ON THE INTEGRALITY OF THE TAYLOR COEFFICIENTS OF MIRROR MAPS, II

C. KRATTENTHALER† AND T. RIVOAL

ABSTRACT. We continue our study begun in *"On the integrality of the Taylor coefficients of mirror maps"* [Duke Math. J. (to appear)] of the fine integrality properties of the Taylor coefficients of the series $\mathbf{q}(z) = z\exp(\mathbf{G}(z)/\mathbf{F}(z))$, where $\mathbf{F}(z)$ and $\mathbf{G}(z)+\log(z)\mathbf{F}(z)$ are specific solutions of certain hypergeometric differential equations with maximal unipotent monodromy at $z = 0$. More precisely, we address the question of finding the largest integer v such that the Taylor coefficients of $(z^{-1}\mathbf{q}(z))^{1/v}$ are still integers. In particular, we determine the Dwork–Kontsevich sequence $(u_N)_{N\geq 1}$, where u_N is the largest integer such that $q_N(z)^{1/u_N}$ is a series with integer coefficients, where $q_N(z) = \exp(G_N(z)/F_N(z))$, $F_N(z) = \sum_{m=0}^{\infty}(Nm)!\,z^m/m!^N$ and $G_N(z) = \sum_{m=1}^{\infty}(H_{Nm}-H_m)(Nm)!\,z^m/m!^N$, with H_n denoting the n-th harmonic number, conditional on the conjecture that there are no prime number p and integer N such that the p-adic valuation of $H_N - 1$ is strictly greater than 3.

1. Introduction and statement of results

The present article is a sequel to our article [6], where we proved general results concerning the integrality properties of mirror maps. We shall prove here stronger integrality assertions for certain special cases that appear frequently in the literature. For any vector $\mathbf{N} = (N,\dots,N)$ (with k occurrences of N), where N is a positive integer, let us define the power series

$$F_{\mathbf{N}}(z) = \sum_{m=0}^{\infty}\frac{(Nm)!^k}{m!^{kN}}z^m \tag{1.1}$$

and

$$G_{\mathbf{N}}(z) = \sum_{m=1}^{\infty}kN(H_{Nm}-H_m)\frac{(Nm)!^k}{m!^{kN}}z^m, \tag{1.2}$$

with $H_n := \sum_{i=1}^{n}\frac{1}{i}$ denoting the n-th harmonic number. The functions $F_{\mathbf{N}}(z)$ and $G_{\mathbf{N}}(z)+\log(z)F_{\mathbf{N}}(z)$ are solutions of the same hypergeometric differential equation with maximal unipotent monodromy at $z = 0$. A basis of solutions with at most logarithmic singularities

Date: September 17, 2009.

2000 *Mathematics Subject Classification.* Primary 11S80; Secondary 11J99 14J32 33C20.

Key words and phrases. Calabi–Yau manifolds, integrality of mirror maps, p-adic analysis, Dwork's theory, harmonic numbers, hypergeometric differential equations.

†Research partially supported by the Austrian Science Foundation FWF, grants Z130-N13 and grant S9607-N13, the latter in the framework of the National Research Network "Analytic Combinatorics and Probabilistic Number Theory".

around $z = 0$ can then be obtained by Frobenius' method; see [11]. That differential equation is the Picard–Fuchs equation of a one parameter family of mirror manifolds W' of a complete intersection W of k hypersurfaces $W_1, \dots, W_k$, all of degree N in $\mathbb{P}^{d+k}(\mathbb{C})$: W is a family of Calabi–Yau manifolds if one chooses d equal to $k(N-1)-1$. The mirrors W' are explicitly constructed in [1, Sec. 5.2]. In the underlying context of mirror symmetry, it is natural to define the canonical coordinate $q_{\mathbf{N}}(z) := z \exp\big(G_{\mathbf{N}}(z)/F_{\mathbf{N}}(z)\big)$ and the mirror map $z_{\mathbf{N}}(q)$, which is the compositional inverse of $q_{\mathbf{N}}(z)$.

In [6], we proved the following result, which settled a conjecture in the folklore of mirror symmetry theory.

