$$\sum_{z \in \mathcal{Z}} q(z|i) \log q(z|i)$$

is finite for all i, and the transition probability matrix Π of the input Markov chain Y, analytically parameterized by $\vec{\varepsilon}$, is strictly positive at $\vec{\varepsilon}_0$, then for the corresponding output discrete hidden Markov chain Z, $H(Z)$ is analytic around $\vec{\varepsilon}_0$. More precisely, all the lemmas above still hold, and one only has to replace the integral sign $\int$ in the main proof with a summation sign $\sum$.

In the case when the channel only has a finite output alphabet, analyticity of $H(Z)$ is already proven by the main result of [4]. The flow of the proof of Theorem 1.1, in fact, mainly follows from that of the proof of the main result of [4]. However, in the proof of Theorem 1.1, based on equality (13), we used the new complex Hilbert metric in a critical way ((13) does not hold for the Euclidean metric, which was employed in the proof of the main result in [4]), and we have to deal with some technical details differently.

4 Proof of Theorem 1.2

In this section, we consider a discrete-time memoryless Gaussian channel as in Theorem 1.2, which was described in Section 1. For simplicity, we assume both the order of the constraint $\mathcal{S}$ and the order of the input Markov chain Y are 1; the higher order case can reduced to order 1 case by the usual "blocking" technique.

Assume that e is the smallest positive integer such that at $\vec{\varepsilon}_0$, Π^e is strictly positive. For the Markov chain Y, define $\tilde{Y} = \{\tilde{Y}_i : i \in \mathbb{Y}\}$ to be a "blocked" process taking values in $\tilde{\mathcal{Y}} = \mathcal{Y}^e$ by

$$\tilde{Y}_i = (Y_{ei}, Y_{ei-1}, \cdots, Y_{ei-e+1});$$

correspondingly, for the hidden Markov chain Z, define $\tilde{Z} = \{\tilde{Z}_i : i \in \mathbb{Z}\}$ to be a "blocked" process taking values in $\tilde{\mathcal{Z}} = \mathcal{Z}^e$ by

$$\tilde{Z}_i = (Z_{ei}, Z_{ei-1}, \cdots, Z_{ei-e+1}).$$

It follows that $H_n(\tilde{Z})/e$ will converge to $H(Z)$ as n goes to ∞, thus to prove the smoothness of $H(Z)$, it suffices to prove that $H_n(\tilde{Z})$ and all its derivatives uniformly converge within a real neighborhood of $\vec{\varepsilon}_0$.

For each $\tilde{z} \in \tilde{\mathcal{Z}}$, define $\Pi(\tilde{z})$ by

$$\Pi(\tilde{z}) = \Pi(\tilde{z}_1)\Pi(\tilde{z}_2)\cdots\Pi(\tilde{z}_e). \tag{22}$$

Similarly as in Section 3, $\Pi(\tilde{z})$ will induce a mapping $f_{\tilde{z}} := f_{\Pi(\tilde{z})}$ from W to W. For any fixed n and $\tilde{z}_{-n}^0$, define

$$\tilde{x}_i = \tilde{x}_i(\tilde{z}_{-n}^i) = p(\tilde{y}_i = \cdot\,|\tilde{z}_i, \tilde{z}_{i-1}, \cdots, \tilde{z}_{-n}), \tag{23}$$

(here $\cdot$ represent the states of the Markov chain $\tilde{Y}$,) then $\{\tilde{x}_i^{\vec{\varepsilon}}\}$ satisfies the random dynamical system

$$\tilde{x}_{i+1} = f_{\tilde{z}_{i+1}}(\tilde{x}_i), \tag{24}$$

starting with

$$\tilde{x}_{-n-1} = \pi(\vec{\varepsilon}). \tag{25}$$

Again similarly, we have

$$p(\tilde{z}_0|\tilde{z}_{-n}) = \tilde{x}_{-1}\Pi(\tilde{z}_0)\mathbf{1}, \tag{26}$$

and

$$p(\tilde{z}_{-n}^0) = \pi(\vec{\varepsilon})\Pi(\tilde{z}_{-n})\Pi(\tilde{z}_{-n+1})\cdots\Pi(\tilde{z}_0)\mathbf{1}. \tag{27}$$

For any fixed $M > 0$, $0 < \alpha < 1$, an instance (with finite length) $\tilde{z}_{-n}^{-1}$ of the above-mentioned $\tilde{Z}$-process, is said to be (M, α)-*typical* if the number of i ($-n \leq i \leq -1$) with $|\tilde{z}_i| \leq M$ (here $|\cdot|_\infty$ denotes ℓ_∞-norm of a sequence) is bigger than αn. Let $T_n^{M,\alpha}$ denote the set of all the (M, α)-typical $\tilde{Z}$-sequences with length n.

