Analyticity of Entropy Rate in Families of Hidden Markov Chains (II) *

Guangyue Han, Brian Marcus
Department of Mathematics
University of British Columbia
Vancouver, B.C., V6T 1Z2
e-mail: ghan, marcus@math.ubc.ca

January 4, 2008

Abstract

We give relaxed sufficient conditions (compared to [2]) for analyticity of the entropy rate of a hidden Markov chain. Several examples of the relaxed conditions are given. A general principle to calculate the domain of analyticity is stated. An example is given to estimate the radius of convergence for the entropy rate. Finally, we prove a "stabilizing" property of "Black Hole" case, which suggests that one can explicitly compute the derivatives and obtain an explicit Taylor series in certain cases, generalizing the results in [13, 12].

1 Introduction

As in [2], let $Y = \{Y_{-\infty}^{\infty}\}$ be a stationary Markov chain with a finite state alphabet $\{1, 2, \cdots, B\}$. A function $Z = \{Z_{-\infty}^{\infty}\}$ of the Markov chain Y with the form $Z = \Phi(Y)$ is called a hidden Markov chain; here Φ is a finite valued function defined on $\{1, 2, \cdots, B\}$, taking values in $\{1, 2, \cdots, A\}$. Let Δ denote the probability transition matrix for Y; it is well known that the entropy rate $H(Y)$ of Y can be analytically expressed using the stationary vector of Y and Δ. Let W be the simplex, comprising the vectors

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

and let W_a be all $w \in W$ with $w_i = 0$ for $\Phi(i) \neq a$. For $a \in A$, let Δ_a denote the $B \times B$ matrix such that $\Delta_a(i,j) = \Delta(i,j)$ for i with $\Phi(i) = a$, and $\Delta_a(i,j) = 0$ otherwise. For $a \in A$, define the function r_a on W by

$$r_a(w) = w\Delta_a\mathbf{1}.$$

*This paper is a continuation of [2], thus we follow the notation in [2]. An earlier version of this paper was submitted to ISIT 2006.

If Y is irreducible, the entropy rate $H(Z)$ of Z is the integral of the function $-\sum_{a \in A} r_a(w) \log r_a(w)$ with respect to a measure, called the Blackwell measure, on W. This measure is defined as the limiting distribution $p(y_0 = \cdot | z_{-\infty}^0)$.

Let $\vec{s}$ be the stationary vector of the Markov chain; for each symbol a of Z and each $w \in W$ such that $r_a(w) > 0$, let

$$f_a(w) = w\Delta_a / r_a(w).$$

Then the support of the Blackwell measure is contained in the limit set of $\{(f_{a_n} \circ f_{a_{n-1}} \circ \cdots \circ f_{a_0})(\vec{s})\}$ (see [2], [1]); in the examples of this paper, the support actually coincides with the limit.

Recently there has been a great deal of work on the entropy rate of a hidden Markov chain [7, 3, 8, 13, 4, 12]. See also closely related work [6, 10, 11].

In Section 2, we relax the conditions given in [2] for analyticity of the entropy rate of a hidden Markov chain. The conditions in Theorem 1.1 (and Remark 3.1) and Theorem 4.7 in [2] fit these conditions; these results are subsumed in Theorem 2.1 of this paper. Several examples are given.

In Section 3, we first state a general principle to determine a domain of analyticity for the entropy rate. As an example, we determine a lower bound on the radius of convergence of the entropy rate for the special case when a binary Markov chain is corrupted by binary symmetric noise.

In Section 4, a "stabilizing" property of the derivatives of the entropy rate for the "Black Hole" case is discussed. According to this property, one can explicitly calculate the derivatives of the entropy rate for this case.

2 Relaxed Conditions

In this section, we consider analyticity of the entropy rate of Z when Δ has a simple maximum eigenvalue 1; this implies that Δ has a unique stationary vector, and so the Markov chain is uniquely defined. For w in W_b, let $f_a'(w)$ denote the first order derivative of the mapping f_a from W_b to W_a at w (restricted to the subspace spanned by directions parallel to the simplex W_b) and let $\|\cdot\|$ denote the Euclidean norm of a linear mapping. We say that $\{f_a : a \in A\}$ is *eventually contracting* at $w \in W_b$ if there exists n such that for any $a_0, a_1, \cdots, a_n \in A$, $\|(f_{a_n} \circ f_{a_{n-1}} \circ \cdots \circ f_{a_0})'(w)\|$ is strictly less than 1. Let L denote the limit set of $\{(f_{a_n} \circ f_{a_{n-1}} \circ \cdots \circ f_{a_0})(\vec{s})\}$.

Theorem 2.1. *If at $\Delta = \hat{\Delta}$,*

1. *1 is a simple eigenvalue for $\hat{\Delta}$,*
2. *For every a and all w in L, $r_a(w) > 0$,*
3. *For every b, $\{f_a : a \in A\}$ is eventually contracting at all w in the convex hull of the intersection of L and W_b,*

then $H(Z)$ is analytic at $\Delta = \hat{\Delta}$.

Proof. Let $\mathcal{X}$ denote the right infinite shift space $\{a_0^\infty : a_i \in A\}$. Let L_δ denote the δ-neighborhood of L. Choose $\delta > 0$ so small that at $\hat{\Delta}$, we have:

- For every $a \in A$ and all w in L_δ, $r_a(w) > 0$
- For every b, $\{f_a : a \in A\}$ is eventually contracting at all w in the convex hull of the intersection of L_δ and W_b.

For simplicity we may assume that for each a, f_a is contracting instead of eventually contracting. For any $c_0^\infty \in \mathcal{X}$, there exists n such that $\{(f_{c_n} \circ f_{c_{n-1}} \circ \cdots \circ f_{c_0})(\vec{s})\} \in L_\delta$. Let $\Omega^n_{c_0^\infty}$ denote the cylinder set $\{a_0^\infty : a_0 = c_0, a_1 = c_1, \cdots, a_n = c_n\}$. Since $\{f_a : a \in A\}$ is contracting, we conclude that for any $a_0^\infty \in \Omega^n_{c_0^\infty}$ and for $m \geq n$, $\{(f_{a_m} \circ f_{a_{m-1}} \circ \cdots \circ f_{a_0})(\vec{s})\} \in L_\delta$. By the compactness of $\mathcal{X}$, we can find finitely many such cylinder sets to cover $\mathcal{X}$. Consequently we can find n such that for $m \geq n$ and any $a_0^\infty \in \mathcal{X}$, we have $\{(f_{a_m} \circ f_{a_{m-1}} \circ \cdots \circ f_{a_0})(\vec{s})\} \in L_\delta$. So similar to the proof of Theorem 1.1 in [2], one can use the contraction (along any symbolic sequence z_{-n}^0) to extend $H_n(Z) = H(Z_0|Z_{-n}^{-1})$ from real to complex and prove the uniform convergence of $H_n(Z)$ to $H(Z)$. □

Remark 2.2.

