Asymptotics of the Entropy Rate for a Hidden Markov Process

Or Zuk[1], Ido Kanter[2], and Eytan Domany[1]

[1] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, 76100, Israel, {*or.zuk/eytan.domany*}*@weizmann.ac.il*
[2] Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel, *kanter@mail.biu.ac.il*

Abstract. We calculate the Shannon entropy rate of a binary Hidden Markov Process (HMP), of given transition rate and noise ϵ (emission), as a series expansion in ϵ. The first two orders are calculated exactly. We then evaluate, for finite histories, simple upper-bounds of Cover and Thomas. Surprisingly, we find that for a fixed order k and history of n steps, the bounds become independent of n for large enough n. This observation is the basis of a conjecture, that the upper-bound obtained for $n \geq (k+3)/2$ gives the exact entropy rate for any desired order k of ϵ.

1 Introduction and Statement of Results

Let $X = \{X_n\}_{n \geq 1}$ be a first order stationary Markov process over a binary alphabet, with a symmetric transition matrix $P \equiv P_{ab}$ given by $P_{00} = P_{11} = p = 1 - P_{01} = 1 - P_{10}$, where $P_{ab} = \Pr(X_n = b | X_{n-1} = a)$, $\forall a, b \in \{0, 1\}$. Consider also a Bernoulli (binary i.i.d.) noise process $E = \{E_n\}_{n \geq 1}$, independent of X, with $\Pr(E_n = 1) = \epsilon = 1 - \Pr(E_n = 0)$. Finally, define the process $Y = \{Y_n\}_{n \geq 1}$ by :

$$Y_n = X_n \oplus E_n, \ \forall n \in \mathbb{N} \tag{1}$$

Where $\oplus$ denotes addition modulo 2 (exclusive-or). We denote a vector of variables $Z_i, \ldots Z_j$ by Z_i^j. Also, $Z_i^j(\overline{k})$ denotes the vector $Z_i, \ldots, \overline{Z_k}, \ldots, Z_j$ where $\overline{Z}$ denotes complement of Z. Uppercase denote r.v.s, and lower case denote their realizations. When possible, we omit the latter. (For example, $\Pr(Z_i^j)$ means $\Pr(Z_i^j = z_i^j)$).
The process Y can be viewed as a noisy observation of X, through a binary symmetric channel. It is one of the simplest examples of a Hidden Markov Process (HMP), and is determined completely by the choice of parameters p and ϵ. More generally, HMP's have a wide variety of applications, in various fields such as speech recognition, machine learning, signal processing, bioinformatics etc. For comprehensive reviews of the literature on HMP's see [3] and [13].
Despite the simplicity of their definition, some very basic questions on the properties of HMP's are still unsolved. Typical examples are the filtering and denoising errors, which are studied, for example, in [6] and [11]. In this paper we concentrate on the Shannon entropy rate of the process, which is also not known to date ([3],[5]). The entropy rate is defined by :

$$H(Y) \equiv H(p, \epsilon) = \lim_{n \to \infty} \frac{1}{n} E[-\log \Pr(Y_1^n)] \tag{2}$$

For simplicity, we use the natural logarithm here, thus the entropy is measured in NATS. Though $H(p,\epsilon)$ has no known closed form, three recent papers ([5],[11],[12]) give the asymptotic behavior of H in several regimes, (for a general binary transition matrix P). This paper extends the work of [5], dealing with the small noise regime (termed 'high SNR') $\epsilon \to 0$. We wish to find the expansion of H in ϵ around zero, when p is treated as a constant parameter (we assume $p \neq 0, 1$ [1]) . Thus, denoting :

$$H_k \equiv H_k(p) = \frac{1}{k!}\frac{\partial^k H(p,\epsilon)}{\partial \epsilon^k}|_{(p,0)},\ \forall k \geq 0 \tag{3}$$