The following lemma says that non-(M, α)-typical sequences only occur with exponentially small probability, thus we only have to focus on (M, α)-typical sequences. The proof uses the fact that the Gaussian channel transition function $q(z|y)$ (see (2)), decreases "very fast" when z goes to ∞.

Lemma 4.1. *Fix* $0 < \alpha < 1$. *For sufficiently large* M, *there exists* $0 < \rho < 1$ *such that*

$$\int_{\tilde{z}_{-n}^{-1} \not\in T_n^{M,\alpha}} p(\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^{-1} = O(\rho^n).$$

Proof. Note that for a given "blocked" hidden Markov sequence $\tilde{z}_{-n}^{-1}$,

$$p(\tilde{z}_{-n}^{-1}) = \sum_{\tilde{y}_{-n}^{-1}} p(\tilde{y}_{-n}^{-1}) p(\tilde{z}_{-n}^{-1}|\tilde{y}_{-n}^{-1}).$$

So we have

$$\int_{\tilde{z}_{-n}^{-1} \not\in T_n^{M,\alpha}} p(\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^{-1} = \int_{\tilde{z}_{-n}^{-1} \not\in T_n^{M,\alpha}} \sum_{\tilde{y}_{-n}^{-1}} p(\tilde{y}_{-n}^{-1}) p(\tilde{z}_{-n}^{-1}|\tilde{y}_{-n}^{-1}) d\tilde{z}_{-n}^{-1} = \sum_{\tilde{y}_{-n}^{-1}} p(\tilde{y}_{-n}^{-1}) \int_{\tilde{z}_{-n}^{-1} \not\in T_n^{M,\alpha}} p(\tilde{z}_{-n}^{-1}|\tilde{y}_{-n}^{-1}) d\tilde{z}_{-n}^{-1}.$$

Let μ denote the largest among all $|\mu_i|$ and let σ denote the smallest among all σ_i, then

$$\left| \int_{\tilde{z}_{-n}^{-1} \not\in T_n^{M,\alpha}} p(\tilde{z}_{-n}^{-1}|\tilde{y}_{-n}^{-1}) d\tilde{z}_{-n}^{-1} \right| \leq \left(\int_{|z|>M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{\sigma^2}} dz \right)^{(1-\alpha)n} (C_n^{(1-\alpha)n} + \cdots + C_n^n);$$

noticing that $C_n^{(1-\alpha)n} + \cdots + C_n^n \leq 2^n$, we then have

$$\left| \int_{z_{-n}^{-1} \notin T_{M,\alpha}} p(z_{-n}^{-1}) dz_{-n}^{-1} \right| \leq |\mathcal{Y}|^n 2^n \left(\int_{|z|>M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{\sigma^2}} dz \right)^{(1-\alpha)n}.$$

It then follows that for sufficiently large M, there exists $0 < \rho < 1$ such that

$$\int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}} p(\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^{-1} = O(\rho^n).$$

□

The above lemma says that non-(M, α)-typical sequences only occur with exponentially small probability, thus we only have to focus on (M, α)-typical sequences. More precisely, define

$$H_n^{M,\alpha}(\tilde{Z}) = \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n}^0) \log p(\tilde{z}_0 | \tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^0.$$

Note that for any z_i^j, we have

$$\min_l \sum_k \frac{\pi_{l,k}}{\sqrt{2\pi}\sigma_k} e^{-(z_j-k)^2/(2\sigma_k^2)} \leq p(z_j | z_i^{j-1})$$

$$= x_{j-1} \Pi_{z_j} \mathbf{1} \leq \max_l \sum_k \frac{\pi_{l,k}}{\sqrt{2\pi}\sigma_k} e^{-(z_j-k)^2/(2\sigma_k^2)}.$$