(1) If $\hat{\Delta}$ has a strictly positive column (or more generally, there is a j such that for all i, there exists n such that $\hat{\Delta}^n_{ij} > 0$), then condition 1 holds by Perron-Frobenius theory.

(2) If for each symbol a, $\hat{\Delta}_a$ is row allowable (i.e., no row is all zero), then $r_a(w) > 0$ for all $w \in W$ and so condition 2 holds.

In Theorem 1.1, and more generally Remark 3.1, of [2], it is proven that $H(Z)$ is analytic assuming a positivity assumption: namely, that for each a, Δ_a is not the zero matrix and each of its columns is either strictly positive or all zero. Theorem 2.1 relaxes the positivity assumption. Indeed given the positivity assumption, by Remark 2.2, condition 1 and 2 hold. For condition 3, first observe that L is contained in $\cup_b f_b(W)$. Using the equivalence of the Euclidean metric and the Hilbert metric, the proof in [2] shows that for every b, $\{f_a : a \in A\}$ is eventually contracting on $f_b(W)$, which is a convex set containing the intersection of L and W_b.

In Theorem 4.7 of [2], necessary and sufficient conditions are given for analyticity of entropy rate of binary hidden Markov chains with unambiguous symbol 0. Theorem 2.1 generalizes the sufficiency part of this result. Indeed the conditions of Theorem 4.7 of [2] imply the conditions of Theorem 2.1. Condition 1 follows from the fact that Δ is assumed irreducible. For conditions 2 and 3, one first notes that the image of f_0 is a single point W_0, and the f_1-orbit of W_0 and f_1-orbit of $\vec{s}$ converge to the same point p_1. It follows that L is the union of W_0, the f_1-orbit of W_0 and p_1. The conditions in Theorem 4.7 of [2] imply that $r_a > 0$ on L (i.e., condition 2 holds) and that for sufficiently large n, f_1^n is contracting on the convex hull of the intersection of L and W_1 (so, condition 3 holds).

Theorem 2.1 also applies to many cases not covered by the results of [2]. For instance, suppose that some column of $\hat{\Delta}$ is strictly positive and each $\hat{\Delta}_a$ is row allowable. By Remark 2.2, Theorem 2.1 applies whenever we can guarantee condition 3. For this, it is sufficient to check that for each a, b, f_a is a contraction, with respect to the Euclidean metric, on the convex hull of the intersection of L with each W_b. This can be done by explicitly computing derivatives.

Example 2.3. Consider a hidden Markov chain Z defined by :

$$\hat{\Delta} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix},$$

with $\Phi(1) = \Phi(2) = 0$ and $\Phi(3) = \Phi(4) = 1$. We assume that some column of $\hat{\Delta}$ is strictly positive and both $\hat{\Delta}_0$ and $\hat{\Delta}_1$ are row allowable.

Parameterize W_0 by $(y, 1-y, 0, 0)$ and parameterize W_1 by $(0, 0, y, 1-y)$ (with $y \in [0, 1]$). We can explicitly compute the derivatives of f_0 and f_1 with respect to y:

$$f_0'|_{(y,1-y,0,0)} = \frac{a_{11}a_{22} - a_{12}a_{21}}{((a_{11} + a_{12} - a_{21} - a_{22})y + a_{21} + a_{22})^2},$$

$$f_0'|_{(0,0,y,1-y)} = \frac{a_{31}a_{42} - a_{32}a_{41}}{((a_{31} + a_{32} - a_{41} - a_{42})y + a_{41} + a_{42})^2},$$

$$f_1'|_{(y,1-y,0,0)} = \frac{a_{13}a_{24} - a_{14}a_{23}}{((a_{13} + a_{14} - a_{23} - a_{24})y + a_{23} + a_{24})^2},$$

$$f_1'|_{(0,0,y,1-y)} = \frac{a_{33}a_{44} - a_{34}a_{43}}{((a_{33} + a_{34} - a_{43} - a_{44})y + a_{43} + a_{44})^2},$$

Note that the row allowability condition guarantees that the denominators in these expressions never vanish.

Choose a_{ij} such that each of these derivatives is less than 1; then we conclude that the entropy rate is analytic at $\hat{\Delta}$. One way to do this is to make each of the 2×2 upper/lower left/right matrices singular.

Or choose the a_{ij} such that

$$\hat{\Delta} = \begin{bmatrix} \alpha_1 & * & \beta_1 & 0 \\ 0 & \alpha_2 & 0 & \beta_2 \\ \lambda_1 & * & \eta_1 & 0 \\ 0 & \lambda_2 & 0 & \eta_2 \end{bmatrix}$$

where $0 < \alpha_1 < \alpha_2$, $0 < \beta_1 < \beta_2$, $0 < \lambda_1 < \lambda_2$, $0 < \eta_1 < \eta_2$ and $*$ denote a real positive numbers. Let (s_2, s_4) be the Perron eigenvalue of the stochastic matrix:

$$\begin{bmatrix} \alpha_2 & \beta_2 \\ \lambda_2 & \eta_2 \end{bmatrix}.$$

Then $\vec{s} = (0, s_2, 0, s_4)$ is the stationary vector of Δ corresponding to the simple eigenvalue 1. Let $w_0 = (0, 1, 0, 0)$ and $w_1 = (0, 0, 0, 1)$. One checks that for $n \geq 0$, $f_{a_n} \circ f_{a_{n-1}} \circ \cdots \circ f_{a_0}(\vec{s}) = w_{a_n}$. Therefore L consists of $\{w_0, w_1\}$. Using the expressions above, we see that

$$f_0'|_{w_0} = \alpha_1/\alpha_2 < 1, f_0'|_{w_1} = \lambda_1/\lambda_2 < 1,$$

$$f_1'|_{w_0} = \beta_1/\beta_2 < 1, f_1'|_{w_1} = \eta_1/\eta_2 < 1.$$

So, f_0 and f_1 are contraction mappings at $\{w_0, w_1\}$, and so condition 3 holds. Thus, the entropy rate $H(Z)$ is analytic at $\hat{\Delta}$.

3 Domain of Analyticity

Suppose Δ is analytically parameterized by a vector variable $\vec{\varepsilon}$, and all the conditions in Theorem 2.1 are satisfied at $\vec{\varepsilon} = \vec{\varepsilon}_0$.

We state a general principle to determine a domain of analyticity for the entropy rate as a function of $\vec{\varepsilon}$. This involves examination of the proof of Theorem 1.1 of [2] and results in the following.