> *For any integers $k \geq 1$ and $N \geq 1$, we have $q_{\mathbf{N}}(z) \in z\mathbb{Z}[[z]]$ and $z_{\mathbf{N}}(q) \in q\mathbb{Z}[[q]]$, where N is repeated k times in the vector $\mathbf{N} = (N, \dots, N)$.* ([1]) (1.3)

Lian and Yau [8, Sec. 5, Theorem 5.5] had proved earlier the particular case of this theorem where $k = 1$ and N is a prime number, and Zudilin [12, Theorem 3] had extended their result to any $k \geq 1$ and any N which is a prime power. Zudilin also formulated a more general conjecture, implying (1.3), which he proved in a particular case. The conjecture was subsequently fully proved as one of the main results in [6].

For $k = 1$, physicists made the observation that, apparently, even the stronger assertion

$$\big(z^{-1} q_{(N)}(z)\big)^{1/N} \in \mathbb{Z}[[z]] \tag{1.4}$$

holds. This was proved by Lian and Yau [9] for any prime number N, thus strengthening their result from [8] mentioned above. The observation (1.4) leads naturally to the more general question of determining the largest integer V such that $(z^{-1}q(z))^{1/V} \in \mathbb{Z}[[z]]$ for mirror maps $q(z)$ such as $q_{\mathbf{N}}(z)$. ([1]) A rather general result in this direction has already been obtained in [6, Theorem 2]. The purpose of the present paper is to sharpen this earlier result for wide classes of special choices of the parameters occurring in [6]. Indeed, our main results, given in Theorems 1 and 2 below, provide values of V for infinite families of mirror(-type) maps, which, conditional on widely believed conjectures on the p-adic valuations of H_N respectively $H_N - 1$, are optimal for these families.

---

[1] In the number-theoretic study undertaken in the present paper, we are interested in the integrality of the coefficients of roots of mirror maps $z(q)$. In that context, $z(q)$ and the corresponding canonical coordinate $q(z)$ play strictly the same role, because $(z^{-1}q(z))^{1/\tau} \in 1 + z\mathbb{Z}[[z]]$ for some integer τ implies that $(q^{-1}z(q))^{1/\tau} \in 1 + q\mathbb{Z}[[q]]$, and conversely; see [7, Introduction]. We shall, in the sequel, formulate our integrality results exclusively for canonical coordinates. By abuse of terminology, we shall often use the term "mirror map" for any canonical coordinate.

[2] Let $q(z)$ be a given power series in $\mathbb{Z}[[z]]$, and let V be the largest integer with the property that $q(z)^{1/V} \in \mathbb{Z}[[z]]$. Then V carries complete information about *all* integers with that property: namely, the set of integers U with $q(z)^{1/U} \in \mathbb{Z}[[z]]$ consists of all divisors of V. Indeed, it is clear that all divisors of V belong to this set. Moreover, if U_1 and U_2 belong to this set, then also $\mathrm{lcm}(U_1, U_2)$ does (cf. [5, Lemma 5] for a simple proof based on Bézout's lemma).

To describe our results, for positive integers L and N, we set

$$G_{L,\mathbf{N}}(z) = \sum_{m=1}^{\infty} H_{Lm} \frac{(Nm)!^k}{m!^{kN}} z^m, \tag{1.5}$$

where again $\mathbf{N} = (N, N, \dots, N)$, with k occurrences of N. We then define the mirror-type map $q_{L,\mathbf{N}}(z) := \exp\big(G_{L,\mathbf{N}}(z)/F_{\mathbf{N}}(z)\big)$. Obviously, the mirror map $q_{\mathbf{N}}(z)$ can be expressed as a product of the series $q_{L,\mathbf{N}}(z)$, namely as

$$q_{\mathbf{N}}(z) = z q_{N,\mathbf{N}}^{kN}(z) q_{1,\mathbf{N}}^{-kN}(z). \tag{1.6}$$

The special case of the afore-mentioned Theorem 2 from [6] where $\mathbf{N} = (N, N, \dots, N)$ (with k occurrences of N) addressed the above question of "maximal integral roots" for the mirror-type map $q_{L,\mathbf{N}}(z)$. It reads as follows:

> *Let $\Theta_L := L!/\gcd(L!, L!\, H_L)$ be the denominator of H_L when written as a reduced fraction. For any positive integers N and L with $L \le N$, we have $q_{L,\mathbf{N}}(z)^{\frac{\Theta_L}{N!^k}} \in \mathbb{Z}[[z]]$.* (1.7)

As we remarked in [6], this result is optimal in the case that $L = 1$; that is, no integer V larger than $N!^k/\Theta_1 = N!^k$ can be found such that $q_{L,\mathbf{N}}(z)^{1/V} \in \mathbb{Z}[[z]]$. However, if $L \ge 2$, improvements may be possible.