It then follows from

$$p(\tilde{z}_0 | \tilde{z}_{-n}^{-1}) \log p(\tilde{z}_0 | \tilde{z}_{-n}^{-1})$$

$$= \prod_{i=-e+1}^{0} p(z_i | z_{-en-e+1}^{i-1}) \sum_{i=-e+1}^{0} \log p(z_i | z_{-en-e+1}^{i-1})$$

that $|p(z_0 | z_{-n}^{-1}) \log p(z_0 | z_{-n}^{-1})|$ is upper bounded by an integrable function $g(\tilde{z}_0)$, which is independent of $\tilde{z}_{-n}^{-1}$. It then follows from Lemma 4.1 that there exists $0 < \rho < 1$ such that

$$|H_n^{M,\alpha}(\tilde{Z}) - H_n(\tilde{Z})| = \left| \int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n}^0) \log p(\tilde{z}_0 | \tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^0 \right|$$

$$= \left| \int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n}^{-1}) p(\tilde{z}_0 | \tilde{z}_{-n}^{-1}) \log p(\tilde{z}_0 | \tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^0 \right|$$

$$\leq \left| \int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^{-1} \int_{\tilde{z}_0} g(\tilde{z}_0) d\tilde{z}_0 \right| = O(\rho^n),$$

which implies that, like $H_n(\tilde{Z})$, $H_n^{M,\alpha}(\tilde{Z})$ converges to $H(\tilde{Z})$, as $n \to \infty$.

To prove smoothness of $H(Z)$ at $\vec{\varepsilon}_0$, it suffices to prove that $H_n^{M,\alpha}(\tilde{Z})$ and all its derivatives uniformly converge on a neighborhood of $\vec{\varepsilon}_0$. In the following, In the following, we use

$\alpha(z_i^j) = \hat{O}(\beta(z_i^j))$ to denote that there exist positive constants C, K, which are independent of z_i^j, such that

$$|\alpha(z_i^j)| \le C\beta(z_i^j)^K.$$

For any smooth function f of $\vec{\varepsilon}$ and $\vec{n} = (n_1, n_2, \cdots, n_m) \in \mathbb{Z}^m$, define

$$f^{(\vec{n})} = \frac{\partial^{|\vec{n}|} f}{\partial \varepsilon_1^{n_1} \partial \varepsilon_2^{n_2} \cdots \partial \varepsilon_m^{n_m}},$$

here $|\vec{n}|$ denotes the order of the $\vec{n}$-th derivative of f with respect to $\vec{\varepsilon}$, and is defined as

$$|\vec{n}| = n_1 + n_2 + \cdots + n_m.$$

We say $\vec{l} \preceq \vec{n}$, if every component of $\vec{l}$ is less or equal to the corresponding one of $\vec{n}$. For any $\vec{l} = (l_1, l_2, \cdots, l_m) \in \mathbb{Z}^m$, $\vec{l}!$ is defined as $\vec{l}! = \prod_{i=1}^m l_i!$. For any $\vec{l} \preceq \vec{n}$, define $C_{\vec{n}}^{\vec{l}} = \frac{\vec{n}!}{\vec{l}!(\vec{n}-\vec{l})!}$.

Proof of Theorem 1.2. In the following we shall prove that $H_n^{M,\alpha}(\tilde{Z})$ and all its derivatives with respect to $\vec{\varepsilon}$ uniformly converge within certain neighborhood of $\vec{\varepsilon}_0$, thus implying smoothness of $H_n^{M,\alpha}(\tilde{Z})$. Although the convergence of $H_n^{M,\alpha}(\tilde{Z})$ and its derivatives can be proven through the same argument at once, we first prove the convergence of $H_n^{M,\alpha}(\tilde{Z})$ only for illustrative purpose.

Using the inequalities

$$|\log\alpha - \log\beta| \le \max\left(|(\alpha-\beta)/\beta)|, |(\beta-\alpha)/\alpha|\right), \tag{28}$$

we have

$$\begin{aligned}
|H_n^{M,\alpha}(\tilde{Z}) - H_{n+1}^{M,\alpha}(\tilde{Z})| &= \left| \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n}^0) \log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^0 - \int_{\tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1}) d\tilde{z}_{-n-1}^0 \right| \\
&= \left| \left(\int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} + \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \notin T_{n+1}^{M,\alpha}, \tilde{z}_0} \right) - p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n-1}^0 \right. \\
&\quad \left. - \left(\int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} + \int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} \right) - p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1}) d\tilde{z}_{-n-1}^0 \right| \\
&\le \left| \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) (\log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) - \log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})) d\tilde{z}_{-n-1}^0 \right| \\
&+ \left| \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \notin T_{n+1}^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n-1}^0 \right| + \left| \int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1}) d\tilde{z}_{-n-1}^0 \right| \\
&\le \left| \int_{\tilde{z}_{-n}^{-1} \in T_{M,\alpha}^n, \tilde{z}_{-n-1}^{-1} \in T_{M,\alpha}^n, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) \max\left(\left| \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) - p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \right|, \left| \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) - p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \right| \right) d\tilde{z}_{-n-1}^0 \right| \\
&+ \left| \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \notin T_{n+1}^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n-1}^0 \right| + \left| \int_{\tilde{z}_{-n}^{-1} \notin T_n^{M,\alpha}, \tilde{z}_{-n-1}^{-1} \in T_{n+1}^{M,\alpha}, \tilde{z}_0} -p(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1}) d\tilde{z}_{-n-1}^0 \right|.
\end{aligned} \tag{29}$$