First we enlarge the convex hull (of the intersection of L and W_b) to a convex complex neighborhood N_b so that for every a, the derivative of f_a on N_b is a ρ-contraction, where $0 < \rho < 1$. Then we choose a small enough compact complex neighborhood $\Omega_{\vec{\varepsilon}_0}$ of $\vec{\varepsilon}_0$ such that for all $\varepsilon \in \Omega_{\vec{\varepsilon}_0}$ and for all n, $f_{a_n} \circ f_{a_{n-1}} \circ \cdots \circ f_{a_0}(\vec{s}) \in N_b$, and for the complexified conditional probability corresponding to any z_1^n, $\sum_{z_n} |p(z_n|z_1^{n-1})| < 1/\rho$ on $\Omega_{\vec{\varepsilon}_0}$. From the proof of Theorem 1.1 of [2], the entropy rate is complex analytic on $\Omega_{\vec{\varepsilon}_0}$.

In the following, we study the special case when a binary Markov chain is corrupted by binary symmetric noise, and give a lower bound on the radius of convergence for the entropy rate.

Consider a binary symmetric channel, characterized by

$$Z_n = Y_n \oplus E_n,$$

where $\{Y_n\}$ is the input Markov chain with transition matrix

$$P = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}, \tag{3.1}$$

$\oplus$ denotes binary addition, E_n denotes the i.i.d. binary noise with $p_E(0) = 1 - \varepsilon$ and $p_E(1) = \varepsilon$, and Z_n denotes the corrupted output hidden Markov chain. Strictly speaking, $\{Z_n\}$ is a hidden Markov chain obtained from the underlying Markov chain $\{(Y_n, E_n)\}$. It follows from Remark 3.1 of [2] that the entropy rate is an analytic function of ε at $\varepsilon = 0$ when $\pi_{ij} > 0$; we shall determine a complex neighborhood of 0 such that the entropy rate, as a function of ε, is analytic on this neighborhood.

Let $a_n = p(y_n = 0|z_1^n)$ and $b_n = p(y_n = 1|z_1^n)$. For $z_{n+1} = 1$ we have

$$a_{n+1} = \frac{\varepsilon(\pi_{00}a_n + \pi_{10}b_n)}{\varepsilon(\pi_{00}a_n + \pi_{10}b_n) + (1-\varepsilon)(\pi_{01}a_n + \pi_{11}b_n)},$$

$$b_{n+1} = \frac{(1-\varepsilon)(\pi_{01}a_n + \pi_{11}b_n)}{\varepsilon(\pi_{00}a_n + \pi_{10}b_n) + (1-\varepsilon)(\pi_{01}a_n + \pi_{11}b_n)}.$$

Since $a_n + b_n = 1$, a_{n+1} is a function of a_n; let g_1 denote this function.

For $z_{n+1} = 0$ we have

$$a_{n+1} = \frac{(1-\varepsilon)(\pi_{00}a_n + \pi_{10}b_n)}{(1-\varepsilon)(\pi_{00}a_n + \pi_{10}b_n) + \varepsilon(\pi_{01}a_n + \pi_{11}b_n)},$$

$$b_{n+1} = \frac{\varepsilon(\pi_{01}a_n + \pi_{11}b_n)}{(1-\varepsilon)(\pi_{00}a_n + \pi_{10}b_n) + \varepsilon(\pi_{01}a_n + \pi_{11}b_n)}.$$

Again, a_{n+1} is a function of a_n; let g_0 denote this function.

And for the conditional probability, we have

$$r_0(a_n, b_n) := p(z_n = 0|z_1^{n-1})$$

$$= ((1-\varepsilon)\pi_{00} + \varepsilon\pi_{01})a_n + ((1-\varepsilon)\pi_{10} + \varepsilon\pi_{11})b_n,$$

$$r_1(a_n, b_n) := p(z_n = 1|z_1^{n-1})$$

$$= (\varepsilon\pi_{00} + (1-\varepsilon)\pi_{01})a_n + (\varepsilon\pi_{10} + (1-\varepsilon)\pi_{11})b_n.$$

Note that g_0, g_1, r_0, r_1 are all implicitly parameterized by ε. The stationary vector (π_0, π_1) of Y, which doesn't depend on ε, is equal to $(\pi_{10}/(\pi_{10} + \pi_{01}), \pi_{01}/(\pi_{10} + \pi_{01}))$.

We shall choose ρ with $0 < \rho < 1$, $r > 0$ and $R > 0$ such that for all ε with $|\varepsilon| < r$

1. g_0 and g_1 are ρ-contraction mappings on R-neighborhoods of 0 and 1 in the complex plane,
2. the set of all $\{g_{a_n} \circ g_{a_{n-1}} \circ \cdots \circ g_{a_1}(\pi_0)\}$) are within the R-neighborhoods of 0 and 1,
3. and $|r_0(w)| + |r_1(w)| < 1/\rho$.

By the general principle above, the entropy rate should be analytic on $|\varepsilon| < r$.

More concretely, condition 1, 2 and 3 translate to (here $\rho < 1$, and recall that $b_n = 1 - a_n$):

1. $|g_0'(a_n)| \le \rho$, $|g_1'(a_n)| \le \rho$ on ($|\varepsilon| < r$ and $|a_n| < R$) and ($|\varepsilon| < r$ and $|b_n| < R$).
2. $\max\{|g_0(0)|, |g_0(1)|, |g_1(0)|, |g_1(1)|\} < R(1-\rho)$ on $|\varepsilon| < r$.
3. $|r_0(a_n, b_n)| + |r_1(a_n, b_n)| < 1/\rho$ on ($|\varepsilon| < r$ and $|a_n| < R$) and ($|\varepsilon| < r$ and $|b_n| < R$).

A straightforward computation shows that the following conditions guarantee condition 1, 2, 3:

$$\frac{r(|-\pi_{00}\pi_{11} + \pi_{10}\pi_{11} + \pi_{10}\pi_{01} - \pi_{10}\pi_{11}|r + |(\pi_{00}\pi_{11} + \pi_{10}\pi_{01})|)}{\pi_{11} - |\pi_{10} - \pi_{11}|r - (|\pi_{00} - \pi_{10} - \pi_{01} + \pi_{11}|r + |\pi_{01} - \pi_{11}|R)} \le \rho,$$

$$\frac{r(|-\pi_{00}\pi_{11} + \pi_{10}\pi_{11} + \pi_{10}\pi_{01} - \pi_{10}\pi_{11}|r + |(\pi_{00}\pi_{11} + \pi_{10}\pi_{01})|)}{\pi_{01} - |\pi_{00} - \pi_{01}|r - (|\pi_{00} - \pi_{10} - \pi_{01} + \pi_{11}|r + |\pi_{01} - \pi_{11}|R)} \le \rho,$$

$$\frac{r(|-\pi_{11}\pi_{00} + \pi_{01}\pi_{00} + \pi_{01}\pi_{10} - \pi_{01}\pi_{00}|r + |\pi_{11}\pi_{00} - \pi_{01}\pi_{10}|)}{\pi_{00} - |\pi_{01} - \pi_{00}|r - (|\pi_{00} - \pi_{10} + \pi_{11} - \pi_{01}|r + |\pi_{10} - \pi_{00}|)R} \le \rho,$$