Our first main result provides such an improvement. In order to state the result, we need to introduce usual notation for p-adic valuation: given a prime number p (and from now on, p will always denote a prime number) and $\alpha \in \mathbf{Q}_p$, $v_p(\alpha)$ denotes the p-adic valuation of α.

Theorem 1. *Let N be a positive integer, $\mathbf{N} = (N, N, \dots, N)$, with k occurrences of N, and let $\Xi_1 = 1$, $\Xi_7 = 1/140$, and, for $N \notin \{1, 7\}$,*

$$\Xi_N := \prod_{p \le N} p^{\min\{2+\xi(p,N), v_p(H_N)\}}, \tag{1.8}$$

where $\xi(p, N) = 1$ if p is a Wolstenholme prime (i.e., a prime p for which $v_p(H_{p-1}) \ge 3$ ([3])) or N is divisible by p, and $\xi(p, N) = 0$ otherwise. Then $q_{N,\mathbf{N}}(z)^{\frac{1}{\Xi_N N!^k}} \in \mathbb{Z}[[z]]$.

Remarks 1. For better comprehension, we discuss the meaning of the statement of Theorem 1 and its implications; in particular, we address some fine points of the definition of Ξ_N.

(a) The case of $N = 1$ is trivial since $q_{1,(1,\dots,1)}(z) = 1/(1-z)$. Furthermore, we have

$$\Xi_7 = \frac{1}{140} = 2^{v_2(H_7)} 5^{v_5(H_7)} 7^{v_7(H_7)},$$

[3]Presently, only two such primes are known, namely 16843 and 2124679, and it is unknown whether there are infinitely many Wolstenholme primes or not.

which differs by a factor of 3 from the right-hand side of (1.8) with $N = 7$ (since $v_3(H_7) = v_3(\frac{363}{140}) = 1$).

(b) Let V_N denotes the largest integer such that $\left(z^{-1}q_{N,\mathbf{N}}(z)\right)^{1/V_N}$ is a series with integer coefficients. Since $q_{N,\mathbf{N}}(z) = 1 + H_N N!^k z + \mathcal{O}(z^2)$, it is clear that $q_{N,\mathbf{N}}(z)^{1/(p^{v_p(H_N)+1}N!^k)} \notin \mathbb{Z}[[z]]$, so that the exponent of p in the prime factorisation of V_N can be at most $v_p(H_N N!^k)$. In Theorem 1, this theoretically maximal exponent is cut down to $v_p(\Xi_N N!^k)$. First of all, the number Ξ_N contains no prime factor $p > N$, whereas the harmonic number H_N may very well do so (and, in practice, always does for $N > 1$). Moreover, for primes p with $p \le N$ and $v_p(H_N) \ge 3$, the definition of Ξ_N cuts the theoretically maximal exponent $v_p(H_N N!^k)$ of p down to $2 + v_p(N!^k)$ respectively $3 + v_p(N!^k)$, depending on whether $\xi(p, N) = 0$ or $\xi(p, N) = 1$. In items (c)–(e) below, we address the question of how serious this cut is expected to be.

(c) Clearly, the minimum appearing in the exponent of p in the definition (1.8) of Ξ_N is $v_p(H_N)$ as long as $v_p(H_N) \le 2$. In other words, the exponent of p in the prime factorisation of Ξ_N depends largely on the p-adic behaviour of H_N. An extensive discussion of this topic, with many interesting results, can be found in [2]. We have as well computed a table of harmonic numbers H_N up to $N = 1000000$. ([4]) Indeed, the data suggest that pairs (p, N) with p prime, $p \le N$, and $v_p(H_N) \ge 3$ are not very frequent. More precisely, so far only five examples are known with $v_p(H_N) = 3$: four for $p = 11$, with $N = 848, 9338, 10583$, and $35464717222689162 72$, and one for $p = 83$ with

$$N = 797810791993600900669891438146765729615283994776995167863779943 70\backslash \\ 78839681692157676915245857235055200779421409821643691818 \tag{1.9}$$

(see [2, p. 289]; the value of N in (1.9), not printed in [2], was kindly communicated to us by David Boyd). There is no example known with $v_p(H_N) \ge 4$. It is, in fact, conjectured that no p and N exist with $v_p(H_N) \ge 4$. Some evidence for this conjecture (beyond mere computation) can be found in [2].