H is given by :

$$H(p,\epsilon) = \sum_{k=0}^{\infty} H_k(p)\epsilon^k \tag{4}$$

First, in Section 2, we give a method for exact computation of any order of the entropy, and demonstrate it for computing H_1, H_2. Our method is based on low-temperature/high-field expansion from statistical mechanics. Next, in Section 3, we use the known bounds [1] on the entropy rate:

$$c^{(n)} \equiv H(Y_n|X_1, Y_1^{n-1}) \leq H(Y) \leq H(Y_n|Y_1^{n-1}) \equiv C^{(n)},\ \forall n \geq 1 \tag{5}$$

which are known to converge to the entropy rate ([1]), i.e. :

$$\lim_{n\to\infty} c^{(n)} = \lim_{n\to\infty} C^{(n)} = H(Y)$$

Using the upper-bounds $C^{(n)}$, we can get an alternative method for computing H_k; rather than computing $H(Y_1^n)$, we evaluate directly the conditional entropies $C^{(n)} = H(Y_n|Y_1^{n-1})$ up to some given order. We demonstrated this for the first order term H_1. We continue in Section 3 to study the upper-bounds $C^{(n)}$, by computing them explicitly ([10]) for $n \leq 8$, and expanding $C^{(n)}$ as a power series in ϵ,

$$C^{(n)} = \sum_{k=0}^{\infty} C_k^{(n)}\epsilon^k \tag{6}$$

This led to the discovery of rather surprising and interesting behavior of the coefficients $C_k^{(n)}$: *they become independent of* n *for* $n \geq \frac{k+3}{2}$. Since $C^{(n)} \to H$ as $n \to \infty$, it follows that $C_k^{(n)} \to H_k, \forall k \in \mathbb{N}$. This behavior was tested to be true for $k = 0, 1, ..., 11$ and $\frac{k+3}{2} \leq n \leq 8$. Therefore we pose the following :

Conjecture 1

$$k \leq 2n - 3 \Rightarrow C_k^{(n)} = H_k,\ \forall k, n \in \mathbb{N} \tag{7}$$

Note that we have computed $C_k^{(n)}$ also for a non-symmetric transition matrix, for the first few orders, up to $k = 7$ and $n = 5$ and $C_k^{(n)}$ also becomes independent of n for $n \geq \frac{k+3}{2}$. In particular, the first order $C_1^{(n)}$ becomes equal to the exact function of the transition probabilities $H_1(p_1, p_2)$, which is computed in [5].This function diverges as

[1] For $p \neq 0, 1$ the entropy is an analytic function of ϵ at $\epsilon = 0$; See Sec. 2

one of the transition probabilities approaches 1, in agreement with [12]. Furthermore, we found that the $C_k^{(n)}$ share some common properties as functions of p. Assuming Conjecture 1 is valid, the H_k's share the same properties (which we checked for $k \leq 11$); we express these as the following :

Conjecture 2 Let $\lambda = 1-2p$ be the 2^{nd} eigenvalue of the transition matrix P. Then, for $k \geq 3$ we have :

$$H_k = \frac{2^{4(k-1)} \sum_{j=0}^{d_k} a_{j,k} \lambda^{2j}}{k(k-1)(1-\lambda^2)^{2(k-1)}} \tag{8}$$

where $d_k \in \mathbb{N}$ are constants and the $a_{j,k} \in \mathbb{Z}$ satisfy the relation $\sum_{j=0}^{d_k} a_{j,k} = (-1)^{k-1}$. In Section 4 we discuss our results, and offer several future directions.

2 Exact Derivation of the First Orders

Here we show how to compute $H(Y_1^n)$ to any finite order in ϵ. We use the Markovian property to write $\Pr(Y_1^n)$ in the form :

$$\Pr(Y_1^n) = \sum_{X_1^n} \Pr(X_1^n, Y_1^n) = \sum_{X_1^n} \Pr(X_1^n) \Pr(Y_1^n | X_1^n) =$$

$$\sum_{X_1^n} \{\Pr(X_1) \prod_{i=1}^{n-1} \Pr(X_{i+1}|X_i) \prod_{i=1}^{n} \Pr(Y_i|X_i)\} \tag{9}$$