We first show that the second and the third terms above are $O(\rho^n)$ for some $0<\rho<1$. Similarly as before, one checks that

$$|p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})\log p(\tilde{z}_0|\tilde{z}_{-n}^{-1})|$$

is upper bounded by an integrable function $g_0(\tilde{z}_0)$, which is independent of $\tilde{z}_{-n}^{-1}$. So we have, by Lemma 4.1, for some $0<\rho<1$,

$$\left|\int_{\tilde{z}_{-n}^{-1}\in T_{M,\alpha}^n,\tilde{z}_{-n-1}^{-1}\notin T_{M,\alpha}^{n+1},\tilde{z}_0} -p(\tilde{z}_{-n-1}^0)\log p(\tilde{z}_0|\tilde{z}_{-n}^{-1})d\tilde{z}_{-n-1}^0\right|$$

$$= \left|\int_{\tilde{z}_{-n}^{-1}\in T_{M,\alpha}^n,\tilde{z}_{-n-1}^{-1}\notin T_{M,\alpha}^{n+1},\tilde{z}_0} -p(\tilde{z}_{-n-1}^{-1})p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})\log p(\tilde{z}_0|\tilde{z}_{-n}^{-1})d\tilde{z}_{-n-1}^0\right|$$

$$\leq \left|\int_{\tilde{z}_{-n-1}^{-1}\notin T_{M,\alpha}^{n+1}} -p(\tilde{z}_{-n-1}^{-1})d\tilde{z}_{-n-1}^{-1}\int_{\tilde{z}_0} g_0(\tilde{z}_0)d\tilde{z}_0\right| = O(\rho^n).$$

Similarly $|p(\tilde{z}_{-n-1}|\tilde{z}_{-n}^{-1})|$, $|p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})\log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})|$ are upper bounded integrable functions $g_1(\tilde{z}_{-n-1})$, $g_2(\tilde{z}_0)$, respectively. So we have, again by Lemma 4.1, for some $0<\rho<1$,

$$\left|\int_{\tilde{z}_{-n}^{-1}\notin T_{M,\alpha}^n,\tilde{z}_{-n-1}^{-1}\in T_{M,\alpha}^{n+1},\tilde{z}_0} -p(\tilde{z}_{-n}^{-1})p(\tilde{z}_{-n-1}|\tilde{z}_{-n}^{-1})p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})\log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})d\tilde{z}_{-n-1}^0\right|$$

$$\leq \int_{\tilde{z}_{-n}^{-1}\notin T_{M,\alpha}^n} p(\tilde{z}_{-n}^{-1})d\tilde{z}_{-n}^{-1}\int_{\tilde{z}_{-n-1}} g_1(\tilde{z}_{-n-1})d\tilde{z}_{-n-1}\int_{\tilde{z}_0} g_2(\tilde{z}_0)d\tilde{z}_0 = O(\rho^n).$$

To show the first term is also $O(\rho^n)$, we need to estimate $|\tilde{x}_i^a-\tilde{x}_i^b|$ where we rewrite $\tilde{x}_i(\tilde{z}_{-n}^i), \tilde{x}_i(\tilde{z}_{-n-1}^i)$ as $\tilde{x}_i^a, \tilde{x}_i^b$, respectively. Note that for $|\tilde{z}_i|_\infty \leq M$, there is a $0<\rho_1<1$ such that

$$d_H(\tilde{x}_i^a,\tilde{x}_i^b) \leq \rho_1 d_H(\tilde{x}_{i-1}^a,\tilde{x}_{i-1}^b),$$

while otherwise trivially we have

$$d_H(\tilde{x}_i^a,\tilde{x}_i^b) \leq d_H(\tilde{x}_{i-1}^a,\tilde{x}_{i-1}^b).$$