(d) Since in all the five examples for which $v_p(H_N) = 3$ we neither have $p \mid N$ (the gigantic number in (1.9) is congruent to 42 modulo 83) nor that the prime p is a Wolstenholme prime, the exponent of p in the prime factorisation of Ξ_N in these cases is 2 instead of $v_p(H_N) = 3$.

(e) On the other hand, should there be a prime p and an integer N with $p \le N$, $v_p(H_N) \ge 3$, p a Wolstenholme prime or $p \mid N$, then the exponent of p in the prime factorisation of Ξ_N would be 3. However, no such examples are known. We conjecture that there are no such pairs (p, N). If this conjecture should turn out to be true, then, given $N \notin \{1, 7\}$, the definition of Ξ_N in (1.8) could be simplified to

$$\Xi_N := \prod_{p \le N} p^{\min\{2, v_p(H_N)\}}. \tag{1.10}$$

[4]The summary of the table is available at `http://www.mat.univie.ac.at/~kratt/artikel/H.html`.

Theorem 1 improves upon (1.7) for $L = N$. Namely, if one compares the definition of Ξ_N in (1.8) with the following alternative way to write the integer Θ_N occurring in (1.7),

$$\Theta_N = \prod_{p \le N} p^{-\min\{0, v_p(H_N)\}}, \tag{1.11}$$

we see that Theorem 1 is *always* at least as strong as [6, Theorem 2], and it is *strictly* stronger if $N \neq 7$ and $v_p(H_N) \geq 1$ for some prime p less than or equal to N. Indeed, the smallest $N \neq 7$ with that property is $N = 20$, in which case $v_5(H_{20}) = 1$.

We remark that strengthenings of [6, Theorem 2] in the spirit of Theorem 1 for more general choices of the parameters can also be obtained by our techniques but are omitted here.

We outline the proof of Theorem 1 in Section 3, with the details being filled in in Sections 4–6. As we explain in Section 7, we conjecture that Theorem 1 cannot be improved if $k = 1$, that is, that for $k = 1$ the largest integer t_N such that $q_{N,(N)}(z)^{1/t_N} \in \mathbb{Z}[[z]]$ is exactly $\Xi_N N!$. Propositions 1 and 2 in Section 7 show that this conjecture would immediately follow if one could prove the conjecture from Remarks 1(c) above that there are no primes p and integers N with $v_p(H_N) \geq 4$.

Even if the series $q_{N,\mathbf{N}}(z)$ appears in the identity (1.6), which relates the mirror map $q_{\mathbf{N}}(z)$ to the series $q_{L,\mathbf{N}}(z)$ (with $L = 1$ and $L = N$), Theorem 1 does not imply an improvement over [6, Corollary 1], which we recall here for convenience.

For all integers $k \geq 1$ and $N \geq 1$, we have $\bigl(z^{-1}q_{\mathbf{N}}(z)\bigr)^{\frac{\Theta_N}{N!^k kN}} \in \mathbb{Z}[[z]]$, where Θ_N is defined in (1.7). (1.12)

The reason is that the coefficient of z in $(z^{-1}q_{\mathbf{N}}(z))^{\Theta_N/(pN!^k kN)}$ is equal to $\frac{\Theta_N}{p}(H_N - 1)$, and, thus, it will not be integral for primes p with $v_p(H_N) > 0$. Still, there is an improvement of (1.12) in the spirit of Theorem 1. It is our second main result, and it involves primes p with $v_p(H_N - 1) > 0$ instead.

Theorem 2. *Let N be a positive integer with $N \geq 2$, and let*

$$\Omega_N := \prod_{p \le N} p^{\min\{2+\omega(p,N), v_p(H_N-1)\}}, \tag{1.13}$$

where $\omega(p, N) = 1$ if p is a Wolstenholme prime or $N \equiv \pm 1 \bmod p$, and $\omega(p, N) = 0$ otherwise. Then $\bigl(z^{-1}q_{\mathbf{N}}(z)\bigr)^{\frac{1}{\Omega_N N!^k kN}} \in \mathbb{Z}[[z]]$.

Remarks 2. Also here, some remarks are in order to get a better understanding of the above theorem.

(a) If $v_p(H_N) < 0$, then $v_p(H_N) = v_p(H_N - 1)$. Hence, differences in the prime factorisations of Ξ_N and Ω_N can only arise for primes p with $v_p(H_N) \geq 0$.