We now use the following change of variables : $\tau_i = (-1)^{X_i}$, $\sigma_i = (-1)^{Y_i}$. Since the process is stationary, we also have $\Pr(X_1 = 1) = \frac{1}{2}$. Thus, eq. 9 becomes ([9],[14]) :

$$\Pr(Y_1^n) = A_0 A_1 \sum_{\tau_1^n} e^{J \sum_{i=1}^{n-1} \tau_i \tau_{i+1} + K \sum_{i=1}^{n} \tau_i \sigma_i} \tag{10}$$

where J and K are related to p and ϵ, respectively, by :

$$e^{-2J} = \frac{p}{1-p}, \qquad e^{-2K} = \frac{\epsilon}{1-\epsilon} \tag{11}$$

and A_0, A_1 are normalizing constants given by :

$$A_0 = \frac{(e^J + e^{-J})^{1-n}}{2}, \qquad A_1 = (e^K + e^{-K})^{-n} \tag{12}$$

In statistical mechanics the form of eq. (10) is referred to as the one-dimensional Ising model [8], and the problem at hand is related to the Ising model in a quenched random field. The leading orders of $H(Y_1^n)$ in ϵ are found by a low-temperature/high-field expansion [2]. Non-analyticity of functions such as $H(p, \epsilon)$ can occur only at phase transitions. In one dimensional systems with short range interactions, at equilibrium, phase transitions can occur only at $p = 0$ or 1.

In order to compute the first and second orders in ϵ we take only realizations τ_1^n's which are different in at most two bits from σ_1^n in the summation in eq. (10). Using the low-temperature/high-field expansion, we obtain the following result :

$$H(Y_1^n) = -\sum_{Y_1^n} \Pr(Y_1^n) \log \Pr(Y_1^n) = n[H_0 + H_1\epsilon + H_2\epsilon^2 + O(\epsilon^3)] + D \tag{13}$$

The term $D = O(1)$ (in n). The coefficients H_k are given by :

$$\begin{aligned} H_0 &= -p\log p - (1-p)\log(1-p) \\ H_1 &= 2(1-2p)\log\left[\frac{1-p}{p}\right] \\ H_2 &= -2(1-2p)\log\left[\frac{1-p}{p}\right] - \frac{(1-2p)^2}{2p^2(1-p)^2} \end{aligned} \tag{14}$$

Although quadratic terms in n appear in intermediate steps of the calculation, they cancel out and we are left with a linear dependency of the entropy on n. This property is true when expanding to any order of ϵ, resulting from the fact that the entropy per-bit converges to a constant.

Any higher orders k can be calculated in a similar way, by allowing in the sum eq. (13) realizations τ_1^n that differ from the fixed σ_1^n in k bits or less. The number of terms to be calculated is, however, related to the number of partitions of k, which is exponential in $\sqrt{k}$ ([4]).

Notice that taking *i.i.d.* source for the X's, with $\Pr(X = 1) = p$, instead of a Markovian source, gives the same zero-order term, but the first order becomes $(1 - 2p)\log\frac{1-p}{p} = \frac{H_1}{2}$. Thus, for small noise, the noise effect on the entropy is roughly doubled.

3 Derivation using Upper-Bound of Cover and Thomas

For a given value of n, the upper-bound $C^{(n)}$ can be explicitly written as a function of p and ϵ, using the fact :

$$C^{(n)} = H(Y_n|Y_1^{n-1}) = H(Y_1^n) - H(Y_1^{n-1})$$

We can express $H(Y_1^n)$ as a function of p and ϵ by using eq. (9) to express $\Pr(Y_1^n)$ in terms of the original variables :

$$\Pr(Y_1^n) = \sum_{X_1^n} (1-p)^{\sum_{i=1}^n 1_{X_i = X_{i+1}}} p^{n - \sum_{i=1}^n 1_{X_i = X_{i+1}}} (1-\epsilon)^{\sum_{i=1}^n 1_{X_i = Y_i}} \epsilon^{n - \sum_{i=1}^n 1_{X_i = Y_i}} \tag{15}$$