$$\frac{r(|-\pi_{11}\pi_{00} + \pi_{01}\pi_{00} + \pi_{01}\pi_{10} - \pi_{01}\pi_{00}|r + |\pi_{11}\pi_{00} - \pi_{01}\pi_{10}|)}{\pi_{10} - |\pi_{11} - \pi_{10}|r - (|\pi_{00} - \pi_{10} + \pi_{11} - \pi_{01}|r + |\pi_{10} - \pi_{00}|)R} \le \rho,$$

$$\frac{r\pi_{00}}{\pi_{01} - |\pi_{00} - \pi_{01}|r} < R(1-\rho), \frac{r\pi_{10}}{\pi_{11} - |\pi_{10} - \pi_{11}|r} < R(1-\rho),$$

$$\frac{r\pi_{11}}{\pi_{10} - |\pi_{11} - \pi_{10}|r} < R(1-\rho), \frac{r\pi_{01}}{\pi_{00} - |\pi_{01} - \pi_{00}|r} < R(1-\rho),$$

$$(|\pi_{00} - \pi_{01} - \pi_{10} + \pi_{11}|r + |\pi_{01} - \pi_{11}|)R + |\pi_{10} - \pi_{11}|r + \pi_{11},$$

$$+(|\pi_{01} - \pi_{00} + \pi_{10} - \pi_{11}|r + |\pi_{00} - \pi_{10}|)R + |\pi_{11} - \pi_{10}|r + \pi_{10} < 1/\rho,$$

$$(|\pi_{10} - \pi_{11} - \pi_{00} + \pi_{01}|r + |\pi_{11} - \pi_{01}|)R + |\pi_{00} - \pi_{01}|r + \pi_{01}$$
$$+(|\pi_{11} - \pi_{10} + \pi_{00} - \pi_{01}|r + |\pi_{10} - \pi_{00}|)R + |\pi_{01} - \pi_{00}|r + \pi_{00} < 1/\rho.$$

In other words, for given ρ with $0 < \rho < 1$, choose r and R to satisfy all the constraints above. Then the entropy rate is an analytic function of ε on $|\varepsilon| < r$.

4 Stabilizing Property of Derivatives in Black Hole Case

Consider the case that for every $a \in A$, Δ_a is a rank one matrix, and every column of Δ_a is either strictly positive or all zeros. For this case, the image of f_a is a single point and each f_a is defined on the whole simplex W. Thus we call this case the Black Hole case. Analyticity of the entropy rate at a Black Hole follows from Remark 3.1 of [2]. In this section we show that, in principle, the coefficients of a Taylor series expansion, centered at a Black Hole, can be explicitly computed. This result was motivated by and generalizes earlier work by Zuk, et. al. [13, 12] and Ordentlich-Weissman [9] on cases of hidden Markov chains obtained by passing a Markov chain through special kinds of channels.

Note that a binary Markov chain corrupted by binary symmetric noise is a Black Hole case when $\varepsilon = 0$ and $\pi_{ij} > 0$; we calculated a lower bound on the radius of convergence for this case in Section 3.

Suppose that Δ is analytically parameterized by a vector variable $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_m)$. Recall that $H_n(Z)$ is defined as

$$H_n(Z) = H(Z_0|Z_{-n}^{-1}).$$

The following theorem says that at a Black Hole, one can calculate the derivatives of $H(Z)$ by taking the derivatives of $H_n(Z)$ for large enough n.

Theorem 4.1. *If at $\varepsilon = \hat{\varepsilon}$, for every $a \in A$, Δ_a is a rank one matrix, and every column of Δ_a is either a positive or a zero column, then*

$$\left.\frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_m} H(Z)}{\partial_{\varepsilon_1}^{\alpha_1}\partial_{\varepsilon_2}^{\alpha_2}\cdots\partial_{\varepsilon_m}^{\alpha_m}}\right|_{\varepsilon=\hat{\varepsilon}} = \left.\frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_m} H_{\alpha_1+\alpha_2+\cdots+\alpha_m}(Z)}{\partial_{\varepsilon_1}^{\alpha_1}\partial_{\varepsilon_2}^{\alpha_2}\cdots\partial_{\varepsilon_m}^{\alpha_m}}\right|_{\varepsilon=\hat{\varepsilon}}$$

In fact, we give a stronger result, Theorem 4.6, later in this section.

Proof. For simplicity we assume that Δ is only parameterized by one real variable ε.

We shall first prove that for all sequences $z_{-\infty}^0$ the n-th derivative of $p(z_0|z_{-\infty}^{-1})$ stabilizes:

$$p^{(n)}(z_0|z_{-\infty}^{-1}) = p^{(n)}(z_0|z_{-n-1}^{-1}) \text{ at } \varepsilon = \hat{\varepsilon}. \tag{4.2}$$

Since $p(z_0|z_{-\infty}^{-1}) = p(y_{-1}|z_{-\infty}^{-1})\Delta_{z_0}\mathbf{1}$, it suffices to prove that for the n-th derivative of $x_i = p(y_i|z_{-\infty}^i)$, we have

$$x_i^{(n)} = p^{(n)}(y_i|z_{-\infty}^i) = p^{(n)}(y_i|z_{i-n}^i) \text{ at } \varepsilon = \hat{\varepsilon}. \tag{4.3}$$

Consider the iteration:

$$x_i = \frac{x_{i-1}\Delta_{z_i}}{x_{i-1}\Delta_{z_i}\mathbf{1}}.$$

In other words, x_i can be viewed as a function of x_{i-1} and Δ_a. Let g denote this function. Since at $\varepsilon = \hat{\varepsilon}$, Δ_{z_i} is a rank one matrix, we conclude that g is a constant as a function of x_{i-1} in this case. Thus at $\varepsilon = \hat{\varepsilon}$

$$x_i = p(y_i|z_{-\infty}^i) = \frac{x_{i-1}\Delta_{z_i}}{x_{i-1}\Delta_{z_i}\mathbf{1}} = \frac{p(y_{i-1})\Delta_{z_i}}{p(y_{i-1})\Delta_{z_i}\mathbf{1}} = p(y_i|z_i). \tag{4.4}$$

For the first order derivative, taking the derivative of g with respect to ε, we have at $\varepsilon = \hat{\varepsilon}$

$$x_i' = \left.\frac{\partial g}{\partial \Delta_{z_i}}\right|_{\varepsilon=\hat{\varepsilon}} (x_{i-1}, \Delta_{z_i})\ \Delta_{z_i}' + \left.\frac{\partial g}{\partial x_{i-1}}\right|_{\varepsilon=\hat{\varepsilon}} (x_{i-1}, \Delta_{z_i})\ x_{i-1}'.$$

Recall that at $\varepsilon = \hat{\varepsilon}$, g is a constant as a function of x_{i-1}, so we have

$$\left.\frac{\partial g}{\partial x_{i-1}}\right|_{\varepsilon=\hat{\varepsilon}} (x_{i-1}, \Delta_{z_i}) = \frac{\partial(\text{a constant vector})}{\partial x_{i-1}} = 0.$$

It then follows from (4.4) that at $\varepsilon = \hat{\varepsilon}$

$$x_i' = p'(y_i|z_{-\infty}^i) = p'(y_i|z_{i-1}^i).$$

Taking higher order derivatives, we have

$$x_i^{(n)} = \left.\frac{\partial g}{\partial x_{i-1}}\right|_{\varepsilon=\hat{\varepsilon}} (x_{i-1}, \Delta_{z_i})\ x_{i-1}^{(n)} + \text{other terms},$$

where "other terms" involve only lower order (than n) derivatives of x_{i-1}. By induction, we conclude that

$$x_i^{(n)} = p^{(n)}(y_i|z_{-\infty}^i) = p^{(n)}(y_i|z_{i-n}^i).$$

at $\varepsilon = \hat{\varepsilon}$. We then have (4.3) as desired.