(b) Since $z^{-1}q_{\mathbf{N}}(z) = 1 + (H_N - 1)N!^k kNz + \mathcal{O}(z^2)$, it is clear that

$$\bigl(z^{-1}q_{\mathbf{N}}(z)\bigr)^{1/(p^{v_p(H_N-1)+1}N!^k kN)} \notin \mathbb{Z}[[z]].$$

If $\widetilde{V}_N$ denotes the largest integer such that $\bigl(z^{-1}q_{\mathbf{N}}(z)\bigr)^{1/\widetilde{V}_N kN}$ is a series with integer coefficients, the exponent of p in $\widetilde{V}_N$ can be at most $v_p\bigl((H_N-1)N!^k\bigr)$. In Theorem 2, this theoretically maximal exponent is cut down to $v_p(\Omega_N N!^k)$. Namely, as is the case for Ξ_N, the number Ω_N in (1.13) contains no prime factor $p > N$, whereas the difference $H_N - 1$ may very well do so (and, in practice, always does for $N > 2$). Moreover, for primes p with $p \le N$ and $v_p(H_N - 1) \ge 3$, the definition of Ω_N cuts the theoretically maximal exponent $v_p\bigl((H_N-1)N!^k\bigr)$ of p down to $2 + v_p(N!^k)$ respectively $3 + v_p(N!^k)$, depending on whether $\omega(p, N) = 0$ or $\omega(p, N) = 1$.

(c) Concerning the question whether there are any primes p and integers N with high values of $v_p(H_N - 1)$, we are not aware of any corresponding literature. Our table of harmonic numbers H_N mentioned in Remarks 1(c) does not contain any pair (p, N) with $v_p(H_N - 1) \ge 3$. ([5]) In "analogy" to the conjecture mentioned in Remarks 1(c), we conjecture that no p and N exist with $v_p(H_N - 1) \ge 4$. It may even be true that there are no p and N with $v_p(H_N - 1) \ge 3$, in which case the definition of Ω_N in (1.13) could be simplified to

$$\Omega_N := \prod_{p \le N} p^{\min\{2, v_p(H_N - 1)\}}. \tag{1.14}$$

In view of (1.11), Theorem 2 improves upon (1.12). Namely, Theorem 2 is *always* at least as strong as (1.12), and it is *strictly* stronger if $v_p(H_N - 1) \ge 1$ for some prime p less than or equal to N. The smallest N with that property is $N = 21$, in which case $v_5(H_{21} - 1) = 1$.

We sketch the proof of Theorem 2 in Section 8. We also explain in that section that we conjecture that Theorem 2 with $k = 1$ is optimal, that is, that for $k = 1$ the largest integer u_N such that $\bigl(z^{-1}q_{(N)}(z)\bigr)^{\frac{1}{Nu_N}} \in \mathbb{Z}[[z]]$ is exactly $\Omega_N N!$. Propositions 3 and 4 in Section 8 show that this conjecture would immediately follow if one could prove the conjecture from Remarks 2(c) above that there are no primes p and integers N with $v_p(H_N - 1) \ge 4$. As a matter of fact, the sequence $(u_{2N})_{N \ge 1}$ appears in the On-Line Encyclopedia of Integer Sequences [10], as sequence `A007757`, contributed around 1995 by R. E. Borcherds under the denomination "Dwork–Kontsevich sequence," without any reference or explicit formula for it, however. ([6])

2. Structure of the paper

We now briefly review the organisation of the rest of the paper. Following the steps of previous authors, our approach for proving Theorems 1 and 2 uses p-adic analysis. In particular, we make essential use of Dwork's p-adic theory (in the spirit of [9]). Since the details of our proofs are involved, we provide brief outlines of the proofs of Theorems 1 and

[5] The summary of the corresponding table, containing pairs (p, N) with $p \le N$ and $v_p(H_N - 1) > 0$, is available at `http://www.mat.univie.ac.at/~kratt/artikel/H1.html`.

[6] In private communication, both, Borcherds and Kontsevich could not remember where exactly this sequence and its denomination came from.