Thus, $\Pr(Y_1^n)$ is given explicitly as a polynomial in p and ϵ with maximal degree n. Collecting its terms gives :

$$\Pr(Y_1^n) = \sum_{i=0}^n Q_i(Y_1^n)\epsilon^i \tag{16}$$

where $Q_i = Q_i(Y)$ are functions of p only.
Substituting this expansion in the definition eq. (2) of H, and expanding $\log \Pr(Y_1^n)$ according to the Taylor series $\log(a+x) = \log(a) - \sum_{k=1}^{\infty} \frac{(-x)^k}{ka^k}$, we get

$$H(Y_1^n) = -\sum_Y \left[\sum_{i=0}^{n} Q_i(Y_1^n)\epsilon^i\right] \left[\log Q_0(Y_1^n) - \sum_{k=1}^{L} \frac{(-\sum_{i=1}^{n} Q_i(Y_1^n)\epsilon^i)^k}{kQ_0(Y_1^n)^k}\right] + O(\epsilon^{L+1}) \tag{17}$$

For $L = 2$ we have

$$H(Y_1^n) = -\sum_Y \Bigg\{ Q_0(Y_1^n) \log Q_0(Y_1^n) + \left[Q_1(Y_1^n)(1 + \log Q_0(Y_1^n))\right] \epsilon +$$
$$\left[\frac{Q_1(Y_1^n)^2}{2Q_0(Y_1^n)} + Q_2(Y_1^n)(1 + \log Q_0(Y_1^n))\right] \epsilon^2 \Bigg\} + O(\epsilon^3) \tag{18}$$

When extended to terms of order ϵ^k, this equation gives us precisely the expansion of the upper-bound $C^{(n)}$ up to the $k-th$ order.
The zeroth and first order terms can be evaluated analytically; beyond first order, we can compute the expansion of $H(Y_1^n)$ symbolically (using maple), for any finite n, as a function of p and ϵ (the computation we have done is exponential in n, but the complexity can be improved).

3.1 Derivation of First Order

We use now the upper-bound $C_1^{(n)}$, as an alternative method for obtaining the first order term in (18) :

$$\frac{\partial H(Y_1^n)}{\partial \epsilon}|_{(p,0)} = -\sum_Y Q_1(Y_1^n)[1 + \log Q_0(Y_1^n)] \tag{19}$$

But, using eq. (15) :

$$\Pr(Y) = \sum_X \Pr(X,Y) = \Pr(X=y,Y) + \sum_{i=1}^{n} \Pr(X = y(\bar{i}), Y) + O(\epsilon^2) =$$
$$(1-n\epsilon)\Pr(X=y) + \epsilon \sum_{i=1}^{n} \Pr(X = y(\bar{i})) + O(\epsilon^2) =$$
$$\Pr(X=y) + \left[-n\Pr(X=y) + \sum_{i=1}^{n} \Pr(X = y(\bar{i}))\right] \epsilon + O(\epsilon^2) \tag{20}$$

So :

$$\sum_y Q_1(y) = \sum_y [-n\Pr(X=y) + \sum_{i=1}^{n} \Pr(X = y(\bar{i}))] = -n + n = 0 \tag{21}$$

Using (3) and (19) we get :

$$C_1^{(n)} = \sum_{Y_1^{n+1}} Q_1(Y_1^{n+1}) \log Q_0(Y_1^{n+1}) - \sum_{Y_1^n} Q_1(Y_1^n) \log Q_0(Y_1^n) \tag{22}$$

In order to prove that $C_1^{(n)}$ is constant, independent of n, we need a finer definition of the orders, which is given by :

$$\Pr(X_n = y_n, Y_1^n) = \sum_{i=0}^{n} Q_i^0(Y_1^n)\epsilon^i, \qquad \Pr(X_n = \overline{y_n}, Y_1^n) = \sum_{i=0}^{n} Q_i^1(Y_1^n)\epsilon^i \tag{23}$$