Then for any sequence $\tilde{z}_{-n}^{-1}\in T_n^{M,\alpha}$, let i_0 denote the smallest index such that $|\tilde{z}_{i_0}|_\infty \leq M$, then we have

$$d_H(\tilde{x}_{-1}^a,\tilde{x}_{-1}^b) \leq \rho_1^{\alpha n-1} d_H(\tilde{x}_{i_0}^a,\tilde{x}_{i_0}^b),$$

which implies that there exists $0<\rho<1$ such that $|\tilde{x}_{-1}^a-\tilde{x}_{-1}^b| \leq O(\rho^n)$. It then follows that there exists $0<\rho<1$ such that $|p(\tilde{z}_0|\tilde{z}_{-n}^{-1})-p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})| \leq \rho^n g_3(\tilde{z}_0)$, where $g_3(\tilde{z}_0)$ is an integrable function of $\tilde{z}_0$. This, together with the fact that $p(\tilde{z}_{-n-1}|\tilde{z}_{-n}^0)$ is upper bounded by $g_4(\tilde{z}_{-n-1})$, which is an integrable function of $\tilde{z}_{-n-1}$, will establish the case when $|\vec{l}|=0$, thus implying that $H_n^{M,\alpha}(Z)$ uniformly converge to $H(\tilde{Z})$.

Apply multivariate Faa Di Bruno formula [2, 6] to the function $f(y)=\log y$, we have for $\vec{l}$ with $|\vec{l}| \neq 0$,

$$f(y)^{(\vec{l})} = \sum D(\vec{a}_1,\vec{a}_2,\cdots,\vec{a}_k)(y^{(\vec{a}_1)}/y)(y^{(\vec{a}_2)}/y)\cdots(y^{(\vec{a}_k)}/y),$$

where the summation is over the set of unordered sequences of non-negative vectors $\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_k$ with $\vec{a}_1 + \vec{a}_2 + \cdots + \vec{a}_k = \vec{l}$ and $D(\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_k)$ is the corresponding coefficient. So for any $\vec{m}$, we have

$$(H_n^{M,\alpha})^{(\vec{m})}(Z) = \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_0} \sum_{\vec{l} \preceq \vec{m}} -C_{\vec{m}}^{\vec{l}} p^{(\vec{m}-\vec{l})}(\tilde{z}_{-n}^0)(\log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}))^{(\vec{l})} d\tilde{z}_{-n}^0$$

$$= \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_0} \sum_{|\vec{l}| \neq 0, \vec{l} \preceq \vec{m}} \sum_{\vec{a}_1 + \vec{a}_2 + \cdots + \vec{a}_k = \vec{l}} -C_{\vec{m}}^{\vec{l}} D(\vec{a}_1, \cdots, \vec{a}_k) p^{(\vec{m}-\vec{l})}(\tilde{z}_{-n}^0) \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \cdots \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} d\tilde{z}_{-n}^0$$

$$+ \int_{\tilde{z}_{-n}^{-1} \in T_n^{M,\alpha}, \tilde{z}_0} -p^{(\vec{m})}(\tilde{z}_{-n}^0) \log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^0.$$

Notice that there exists a positive constant C_m such that for any $\tilde{z}_{-n}^0$, we have

$$|p^{(m)}(\tilde{z}_{-n}^0)| \leq C_m n^{|\vec{m}|} p(\tilde{z}_{-n}^0).$$

Again using the inequalities (28) and with a parallel argument through replacing $p(\tilde{z}_{-n}^0), p(\tilde{z}_{-n-1}^0)$ by $p^{(m)}(\tilde{z}_{-n}^0), p^{(m)}(\tilde{z}_{-n-1}^0)$, respectively, we can show that

$$\left| \int_{\tilde{z}_{-n}^{-1} \in T_n^{\alpha}, \tilde{z}_0} -p^{(\vec{m})}(\tilde{z}_{-n}^0) \log p(\tilde{z}_0|\tilde{z}_{-n}^{-1}) d\tilde{z}_{-n}^0 - \int_{\tilde{z}_{-n-1}^{-1} \in T_{n+1}^{\alpha}, \tilde{z}_0} -p^{(\vec{m})}(\tilde{z}_{-n-1}^0) \log p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1}) d\tilde{z}_{-n-1}^0 \right| = O(\rho^n).$$