2 in separate sections. Namely, Section 3 provides an outline of the proof of Theorem 1, while Section 8 contains a sketch of the proof of Theorem 2. Both follow closely the chain of arguments used in the proof of [6, Theorem 2]. In the outline, respectively sketch, the proofs are reduced to a certain number of lemmas. The lemmas which are necessary for the proof of Theorem 1 are established in Sections 4–6, while the lemmas which are necessary for the proof of Theorem 2 are contained in Section 8. Section 7 reports on the evidence to believe (or not to believe) that the value t_N (defined in the next-to-last paragraph before Theorem 2) is given by $t_N = \Xi_N N!$, $N = 1, 2, \dots$, while Section 9 addresses the question of whether the Dwork–Kontsevich sequence $(u_N)_{N\geq 1}$ (defined in the last paragraph of the Introduction) is (or is not) given by $u_N = \Omega_N N!$, $N = 1, 2, \dots$.

3. Outline of the proof of Theorem 1

In this section, we provide a brief outline of the proof of Theorem 1, reducing it to Lemmas 3–14 and Corollaries 1 and 2, the proofs of which are postponed to Sections 4–6, except that Lemma 4 has already been established in [6, Lemma 6].

The whole proof is heavily based on Dwork's p-adic theory (as presented by Lian and Yau in [9]), enhanced in the spirit of [6]. In particular, there we used the following result (cf. [6, Lemma 10]).

Lemma 1. *Given two formal power series $f(z) \in 1+z\mathbb{Z}[[z]]$ and $g(z) \in z\mathbb{Q}[[z]]$, an integer $\tau \geq 1$ and a prime number p, we have $\exp\big(g(z)/(\tau f(z))\big) \in 1 + z\mathbb{Z}_p[[z]]$ if and only if*

$$f(z)g(z^p) - p\, f(z^p)g(z) \in p\tau z\mathbb{Z}_p[[z]]. \tag{3.1}$$

It follows from Lemma 1 that Theorem 1 can be reduced to the following statement: for any prime number p, we have

$$F_{\mathbf{N}}(z)G_{N,\mathbf{N}}(z^p) - pF_{\mathbf{N}}(z^p)G_{N,\mathbf{N}}(z) \in p\Xi_N N!^k z\mathbb{Z}_p[[z]].$$

We now follow the presentation in [6] and let $0 \leq a < p$ and $K \geq 0$. The $(a + Kp)$-th Taylor coefficient of $F_{\mathbf{N}}(z)G_{N,\mathbf{N}}(z^p) - pF_{\mathbf{N}}(z^p)G_{N,\mathbf{N}}(z)$ is

$$C(a + Kp) := \sum_{j=0}^{K} B_{\mathbf{N}}(a + jp)B_{\mathbf{N}}(K - j)(H_{N(K-j)} - pH_{Na+Njp}), \tag{3.2}$$

where $B_{\mathbf{N}}(m) = \frac{(Nm)!^k}{m!^{kN}}$. As said above, proving Theorem 1 turns out to be equivalent to proving that

$$C(a + Kp) \in p\Xi_N N!^k\mathbb{Z}_p \tag{3.3}$$

for all primes p and non-negative integers a and K with $0 \leq a < p$.

There are two cases which can be treated directly: if $K = a = 0$, then $C(0) = 0$, and thus (3.3) holds trivially, whereas if $K = 0$ and $a = 1$, then $C(1) = -pB_{\mathbf{N}}(1)H_N = -pN!^k H_N$, and thus (3.3) holds by definition of Ξ_N. We therefore assume in addition $a + Kp \geq 2$ for the remainder of this section.

The following lemma will be useful. Its (simple) proof can be found in [6, Lemma 4].

Lemma 2. *For all integers $m \geq 1$ and $N \geq 1$, we have*

$$\frac{(Nm)!}{m!^N} \in N!\,\mathbb{Z}.$$

We deduce in particular that $B_{\mathbf{N}}(m) \in N!^k\mathbb{Z}$ for any $m \geq 1$.

Since

$$H_J = \sum_{j=1}^{\lfloor J/p \rfloor} \frac{1}{pj} + \sum_{\substack{j=1 \\ p\nmid j}}^{J} \frac{1}{j},$$

we have

$$pH_J \equiv H_{\lfloor J/p \rfloor} \mod p\mathbb{Z}_p. \tag{3.4}$$

Applying this to $J = Na + Njp$, we get

$$pH_{Na+Njp} \equiv H_{\lfloor Na/p \rfloor + Nj} \mod p\mathbb{Z}_p.$$

By Lemma 2, we infer

$$C(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)(H_{N(K-j)} - H_{\lfloor Na/p \rfloor + Nj}) \mod pN!^k\mathbb{Z}_p. \tag{3.5}$$

Indeed, if $K \geq 1$ or $a \geq 1$, this is because $a+jp$ and $K-j$ cannot be simultaneously zero and therefore at least one of $B_{\mathbf{N}}(a+jp)$ or $B_{\mathbf{N}}(K-j)$ is divisible by $N!^k$ by Lemma 2.