By the proof of Theorem 1.1 of [2] (see Remark 3.1 of [2]), the complexified $H_n(Z)$ uniformly converges to the complexified $H(Z)$, and so we can switch the limit operation and the derivative operation. Thus, at all ε,

$$H'(Z) = (\lim_{k\to\infty} \sum_{z_{-k}^0} (p(z_{-k}^0)\log p(z_0|z_{-k}^{-1}))'$$

$$= \lim_{k\to\infty} \sum_{z_{-k}^0} (p'(z_{-k}^0)\log p(z_0|z_{-k}^{-1}) + p(z_{-k}^0)\frac{p'(z_0|z_{-k}^{-1})}{p(z_0|z_{-k}^{-1})})$$

Since

$$\sum_{z_0} p(z_{-k}^0)\frac{p'(z_0|z_{-k}^{-1})}{p(z_0|z_{-k}^{-1})} = \sum_{z_0} p(z_{-k}^{-1})p'(z_0|z_{-k}^{-1}) = 0,$$

we have for all ε

$$H'(Z) = \lim_{k\to\infty} \sum_{z_{-k}^{0}} (p'(z_{-k}^{0}) \log p(z_0|z_{-k}^{-1})). \tag{4.5}$$

At $\varepsilon = \hat{\varepsilon}$, we obtain:

$$H'(Z) = \lim_{k\to\infty} \sum_{z_{-k}^{0}} (p'(z_{-k}^{0}) \log p(z_0|z_{-1}))$$

$$= \sum_{z_{-1}^{0}} (p'(z_{-1}^{0}) \log p(z_0|z_{-1})) = H_1'(Z).$$

For higher order derivatives, again using the fact that we can interchange the order of limit and derivative operations and using (4.5), we have for all ε

$$H^{(n)}(Z) = \lim_{k\to\infty} \sum_{z_{-k}^{0}} \sum_{l=1}^{n} C_{n-1}^{l-1} p^{(l)}(z_{-k}^{0}) (\log p(z_0|z_{-k}^{-1}))^{(n-l)}$$

(the use of (4.5) accounts for the fact that there is no $l = 0$ term in this expression). Note that the term $(\log p(z_0|z_{-k}^{-1}))^{(n-l)}$ involves only the lower order (less than or equal to $n-1$) derivatives of $p(z_0|z_{-k}^{-1})$, which are already "stabilizing" in the sense of (4.2); so, we have

$$H^{(n)}(Z) = \lim_{k\to\infty} \sum_{z_{-k}^{0}} \sum_{l=1}^{n} C_{n-1}^{l-1} p^{(l)}(z_{-k}^{0}) (\log p(z_0|z_{-n}^{-1}))^{(n-l)}$$

$$= \sum_{z_{-n}^{0}} \sum_{l=1}^{n} C_{n-1}^{l-1} p^{(l)}(z_{-n}^{0}) (\log p(z_0|z_{-n}^{-1}))^{(n-l)} = H_n^{(n)}(Z).$$

We thus prove the theorem. □

Remark 4.2. It follows from (4.4) that a hidden Markov chain at a Black Hole is, in fact, a Markov chain. Note that in the argument above the proof of the stabilizing property of the first derivative (as opposed to higher derivatives) requires only that the hidden Markov chain is Markov and that we can interchange the order of limit and derivative operations (instead of the stronger Black Hole property). Therefore if a hidden Markov chain Z defined by $\hat{\Delta}$ and Φ is in fact a Markov chain, and the complexified $H_n(Z)$ uniformly converges to $H(Z)$ on some neighborhood of $\hat{\Delta}$ (e.g., if the conditions of Theorem 2.1 or the conditions of Remark 3.1 of [2] hold), then at $\hat{\Delta}$, we have

$$H'(Z) = H_1'(Z).$$

In the cases studied in [13, 12, 9], the authors obtained, using a finer analysis, a shorter "stabilizing length." This shorter length can be derived for the Black Hole case as well, as shown in Theorem 4.6 below.

By induction, one can prove that the formal derivative of $y\log y$ takes the following form:

$$(y\log y)^{(N)} = \sum_{a_1\geq a_2\geq\cdots\geq a_{m+1}: a_1+a_2+\cdots+a_{m+1}=N} E_{[a_1,a_2,\cdots,a_{m+1}]}\frac{y^{(a_1)}y^{(a_2)}\cdots y^{(a_{m+1})}}{y^m} + y^{(N)}(\log y+1)$$

$$= \sum_{i=1}^{N-1} y^{(a_1=i)} \sum_{a_2\geq a_3\geq\cdots\geq a_{m+1}} E_{[a_1,a_2,\cdots,a_{m+1}]}\frac{y^{(a_2)}y^{(a_3)}\cdots y^{(a_{m+1})}}{y^m} + y^{(N)}(\log y+1).$$

Let $q_i[y]$ denote the "coefficient" of $y^{(i)}$, which is a function of y and its formal derivatives (up to the i-th order derivative). Thus we have

$$(y\log y)^{(N)} = \sum_{i=1}^{N} q_i[y]y^{(i)} = \text{High}_N[y] + \text{Low}_N[y],$$

where $\text{High}_N[y] = \sum_{\lceil (N+1)/2\rceil}^{N} q_i[y]y^{(i)}$ and $\text{Low}_N[y] = \sum_{i=1}^{\lceil (N-1)/2\rceil} q_i[y]y^{(i)}$.