Thus, Q_i^0 (Q_i^1) is the i-th order of the fraction of $\Pr(Y_1^n)$ for which the last bit is equal (different) to the source bit. Clearly, $Q_i^0 + Q_i^1 = Q_i$, $\forall i \in \mathbb{N}$. Using the above definitions, and noting that $Q_0^1 \equiv 0$ (thus $Q_0^0 = Q_0$), we get a relation between the terms for n and $n+1$ bits :

$$\begin{aligned}\Pr(Y_1^n, Y_{n+1} = y_n) = ((1-p)(1-\epsilon) + p\epsilon)(Q_0(Y_1^n) + \epsilon Q_1^0(Y_1^n)) + (p(1-\epsilon)+ \\ (1-p)\epsilon)\epsilon Q_1^1(Y_1^n) = (1-p)Q_0(Y_1^n) + \epsilon[(2p-1)Q_0(Y_1^n) + (1-p)Q_1^0(Y_1^n)+ \\ pQ_1^1(Y_1^n)] + O(\epsilon^2)\end{aligned} \tag{24}$$

And similarly :

$$\begin{aligned}\Pr(Y_1^n, Y_{n+1} = \overline{y_n}) = pP_0^0(Y_1^n) + \epsilon[(1-2p)Q_0(Y_1^n)+ \\ pQ_1^0(Y_1^n) + (1-p)Q_1^1(Y_1^n)] + O(\epsilon^2)\end{aligned} \tag{25}$$

By substituting in eq. (22), we can verify that :

$$\begin{aligned}C_1^{(n)} = \sum_{Y_1^n} \Big\{ [(2p-1)Q_0(Y_1^n) + (1-p)Q_1^0(Y_1^n) + pQ_1^1(Y_1^n)] \log((1-p)Q_0(Y_1^n))+ \\ [(1-2p)Q_0(Y_1^n) + pQ_1^0(Y_1^n) + (1-p)Q_1^1(Y_1^n)] \log(pQ_0(Y_1^n)) \Big\} \\ - \sum_{Y_1^n} (Q_1^0(Y_1^n) + Q_1^1(Y_1^n)) \log Q_0(Y_1^n) = \\ \sum_{Y_1^n} \Big\{ [(2p-1)Q_0^0(Y_1^n) + (1-p)Q_1^0(Y_1^n) + pQ_1^1(Y_1^n)] \log(1-p)+ \\ [(1-2p)Q_0(Y_1^n) + pQ_1^0(Y_1^n) + (1-p)Q_1^1(Y_1^n)] \log p \Big\} = \\ ((1-p)\log p + p\log(1-p)) \sum_{Y} Q_1^1(Y_1^n)+\end{aligned}$$

$$(p\log p + (1-p)\log(1-p))\sum_Y Q_1^0(Y_1^n) - (2p-1)\log\frac{p}{1-p}\sum_Y Q_0(Y_1^n) \qquad (26)$$

Now, noting that :

$$1 = \sum_Y Q_0^0(Y) = \sum_Y Q_1^1(Y) = 1 - \sum_Y Q_1^0(Y)$$

and substituting in 26, gives :

$$C_1^{(n)} = 2(1-2p)\log\frac{1-p}{p} \qquad (27)$$

which is identical to H_1 in eq (14).

3.2 Higher Order Terms

Symbolic computation of higher order terms yielded the same independence of n for large enough n, as proved above for $k = 1$. For example, computing $C_2^{(n)}$ we found that its value for $n = 3, 4, ..., 11$ is independent of n and given by the exact H_2 of eq. (14). Similarly, $C_3^{(n)}$ settles, for $n \geq 3$, at the value denoted by H_3 in the Appendix, and so on.
The first orders up to H_{11} are given in the Appendix, as functions of $\lambda = 1 - 2p$, for better readability. The values of H_0, H_1 and H_2 coincide with the results that were derived rigorously from the low-temperature/high-field expansion, thus giving us a clue for postulating Conjecture 1.
Interestingly, the nominators have a simpler expression when considering them as a function of λ, which is the 2^{nd} eigenvalue of the Markov transition matrix P. Note that only even powers of λ appear. Another interesting observation is that the free element in $[p(1-p)]^{2(k-1)}H_k$ (when treated as a polynomial in p), is $\frac{(-1)^k}{k(k-1)}$, which might suggest some role for the function $\log(1 + \frac{\epsilon}{[2p(1-p)]^2})$ in the first derivative of H. All of the above observations are summarized in writing Conjecture 2.