As long as $v_p(H_N) \leq 0$, Equation (3.5) implies

$$C(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)(H_{N(K-j)} - H_{\lfloor Na/p \rfloor + Nj}) \mod p\Xi_N N!^k\mathbb{Z}_p. \tag{3.6}$$

We claim that this congruence holds also for p and N with $v_p(H_N) > 0$.

For $p = 2$, we observe that we have always $v_2(H_N) \leq 0$ because of Lemma 8, so that no improvement over (3.5) is needed in this case. Furthermore, Lemma 9 together with Remarks 1(a) in the Introduction tells us that, if $p = 3$, we need an improvement only if $N = 22$. To be precise, for $N = 22$ we need to show that

$$C(a+3K) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+3j)B_{\mathbf{N}}(K-j)(H_{N(K-j)} - H_{\lfloor Na/3 \rfloor + Nj}) \mod 3^2N!^k\mathbb{Z}_3. \tag{3.7}$$

For $p \geq 5$, we should prove

$$C(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)(H_{N(K-j)} - H_{\lfloor Na/p \rfloor + Nj}) \mod p^3N!^k\mathbb{Z}_p, \tag{3.8}$$

and if, in addition, $v_p(H_N) \geq 3$ and p is a Wolstenholme prime or if $v_p(H_N) \geq 3$ and $p \mid N$ (the reader should recall the definition (1.8) of Ξ_N), then we need to show the even

stronger assertion that

$$C(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)(H_{N(K-j)} - H_{\lfloor Na/p \rfloor + Nj}) \mod p^4 N!^k \mathbb{Z}_p. \quad (3.9)$$

In order to see (3.7)–(3.9), we begin by combining our assumption $v_p(H_N) > 0$ with Corollary 1, Lemma 11(1), and (3.4), to obtain that

$$B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)pH_{Na+Njp} \equiv B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)H_{\lfloor Na/p \rfloor + Nj} \mod p^3 N!^k \mathbb{Z}_p$$

as long as $a + Kp \geq 2$ and $a \neq 0$. Moreover, due to Lemma 2 and Corollary 2, the above congruence is even true if $a = 0$ and $p \geq 5$. This implies (3.8). The congruence (3.7) follows in the same way by the slightly weaker assertion for $p = 3$ in Corollary 2.

For the congruence (3.9), one needs to combine Corollary 1 and Lemma 11(4), to see that

$$B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)pH_{Na+Njp} \equiv B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)H_{\lfloor Na/p \rfloor + Nj} \mod p^4 N!^k \mathbb{Z}_p \quad (3.10)$$

as long as $a + Kp \geq 2$ and $a \neq 0$. If $a = 0$, then the congruence (3.10) still holds as long as $j < K$ because of Lemma 2, Corollary 2, and the fact that the term $B_{\mathbf{N}}(K-j)$ contributes at least one factor p. The only remaining case to be discussed is $a = 0$ and $j = K$. If we apply the simple observation that $v_p(B_{\mathbf{N}}(p^e h)) = v_p(B_{\mathbf{N}}(h))$ for any positive integers e and h to $B_{\mathbf{N}}(a+jp) = B_{\mathbf{N}}(jp)$, then, making again appeal to Corollary 1 and Lemma 11(4), we see that (3.10) holds as well as long as $j = K$ is no prime power. Finally, let $a = 0$ and $j = K$ be a prime power, $j = K = p^f$ say. If $f \geq 1$, then we may use Lemma 13 with $J = a + jp = a + Kp = p^{f+1}$ and $v_p(B_{\mathbf{N}}(p^{f+1})) = v_p(B_{\mathbf{N}}(1)) = v_p(N!^k)$ to conclude that (3.10) also holds in this case. On the other hand, if $j = K = 1$, then Lemma 14, the assumption that p is a Wolstenholme prime or that p divides N, the fact that $v_p(B_{\mathbf{N}}(p)) = v_p(B_{\mathbf{N}}(1)) = v_p(N!^k)$, altogether yield the congruence (3.10) in this case as well.