In following, let $P(a_1,a_2,\cdots,a_m)$ denote the number of distinct sequences obtained by permuting the coordinates of the sequence $(a_1,a_2,\cdots,a_m)$. Namely if

$$a_1 = a_2 = \cdots = a_{m_1} > a_{m_1+1} = \cdots = a_{m_1+m_2} > \cdots > a_{m_1+m_2+\cdots+m_{j-1}+1} = \cdots = a_{m_1+m_2+\cdots+m_j} = a_m, \tag{4.6}$$

then

$$P(a_1,a_2,\cdots,a_m) = \frac{m!}{m_1!m_2!\cdots m_j!}.$$

Lemma 4.3.

$$(y'/y)^{(n)} = \sum_{a_1\geq a_2\geq\cdots\geq a_m\geq 1: a_1+a_2+\cdots+a_m=n+1} C_{[a_1,a_2,\cdots,a_m]}(y^{(a_1)}y^{(a_2)}\cdots y^{(a_m)})/y^m,$$

where $C_{[a_1,a_2,\cdots,a_m]} = (-1)^{m+1}\frac{1}{m}P(a_1,a_2,\cdots,a_m)\frac{(a_1+a_2+\cdots+a_m)!}{a_1!a_2!\cdots a_m!}$.

Proof. One checks that $C_{[1]} = 1$ and $C_{[a_1,a_2,\cdots,a_m]}$ satisfies the following recursion relationship:

For $a_1\geq a_2\geq\cdots\geq a_m\geq 2$,

$$C_{[a_1,a_2,\cdots,a_m]} = \sum D(a_1,a_2,\cdots,a_m;b_1,b_2,\cdots,b_m)C_{[b_1,b_2,\cdots,b_m]}, \tag{4.7}$$

where the summation is over all $b_1\geq b_2\geq\cdots\geq b_m\geq 1$, and all b_i is equal to a_i except for one of them, say $b_k = a_k - 1$, and $D(a_1,a_2,\cdots,a_m;b_1,b_2,\cdots,b_m)$ is defined to the number of b_k occurring in the sequence of $b_1,b_2,\cdots,b_m$. For $a_1\geq a_2\geq\cdots\geq a_m = 1$,

$$C_{[a_1,a_2,\cdots,a_m]} = \sum D(a_1,a_2,\cdots,a_m;b_1,b_2,\cdots,b_m)C_{[b_1,b_2,\cdots,b_m]} - (m-1)C_{[a_1,a_2,\cdots,a_{m-1}]}; \tag{4.8}$$

again here the summation is over all $b_1\geq b_2\geq\cdots\geq b_m\geq 1$, and all b_i is equal to a_i except for one of them, say $b_k = a_k - 1$, and $D(a_1,a_2,\cdots,a_m;b_1,b_2,\cdots,b_m)$ is defined to the number of b_k occurring in the sequence of $b_1,b_2,\cdots,b_m$.

One checks that

$$(-1)^{m+1}\frac{1}{m}P(a_1,a_2,\cdots,a_m)\frac{(a_1+a_2+\cdots+a_m)!}{a_1!a_2!\cdots a_m!},$$

satisfies the initial value and recursion (4.7) and (4.8). Since the initial value and recursion uniquely determine the sequence, the theorem then follows. $\square$

Lemma 4.4. *For $i = \lceil (N+1)/2 \rceil, \cdots, N$, q_i is proportional to $(\log y + 1)^{(N-i)}$. More specifically, we have*

$$q_i[y] = C_{i,N}(\log y + 1)^{(N-i)},$$

where $C_{i,N}$ is an integer.

Proof. We first prove that for $N = 2k+1$, the coefficient of $y^{(k+1)}$ is proportional to $z^{(k-1)}$, where $z = (\log y + 1)' = y'/y$. According to Leibnitz formula, we have

$$(y \log y)^{(2k+1)} = (y'(\log y + 1))^{(2k)} = \sum_{l=0}^{2k} C_{2k}^l y^{(l+1)} (\log y + 1)^{(2k-l)}$$

$$= y^{(2k+1)}(\log y + 1) + \sum_{l=0}^{2k-1} C_{2k}^l y^{(l+1)} z^{(2k-l-1)}.$$

It suffices to prove that the coefficient of $y^{(k+1)}$ of

$$C_{2k}^{k+1} y^{(k)} z^{(k)} + C_{2k}^{k+2} y^{(k-1)} z^{(k+1)} + \cdots + C_{2k}^{2k} y' z^{(2k-1)}$$

is $C_{2k}^{k+1} z^{(k-1)}$. Applying Lemma 4.3 and collecting terms, we have the coefficient of $y^{(k+1)}$ equal to

$$C_{2k}^{k+1} C_{[k+1]} y^{(k)}/y + C_{2k}^{k+2} C_{[k+1,1]} (y^{(k-1)} y^{(1)})/y^2$$
$$+ C_{2k}^{k+3} C_{[k+1,2]} (y^{(k-2)} y^{(2)})/y^2 + C_{2k}^{k+3} C_{[k+1,1,1]} (y^{(k-2)} y^{(1)} y^{(1)})/y^3$$
$$+ C_{2k}^{k+4} C_{[k+1,3]} (y^{(k-3)} y^{(3)})/y^2 + C_{2k}^{k+4} C_{[k+1,2,1]} (y^{(k-3)} y^{(2)} y^{(1)})/y^3 + C_{2k}^{k+4} C_{[k+1,1,1,1]} (y^{(k-3)} y^{(1)} y^{(1)} y^{(1)})/y^4$$
$$+ \cdots + C_{2k}^{2k} C_{[k+1,k-1]} (y^{(1)} y^{(k-1)}/y^2 + \cdots + C_{2k}^{2k} C_{[k+1,1,\cdots,1]} (y^{(1)} y^{(1)} \cdots y^{(1)})/y^k.$$

Consider the term $(y^{(a_1)} y^{(a_2)} \cdots y^{(a_m)})/y^m$ (here $a_1 + a_2 + \cdots + a_m = k$) and compute its coefficient in the expression above. Assuming that $a_1 \geq a_2 \geq \cdots \geq a_m$ satisfying (4.6), we have the coefficient

$$C_{2k}^{2k+1-a_1} C_{[k+1,a_2,\cdots,a_m]} + C_{2k}^{2k+1-a_{m_1+1}} C_{[k+1,a_1,\cdots,a_{m_1},a_{m_1+2},\cdots,a_m]} + \cdots$$
$$+ C_{2k}^{2k+1-a_{m_1+m_2+\cdots+m_{j-1}+1}} C_{[k+1,a_1,\cdots,a_{m_1+m_2+\cdots+m_{j-1}},a_{m_1+m_2+\cdots+m_{j-1}+2},\cdots,a_m]}$$
$$= (-1)^{m+1} \frac{1}{m} \left(\frac{(2k)!}{(2k+1-a_1)!(a_1-1)!} \frac{(2k+1-a_1)!}{(k+1)! a_2! \cdots a_m!} \frac{m!}{(m_1-1)! m_2! \cdots m_j!} \right.$$
$$+ \frac{(2k)!}{(2k+1-a_{m_1+1})!(a_{m_1+1}-1)!} \frac{(2k+1-a_{m_1+1})!}{a_1! \cdots a_{m_1}!(k+1)! a_{m_1+2}! \cdots a_m!} \frac{m!}{m_1!(m_2-1)! \cdots m_j!} + \cdots +$$
$$\left. + \frac{(2k)!}{(2k+1-a_{m_1+\cdots+m_{j-1}+1})!(a_{m_1+\cdots+m_{j-1}+1}-1)!} \frac{(2k+1-a_{m_1+\cdots+m_{j-1}+1})!}{a_1! \cdots a_{m_1+\cdots+m_{j-1}}!(k+1)! a_{m_1+\cdots+m_{j_1}+2}! \cdots a_m!} \frac{m!}{m_1! m_2! \cdots (m_j-1)!} \right)$$
$$= (-1)^{m+1} \frac{1}{m} \frac{(2k)!}{(k+1)!} \frac{m!}{m_1! \cdots m_j!} \frac{m_1 a_1 + m_2 a_{m_1+1} + \cdots + m_j a_{m_1+\cdots+m_{j-1}+1}}{a_1! a_2! \cdots a_m!}$$
$$= (-1)^{m+1} \frac{1}{m} \frac{(2k)!}{(k+1)!(k-1)!} \frac{(a_1 + a_2 + \cdots + a_m)!}{a_1! a_2! \cdots a_m!} \frac{m!}{m_1! m_2! \cdots m_j!}$$