4 Discussion

We have shown a method for calculating arbitrary orders of the expansion of the entropy rate in the noise variable for binary HMPs. A practical issue concerns the radius of convergence $R(p)$ of the series (4). This topic is under current study; we have shown that $R(p) << 1$ for small p and increases with p [15]. The validity of Conjecture 1 on the upper-bounds $C_k^{(n)}$ settling to a fixed value for large enough n needs to be better understood, as it might reveal new insights on the model. Another direction for future research is looking for more general HMPs , with arbitrary transition and omission matrices, for which even the first order in ϵ are not given by the upper-bounds. Another interesting regime, not addressed in [5] and [11] is $p \to 0$, with ϵ fixed. This delicate limit is under consideration and differs from the situation ($\epsilon \to 0$) discussed here, which gives further evidence for the non-symmetric character of the function H with respect to the parameters p and ϵ. Needless to say, the ultimate goal

is to find a closed form expression for the function $H(p, \epsilon)$ (or to prove that such expression does not exists.) Other subjects of interest are the index of coincidence for two independent identical HMPs, and the Kullback-Leibler divergence-rate between the Markov process X and its noisy observation Y; both properties seem to relate to the entropy rate. An analogue of Conjecture 1 holds (at least for small n and k) also for higher order HMPs, and will be addressed in [15].
Note that Conjecture 1, if true, may be used for bounding the error in the approximation using the upper-bounds. Since we can assume $\epsilon \leq \frac{1}{2}$, we get that the error in the $n-th$ term is no more that $2^{-2n}C(p)$, where the constant can be viewed as function of λ, which, not surprisingly, relates to the mixing time of the Markov chain.

Acknowledgments

I.K. thanks N. Merhav for very helpful comments, and the Einstein Center for Theoretical Physics for partial support. This work was partially supported by grants from the Minerva Foundation and by the European Community's Human Potential Programme under contract HPRN-CT-2002-00319, STIPCO.

References

1. T. M. Cover and J. A. Thomas, "Elements of Information Theory", Wiley, New York, 1991.
2. C. Domb, "Phase Transitions and Critical Phenomena", eds. C. Domb and M. Green, vol. 6 p. 357, Academic Press, London, New York , 1974.
3. Y. Ephraim and N. Merhav, "Hidden Markov processes", IEEE Trans. Inform. Theory, vol. 48, p. 1518-1569, June 2002.
4. G.H. Hardy and S. Ramanujan, "Asymptotic Formulae in Combinatory Analysis", Proc. London Math. Soc. 17, p. 75-115, 1918
5. P. Jacquet, G. Seroussi and W. Szpankowski, "On the Entropy of a Hidden Markov Process", Data Compression Conference , Snowbird, 2004.
6. R. Khasminskii and O. Zeitouni, "Asymptotic filtering for finite state Markov chains", Stoch. Proc. Appl. 63, p. 1-10, 1996.
7. L.D. Landau and E.M. Lifshitz, "Statistical Physics Part 1", 3rd ed. vol. 5 p. 537, Pergamon Press, Oxford, 1980.
8. E.H.Lieb and D.C. Mattis, "Mathematical Physics in one Dimension", Academic Press, New York, 1961.
9. D. J.C. MacKay, "Equivalence of Boltzmann Chains and Hidden Markov Models", Neural Computation, vol. 8 (1), p. 178-181, Jan 1996.
10. http://www.maplesoft.com/
11. E. Ordentlich and T. Weissman, "On the optimality of symbol-by-symbol filtering and denoising", submitted to IEEE Tra. Inf. Th.
12. E. Ordentlich and T. Weissman, "New Bounds on the Entropy Rate of Hidden Markov Processes", San Antonio Information Theory Workshop, Oct 2004.