And using the identity

$$\alpha_1\alpha_2\cdots\alpha_n - \beta_1\beta_2\cdots\beta_n = (\alpha_1-\beta_1)\alpha_2\cdots\alpha_n + \beta_1(\alpha_2-\beta_2)\alpha_3\cdots\alpha_n + \cdots + \beta_1\cdots\beta_{n-1}(\alpha_n-\beta_n),$$

we have

$$\left| \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \cdots \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} - \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \cdots \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \right|$$

$$\leq \left| \left(\frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} - \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \right) \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_2)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \cdots \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \right|$$

$$+ \left| \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \left(\frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_2)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} - \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_2)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \right) \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_3)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \cdots \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} \right| + \cdots$$

$$+ \left| \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_1)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \cdots \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_{k-1})}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \left(\frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} - \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_k)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \right) \right|.$$

Now apply the inequality

$$\left| \frac{\beta_1}{\alpha_1} - \frac{\beta_2}{\alpha_2} \right| = \left| \frac{\beta_1}{\alpha_1} - \frac{\beta_1}{\alpha_2} + \frac{\beta_1}{\alpha_2} - \frac{\beta_2}{\alpha_2} \right| \leq |\beta_1/(\alpha_1\alpha_2)||\alpha_1 - \alpha_2| + |1/(\alpha_2)||\beta_1 - \beta_2|,$$

we have for any feasible i,

$$\left| \frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_i)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})} - \frac{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_i)}}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})} \right|$$

$$\leq \left|\frac{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_i)}}{p(\tilde{z}_0|\tilde{z}_{-n}^{-1})p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})}\right| |p(\tilde{z}_0|\tilde{z}_{-n}^{-1})-p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})|+\left|\frac{1}{p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})}\right| |p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_i)}-p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_i)}|.$$

To estimate $|p(\tilde{z}_0|\tilde{z}_{-n}^{-1})^{(\vec{a}_i)} - p(\tilde{z}_0|\tilde{z}_{-n-1}^{-1})^{(\vec{a}_i)}|$, we shall first estimate the difference between the derivatives of $\tilde{x}_i^a$ and $\tilde{x}_i^b$.

In what follows, we prove that for any $\vec{k}$ with $|\vec{k}| = 1$, the $\vec{k}$-th order derivatives of $\tilde{x}_i^a, \tilde{x}_i^b$ get exponentially close with respect to n along any (M, α)-typical sequence $\tilde{z}_{-n}^{-1}$ (up to certain integrable functions). Note that

$$\tilde{x}_{i+1,a}^{(\vec{k})} = \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,a}) + \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,a})\tilde{x}_{i,a}^{(\vec{k})}.$$

Similarly, we also have

$$\tilde{x}_{i+1,b}^{(\vec{k})} = \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,b}) + \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})\tilde{x}_{i,b}^{(\vec{k})}.$$

Note that

$$\left|\frac{\partial f_{\tilde{z}}}{\partial x}(\vec{\varepsilon}, x)\right|, \left|\frac{\partial f_{\tilde{z}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, x)\right| = \hat{O}(e^{|\tilde{z}|_1}),$$

where $|\cdot|_1$ denotes the ℓ_1 norm and the constants in $\hat{O}(e^{|\tilde{z}|_1})$ is independent of all $\tilde{z} \in \tilde{\mathcal{Z}}$ and all $x \in \mathcal{W}$. It then follows from an inductive argument, using the iterations above, that for any $\tilde{z}_{-n}^{-1}$,

$$|\tilde{x}_{-1,a}^{(\vec{k})}|, |\tilde{x}_{-1,b}^{(\vec{k})}| = \hat{O}(ne^{|\tilde{z}_{-n}|_1+|\tilde{z}_{-n+1}|_1+\cdots+|\tilde{z}_{-1}|_1}),$$

where the constants in $\hat{O}$ only depend on $\vec{k}$.

Take the difference of the iterative equations, we then have

$$\begin{aligned}
\tilde{x}_{i+1,a}^{(\vec{k})} - \tilde{x}_{i+1,b}^{(\vec{k})} &= \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,a}) - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,b}) + \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,a})\tilde{x}_{i,a}^{(\vec{k})} - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})\tilde{x}_{i,b}^{(\vec{k})} \\
&= \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,a}) - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,b}) + \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,a})\tilde{x}_{i,a}^{(\vec{k})} - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})\tilde{x}_{i,a}^{(\vec{k})} \\
&\quad + \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})\tilde{x}_{i,a}^{(\vec{k})} - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})\tilde{x}_{i,b}^{(\vec{k})} \\
&= \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})(\tilde{x}_{i,a}^{(\vec{k})} - \tilde{x}_{i,b}^{(\vec{k})}) + g_{i+1}(\tilde{z}_{-n}^{i+1}),
\end{aligned}$$

where we defined

$$g_{i+1}(\tilde{z}_{-n}^{i+1}) = \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,a}) - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial \vec{\varepsilon}}(\vec{\varepsilon}, \tilde{x}_{i,b}) + \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,a})\tilde{x}_{i,a}^{(\vec{k})} - \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})\tilde{x}_{i,a}^{(\vec{k})}.$$