$$= C_{2k}^{k+1} C_{[a_1, a_2, \cdots, a_m]}.$$

It then follows that the coefficient of $y^{(k+1)}$ is equal to $C_{2k}^{k+1} z^{(k-1)}$.

One can do similar computations to prove that for $N = 2k, 2k+1$, this lemma holds for other derivatives. An alternative approach is to use induction. Using the fact that the coefficient of $y^{(k+1)}$ is proportional to $z^{(k-1)}$ (established above), one can prove by induction that for the $2k$-th order derivative of $y \log y$, the coefficient of $y^{(l)}$ is proportional to $(\log y + 1)^{(2k-l)}$ for l with $k+1 \leq l \leq 2k$; and for $2k+1$-th order derivative of $y \log y$, the coefficient of $y^{(l)}$ is proportional to $(\log y + 1)^{(2k+1-l)}$ for l with $k+2 \leq l \leq 2k+1$. □

Lemma 4.5.

$$\mathrm{Low}_N[ax] = \sum_{i=0}^{\lceil (N-1)/2 \rceil} r_i[a] x^{(i)} + \sum_{i=0}^{\lceil (N-1)/2 \rceil} s_i[x] a^{(i)},$$

where $r_i[a]$ is a function of a and its derivatives (up to order $\lceil (N-1)/2 \rceil$), and $s_i[x]$ is a function of x and its derivatives (up to order $\lceil (N-1)/2 \rceil$). Also,

$$s_0 = \mathrm{Low}_N[x].$$

Proof. By Leibnitz formula, we have

$$((ax)\log(ax))^{(N)} = \sum_{i=0}^{N} C_N^i (ax)^{(i)} (\log(ax))^{(N-i)}$$

$$= \sum_{i=0}^{N} C_N^i \sum_{j=0}^{i} C_i^j a^{(j)} x^{(i-j)} (\log a + \log x)^{(N-i)}.$$

Thus there exist a function of a and its derivatives $t_i[a]$, and a function of x and its derivatives $w_i[x]$ such that

$$((ax)\log(ax))^{(N)} = \sum_{i=0}^{N} t_i[a] x^{(i)} + \sum_{i=0}^{N} w_i[x] a^{(i)},$$

with $w_0[x] = (x \log x)^{(N)}$.

By Lemma 4.4, we have

$$\mathrm{High}_N[ax] = \sum_{i=\lceil (N+1)/2 \rceil}^{N} q_i[ax](ax)^{(i)} = \sum_{i=\lceil (N+1)/2 \rceil}^{N} C_{i,N} (\log a + \log x + 1)^{(N-i)} (ax)^{(i)}$$

Thus we conclude that there exist a function of a and its derivatives $u_i[a]$, and a function of x and its derivatives $v_i[x]$ such that

$$\mathrm{High}_N[ax] = \sum_{i=\lceil (N+1)/2 \rceil}^{N} u_i[a] x^{(i)} + \sum_{i=\lceil (N+1)/2 \rceil}^{N} v_i[x] a^{(i)},$$

with $v_0[x] = \mathrm{High}_N[x]$. Since

$$\mathrm{Low}_N[ax] = ((ax)\log(ax))^{(N)} - \mathrm{High}_N[ax],$$

existence of $r_i[a]$ and $s_i[x]$ then follows, and they depend on the derivatives only up to $\lceil (N-1)/2 \rceil$, and $s_0[x] = \mathrm{Low}_N[x]$. □

Theorem 4.6. *If at $\varepsilon = \hat{\varepsilon}$, for every $a \in A$, Δ_a is a rank one matrix, and every column of Δ_a is either a positive or a zero column, then*

$$\left.\frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_m} H(Z)}{\partial_{\varepsilon_1}^{\alpha_1}\partial_{\varepsilon_2}^{\alpha_2}\cdots\partial_{\varepsilon_m}^{\alpha_m}}\right|_{\varepsilon=\hat{\varepsilon}} = \left.\frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_m} H_{\lceil(\alpha_1+\alpha_2+\cdots+\alpha_m+1)/2\rceil}(Z)}{\partial_{\varepsilon_1}^{\alpha_1}\partial_{\varepsilon_2}^{\alpha_2}\cdots\partial_{\varepsilon_m}^{\alpha_m}}\right|_{\varepsilon=\hat{\varepsilon}}$$

Proof. For simplicity we assume that Δ is only parameterized by only one variable ε, and we drop ε when the implication is clear in the context. Recall that

$$H_n(Z) = -\sum_{z_{-n}^0} p(z_{-n}^0)\log p(z_0|z_{-n}^{-1}) = -(\sum_{z_{-n}^0} p(z_{-n}^0)\log p(z_{-n}^0) - \sum_{z_{-n}^{-1}} p(z_n^{-1})\log p(z_n^{-1})).$$

With slight abuse of notation (by replacing the formal derivative with the derivative with respect to ε, we can define $\text{High}_N[p(z_{-n}^0)] = \text{High}_N[p^\varepsilon(z_{-n}^0)]$. Similarly for $\text{Low}_N[p(z_{-n}^0)]$, etc.),

$$(p(z_{-n}^0)\log p(z_{-n}^0))^{(N)} = \text{High}_N[p(z_{-n}^0)] + \text{Low}_N[p(z_{-n}^0)]$$
$$(p(z_{-n}^{-1})\log p(z_{-n}^{-1}))^{(N)} = \text{High}_N[p(z_{-n}^{-1})] + \text{Low}_N[p(z_{-n}^{-1})]$$

Note that by Lemma 4.4, we have

$$\text{High}_N[p(z_{-n}^0)] = \sum_{i=\lceil(N+1)/2\rceil}^{N} C_{i,N}(\log p(z_0|z_{-n}^{-1}) + \log p(z_{-n}^{-1}) + 1)^{(N-i)} p(z_{-n}^0)^{(i)},$$

and

$$\text{High}_N[p(z_{-n}^{-1})] = \sum_{i=\lceil(N+1)/2\rceil}^{N} C_{i,N}(\log p(z_{-n}^{-1}) + 1)^{(N-i)} p(z_{-n}^{-1})^{(i)}.$$