13. L. R. Rabiner, ”A tutorial on hidden Markov models and selected applications in speech recognition”, Proc. IEEE, vol. 77, p. 257286, Feb 1989.
14. L. K. Saul and M. I. Jordan, ”Boltzmann chains and hidden Markov models”, Advances in Neural Information Processing Systems 7, MIT Press, 1994.
15. O. Zuk, I. Kanter and E. Domany, in preparation.

Appendix

Orders three to eleven, as function of $\lambda = 1 - 2p$. (Orders $0 - 2$ are given in eq. 14) :

$$H_3 = \frac{-16(5\lambda^4 - 10\lambda^2 - 3)\lambda^2}{3(1-\lambda^2)^4}$$

$$H_4 = \frac{8(109\lambda^8 + 20\lambda^6 - 114\lambda^4 - 140\lambda^2 - 3)\lambda^2}{3(1-\lambda^2)^6}$$

$$H_5 = \frac{-128(95\lambda^{10} + 336\lambda^8 + 762\lambda^6 - 708\lambda^4 - 769\lambda^2 - 100)\lambda^4}{15(1-\lambda^2)^8}$$

$$H_6 = 128(125\lambda^{14} - 321\lambda^{12} + 9525\lambda^{10} + 16511\lambda^8 - 7825\lambda^6 - \\ 17995\lambda^4 - 4001\lambda^2 - 115)\lambda^4/15(1-\lambda^2)^{10}$$

$$H_7 = -256(280\lambda^{18} - 45941\lambda^{16} - 110888\lambda^{14} + 666580\lambda^{12} + 1628568\lambda^{10} - \\ 270014\lambda^8 - 1470296\lambda^6 - 524588\lambda^4 - 37296\lambda^2 - 245)\lambda^4/105(1-\lambda^2)^{12}$$

$$H_8 = 64(56\lambda^{22} - 169169\lambda^{20} - 2072958\lambda^{18} - 5222301\lambda^{16} + 12116328\lambda^{14} + \\ 35666574\lambda^{12} + 3658284\lambda^{10} - 29072946\lambda^8 - 14556080\lambda^6 - \\ 1872317\lambda^4 - 48286\lambda^2 - 49)\lambda^4/21(1-\lambda^2)^{14}$$

$$H_9 = 2048(37527\lambda^{22} + 968829\lambda^{20} + 8819501\lambda^{18} + 20135431\lambda^{16} - 23482698\lambda^{14} - \\ 97554574\lambda^{12} - 30319318\lambda^{10} + 67137630\lambda^8 + 46641379\lambda^6 + 8950625\lambda^4 + \\ 495993\lambda^2 + 4683)\lambda^6/63(1-\lambda^2)^{16}$$

$$H_{10} = -2048(38757\lambda^{26} + 1394199\lambda^{24} + 31894966\lambda^{22} + 243826482\lambda^{20} + \\ 571835031\lambda^{18} - 326987427\lambda^{16} - 2068579420\lambda^{14} - 1054659252\lambda^{12} + \\ 1173787011\lambda^{10} + 1120170657\lambda^8 + 296483526\lambda^6 + 26886370\lambda^4 +$$

$$684129\lambda^2 + 2187)\lambda^6/45(1-\lambda^2)^{18}$$

$$H_{11} = 8192(98142\lambda^{30} - 1899975\lambda^{28} + 92425520\lambda^{26} + 3095961215\lambda^{24} +$$

$$25070557898\lambda^{22} + 59810870313\lambda^{20} - 11635283900\lambda^{18} - 173686662185\lambda^{16} -$$

$$120533821070\lambda^{14} + 74948247123\lambda^{12} + 102982107048\lambda^{10} + 35567469125\lambda^{8} +$$

$$4673872550\lambda^6 + 217466315\lambda^4 + 2569380\lambda^2 + 2277)\lambda^6/495(1-\lambda^2)^{20}$$