Applying the mean value theorem, one checks that

$$g_{i+1}(\tilde{z}_{-n}^{i+1}) = \hat{O}(e^{|\tilde{z}_{-n}|_1+|\tilde{z}_{-n+1}|_1+\cdots+|\tilde{z}_{-1}|_1})(\tilde{x}_{i,a} - \tilde{x}_{i,b}).$$

Letting the dynamical system evolve, we obtain

$$\tilde{x}_{-1,a}^{(\vec{k})} - \tilde{x}_{-1,b}^{(\vec{k})} = \prod_{i=-n}^{-2} \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})(\tilde{x}_{-n,a}^{(\vec{k})} - \tilde{x}_{-n,b}^{(\vec{k})}) + \prod_{i=-n+1}^{-2} \frac{\partial f_{\tilde{z}_{i+1}}}{\partial x}(\vec{\varepsilon}, \tilde{x}_{i,b})g_{-n+1}(\tilde{z}_{-n}^{-n+1})$$

$$+\cdots+(\frac{\partial f_{\tilde z_{-1}}}{\partial x}(\vec\varepsilon,\tilde x_{-2,b}))g_{-2}(\tilde z_{-n}^{-2})+g_{-1}(\tilde z_{-n}^{-1}).$$

Consider any individual term above,

$$\prod_{i=j}^{-2}\frac{\partial f_{\tilde z_{i+1}}}{\partial x}(\vec\varepsilon,\tilde x_{i,b})g_j(\tilde z_{-n}^{j}).$$

When $j+n$ is larger than $2(1-\alpha)n$, $|\tilde x_{j-1,a}-\tilde x_{j-1,b}|=O(\rho_1^n)$ for some $0<\rho_1<1$; while $j+n$ is smaller than $2(1-\alpha)n$, it follows from the fact that $f_{\tilde z_{-1}}\circ\cdots\circ f_{\tilde z_{-n+1}}\circ f_{\tilde z_{-n}}$ is a contraction mapping under Hilbert metric and Lemma 3.3 that for some $0<\rho_1<1$, we have

$$\prod_{i=j}^{-2}\frac{\partial f_{\tilde z_{i+1}}}{\partial x}(\vec\varepsilon,\tilde x_{i,b})\le\rho_1^n\hat O(|\tilde z_{j+1}|_1+\cdots+|\tilde z_{-1}|_1).$$

In any case, there exists $0<\rho_1<1$ such that

$$\prod_{i=j}^{-2}\frac{\partial f_{\tilde z_{i+1}}}{\partial x}(\vec\varepsilon,\tilde x_{i,b})g_j(\tilde z_{-n}^{j})=\rho_1^n\hat O(e^{|\tilde z_{-n-1}|_1+\cdots+|\tilde z_{-1}|_1}),$$

which implies that there exists $0<\rho_1<1$ such that

$$|\tilde x_{-1,a}^{(\vec k)}-\tilde x_{-1,b}^{(\vec k)}|=\rho_1^n\hat O(e^{|\tilde z_{-n-1}|_1+\cdots+|\tilde z_{-1}|_1}).$$

Most importantly, one checks that there exists $0<\rho_1<1$ such that

$$\int_{\tilde z_{-n}^{-1}\in T_n^{M,\alpha},\tilde z_{-n-1}^{-1}\in T_n^{M,\alpha},\tilde z_0}p^{(m-l)}(\tilde z_{-n-1}^{0})\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_1)}}{p(\tilde z_0|\tilde z_{-n-1}^{-1})}\cdots\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_{i-1})}}{p(\tilde z_0|\tilde z_{-n-1}^{-1})}$$

$$\left|\frac{p(\tilde z_0|\tilde z_{-n}^{-1})^{(\vec a_i)}}{p(\tilde z_0|\tilde z_{-n}^{-1})p(\tilde z_0|\tilde z_{-n-1}^{-1})}\right||p(\tilde z_0|\tilde z_{-n}^{-1})-p(\tilde z_0|\tilde z_{-n-1}^{-1})|\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_{i+1})}}{p(\tilde z_0|\tilde z_{-n}^{-1})}\cdots\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_m)}}{p(\tilde z_0|\tilde z_{-n}^{-1})}d\tilde z_{-n-1}^{0}\le\rho_1^n,$$