Thus

$$\sum_{z_{-n}^0}\text{High}_N[p(z_{-n}^0)] - \sum_{z_{-n}^{-1}}\text{High}_N[p(z_{-n}^{-1})]$$

$$= \sum_{z_{-n}^0}\sum_{i=\lceil(N+1)/2\rceil}^{N} C_{i,N}(\log p(z_0|z_{-n}^{-1}) + \log p(z_{-n}^{-1}) - \log p(z_{-n}^{-1}))^{(N-i)} p(z_{-n}^0)^{(i)}$$

$$= \sum_{z_{-n}^0}\sum_{i=\lceil(N+1)/2\rceil}^{N} C_{i,N}(\log p(z_0|z_{-n}^{-1}))^{(N-i)} p(z_{-n}^0)^{(i)}$$

$$= \sum_{z_{-n}^0}\sum_{i=\lceil(N+1)/2\rceil}^{N} C_{i,N}(\log p(z_0|z_{-\lceil(N+1)/2\rceil}^{-1}))^{(N-i)} p(z_{-\lceil(N+1)/2\rceil}^0)^{(i)}$$

So the higher derivative part stabilizes at $\lceil(N+1)/2\rceil$, namely for any $n \geq \lceil(N+1)/2\rceil$, $\sum_{z_{-n}^0}\text{High}_N[p(z_{-n}^0)] - \sum_{z_{-n}^{-1}}\text{High}_N[p(z_{-n}^{-1})]$ is equal to $\sum_{z_{-\lceil(N+1)/2\rceil}^0}\text{High}_N[p(z_{-\lceil(N+1)/2\rceil}^0)] - \sum_{z_{-\lceil(N+1)/2\rceil}^{-1}}\text{High}_N[p(z_{-\lceil(N+1)/2\rceil}^{-1})]$. And by Lemma 4.5, we have

$$\text{Low}_N[p(z_{-n}^0)] = \sum_{i=0}^{\lceil(N-1)/2\rceil} r_i[p(z_0|z_{-n}^{-1})]p(z_{-n}^{-1})^{(i)} + \sum_{i=0}^{\lceil(N-1)/2\rceil} s_i[p(z_{-n}^{-1}))]p(z_0|z_{-n}^{-1})^{(i)},$$

with $s_0[p(z_{-n}^{-1}))] = \text{Low}_N[p(z_{-n}^{-1})]$. Thus,

$$\sum_{z_{-n}^0} \text{Low}_N[p(z_{-n}^0)] - \sum_{z_{-n}^{-1}} \text{Low}_N[p(z_{-n}^{-1})]$$

$$= \sum_{z_{-n}^0} \sum_{i=0}^{\lceil (N-1)/2 \rceil} r_i[p(z_0|z_{-n}^{-1})]p(z_{-n}^{-1})^{(i)}.$$

$$= \sum_{z_{-n}^0} \sum_{i=0}^{\lceil (N-1)/2 \rceil} r_i[p(z_0|z_{-\lceil (N+1)/2 \rceil}^{-1})]p(z_{-\lceil (N+1)/2 \rceil}^{-1})^{(i)}.$$

Consequently the lower derivative part stabilizes at $\lceil (N+1)/2 \rceil$ as well, namely for any $n \geq \lceil (N+1)/2 \rceil$, $\sum_{z_{-n}^0} \text{Low}_N[p(z_{-n}^0)] - \sum_{z_{-n}^{-1}} \text{Low}_N[p(z_{-n}^{-1})]$ is equal to $\sum_{z_{-\lceil (N+1)/2 \rceil}^0} \text{Low}_N[p(z_{-\lceil (N+1)/2 \rceil}^0)] - \sum_{z_{-\lceil (N+1)/2 \rceil}^{-1}} \text{Low}_N[p(z_{-\lceil (N+1)/2 \rceil}^{-1})]$. The theorem then follows.

□

Remark 4.7. For an irreducible stationary Markov chain Y with probability transition matrix Δ, let Y^{-1} denote its reverse Markov chain. It is well known that probability transition matrix of Y^{-1} is $\text{diag}\,(\pi_1^{-1}, \pi_2^{-1}, \cdots, \pi_B^{-1})\Delta^t \text{diag}\,(\pi_1, \pi_2, \cdots, \pi_B)$, where Δ^t denotes the transpose of Δ and $(\pi_1, \pi_2, \cdots, \pi_B)$ is the stationary vector of Y. Therefore if Δ^t is a Black Hole case, the derivatives of $H(Z^{-1})$ (here, Z^{-1} is the reverse hidden Markov chain defined by $Z^{-1} = \Phi(Y^{-1})$) also stabilize. It then follows from $H(Z) = H(Z^{-1})$ that the derivatives of $H(Z)$ also stabilize.

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. Han and B. Marcus, Analyticity of Entropy Rate in Families of Hidden Markov Chains, July 2005 (revised Nov. 2005) http://front.math.ucdavis.edu/math.PR/0507235. Submitted to IEEE Transactions on Information Theory, a preliminary version can be found in Proc. of IEEE International Symposium on Information Theory, Adelaide, Australia September 4-September 9 2005, pages 2193-2197.

[3] T. Holliday, A. Goldsmith, and P. Glynn. On entropy and lyapunov exponents for finite state channels. 2003. Available at http://wsl.stanford.edu/Publications/THolliday/Lyapunov.pdf.

[4] P. Jacquet, G. seroussi, and W. Szpankowski. On the entropy of a hidden markov process. In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, page 10, Chicago, U.S.A., 2004.

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

[6] B. Marcus and K. Petersen and S. Williams. Transmission rates and factors of markov chains. *Contemporary Mathematics*, 26:279–294, 1984.

[7] E. Ordentlich and T. Weissman. On the optimality of symbol by symbol filtering and denoising. Preprint, 2003.

[8] E. Ordentlich and T. Weissman. New bounds on the entropy rate of hidden markov process. Preprint, 2004.

[9] E. Ordentlich and T. Weissman, personal communication.

[10] Y. Peres. *Analytic dependence of Lyapunov exponents on transition probabilities*, volume 1486 of *Lecture Notes in Mathematics, Lyapunov's exponents, Proceedings of a Workshop*. Springer Verlag, 1990.

[11] Y. Peres. Domains of analytic continuation for the top Lyapunov exponent. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(1):131–148, 1992.

[12] O. Zuk, E. Domany, I. Kanter, and M. Aizenman. Taylor series expansions for the entropy rate of Hidden Markov Processes. Preprint, 2005.

[13] O. Zuk, I. Kanter, and E. Domany. Asymptotics of the entropy rate for a hidden markov process. Preprint, 2004.