and

$$\int_{\tilde z_{-n}^{-1}\in T_n^{M,\alpha},\tilde z_{-n-1}^{-1}\in T_n^{M,\alpha},\tilde z_0}p^{(m-l)}(\tilde z_{-n-1}^{0})\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_1)}}{p(\tilde z_0|\tilde z_{-n-1}^{-1})}\cdots\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_{i-1})}}{p(\tilde z_0|\tilde z_{-n-1}^{-1})}$$

$$\left|\frac{1}{p(\tilde z_0|\tilde z_{-n-1}^{-1})}\right||p(\tilde z_0|\tilde z_{-n}^{-1})^{(\vec a_i)}-p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_i)}|\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_{i+1})}}{p(\tilde z_0|\tilde z_{-n}^{-1})}\cdots\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_m)}}{p(\tilde z_0|\tilde z_{-n}^{-1})}d\tilde z_{-n-1}^{0}\le\rho_1^n,$$

where we used the facts that there exists $C_{|\vec m-\vec l|}$ such that

$$|p^{(\vec m-\vec l)}(\tilde z_{-n-1}^{0})|\le C_{|\vec m-\vec l|}p(\tilde z_{-n-1}^{0});$$

and there exists $0<\rho_2<1$ such that

$$|p(\tilde z_0|\tilde z_{-n}^{-1})^{(\vec a_i)}-p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_i)}|=\rho_2^n p(\tilde z_0)\hat O(e^{|\tilde z_{-n-1}|+\cdots+|\tilde z_0}|);$$

and

$$\frac{p(\tilde z_0|\tilde z_{-n-1}^{-1})^{(\vec a_i)}}{p(\tilde z_0|\tilde z_{-n}^{-1})}=\hat O(e^{|\tilde z_{-n}|_1+\cdots+|\tilde z_{-1}|_1}).$$

A completely parallel yet more tedious argument can be applied to higher derivatives to establish that for any $\vec{m}$, we have

$$|(H_n^{M,\alpha})^{(\vec{m})}(\tilde{Z}) - (H_n^{M,\alpha})^{(\vec{m})}(\tilde{Z})| = O(\rho^n),$$

which implies that $H_n^{M,\alpha}(\tilde{Z})$ and its derivatives with respect to $\vec{\varepsilon}$ uniformly converge to the function $H(\tilde{Z})$ and correspondingly its derivatives on some neighborhood of $\vec{\varepsilon}_0$, which implies the smoothness of $H(\tilde{Z})$.

□

Remark 4.2. Unlike Theorem 1.1, the complex Hilbert metric can not be applied because of the possible zero entries of Π.

Problem 4.3. *Can Blackwell's measure be generalized to continuous-state hidden Markov chains? It seems like doable when the transition probability matrix of the Markov chain is strictly positive, since we have strict contract along any hidden Markov chain sequence.*

For analyticity, what about treating entropy rate as a function of channel parameters. It appears that given the positivity assumptions, entropy rate should be analytic with respect to channel parameters too. For this, we may have to prove perturbation of the channel parameters will uniformly act on the real simplex.

What about asymptotic behavior of entropy rate? Remember BSC is an extreme version of Gaussian channel.

References

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

[2] G. Constantine and T. Savits. A Multivariate Faa Di Bruno Formula With Applications. *Transactions of the American Mathematical Society*, Vol. 348, No. 2., Feb, 1996, pp. 503-520.

[3] S. Ihara. *Information theory for continuous systems.* World Scientific, 1993.

[4] G. Han and B. Marcus. Analyticity of entropy rate of hidden Markov chains. *IEEE Transactions on Information Theory*, Volume 52, Issue 12, December, 2006, pages: 5251-5266.

[5] G. Han and B. Marcus and Y. Peres. A note on a complex Hilbert metric with application to domain of analyticity for entropy rate of hidden Markov processes. To appear in *Entropy of Hidden Markov Processes and Connections to Dynamical Systems.*

[6] R. Leipnik and T. Reid. Multivariable Faa di Bruno Formulas. *Electronic Proceedings of the Ninth Annual International Conference on Technology in Collegiate Mathematics*, http://archives.math.utk.edu/ICTCM/EP-9.html#C23 .

[7] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.

[8] B. Marcus, R. Roth and P. Siegel. Constrained Systems and Coding for Recording Channels. Chap. 20 in *Handbook of Coding Theory* (eds. V. S. Pless and W. C. Huffman), Elsevier Science, 1998.

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