We now want to transform the sum on the right-hand side of (3.6) to a more manageable expression. In particular, we want to get rid of the floor function $\lfloor Na/p \rfloor$. In order to achieve this, we will prove the following lemma in Section 4.

Lemma 3. *For any prime p, non-negative integers a, j, K with $0 \leq a < p$ and $K \geq 1$. If $a \neq 1$ or $j \geq 1$, we have*

$$B_{\mathbf{N}}(a+pj)\left(H_{Nj+\lfloor Na/p \rfloor} - H_{Nj}\right) \in p\Xi_N N!^k \mathbb{Z}_p \ , \quad (3.11)$$

and, in the case that $a = 1$ and $j = 0$, we have

$$B_{\mathbf{N}}(1)B_{\mathbf{N}}(K)H_{\lfloor N/p \rfloor} \in p\Xi_N N!^k \mathbb{Z}_p \ , \quad (3.12)$$

where $\mathbf{N} = (N, N, \ldots, N)$, with k occurrences of N.

It follows from Eq. (3.6) and Lemma 3 that

$$C(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)\Big(H_{N(K-j)} - H_{Nj}\Big) \mod p\Xi_N N!^k \mathbb{Z}_p \ ,$$

which can be rewritten as

$$C(a+Kp) \equiv -\sum_{j=0}^{K} H_{Nj}\Big(B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)-B_{\mathbf{N}}(j)B_{\mathbf{N}}(a+(K-j)p)\Big) \mod p\Xi_N N!^k\mathbb{Z}_p\,. \tag{3.13}$$

We now use a combinatorial lemma due to Dwork (see [3, Lemma 4.2]) which provides an alternative way to write the sum on the right-hand side of (3.13): namely, we have

$$\sum_{j=0}^{K} H_{Nj}\Big(B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j) - B_{\mathbf{N}}(j)B_{\mathbf{N}}(a+(K-j)p)\Big) = \sum_{s=0}^{r}\sum_{m=0}^{p^{r+1-s}-1} Y_{m,s}, \tag{3.14}$$

where r is such that $K < p^r$, and

$$Y_{m,s} := \big(H_{Nmp^s} - H_{N\lfloor m/p\rfloor p^{s+1}}\big)S(a,K,s,p,m),$$

the expression $S(a,K,s,p,m)$ being defined by

$$S(a,K,s,p,m) := \sum_{j=mp^s}^{(m+1)p^s-1} \big(B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j) - B_{\mathbf{N}}(j)B_{\mathbf{N}}(a+(K-j)p)\big).$$

In this expression for $S(a,K,s,p,m)$, it is assumed that $B_{\mathbf{N}}(n) = 0$ for negative integers n.

It would suffice to prove that

$$Y_{m,s} \in p\Xi_N N!^k\mathbb{Z}_p \tag{3.15}$$

because (3.13) and (3.14) would then imply that $C(a+Kp) \in p\Xi_N N!^k\mathbb{Z}_p$, as desired.

Equation (3.15) can be proved in the following manner. The expression for $S(a,K,s,p,m)$ is of a form which can be treated by Dwork's formal congruence theorem (see [3, Theorem 1.1]). This enables us to get the following lemma, whose proof is given in [6] and will not be repeated here.

Lemma 4. *For all primes p and non-negative integers a,m,s,K with $0 \le a < p$, we have*

$$S(a,K,s,p,m) \in p^{s+1}B_{\mathbf{N}}(m)\mathbb{Z}_p. \tag{3.16}$$

Furthermore, in Section 5 we shall prove the following lemma.

Lemma 5. *For all primes p, non-negative integers m, and positive integers N, we have*

$$B_{\mathbf{N}}(m)\big(H_{Nmp^s} - H_{N\lfloor m/p\rfloor p^{s+1}}\big) \in p^{-s}\Xi_N N!^k\mathbb{Z}_p\ , \tag{3.17}$$

where again $\mathbf{N} = (N,N,\ldots,N)$, with k occurrences of N.

It is clear that (3.16) and (3.17) imply (3.15). This completes the outline of the proof of Theorem 1.

4. Proof of Lemma 3

The assertion is trivially true if $\lfloor Na/p\rfloor = 0$, that is, if $0 \le a < p/N$. We may hence assume that $p/N \le a < p$ from now on.

4.1. **First part: a weak version of Lemma 3.** From [6, Eq. (9.1)] with $\mathbf{N} = (N, N, \dots, N)$, we know that

$$B_{\mathbf{N}}(a+pj)\left(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\right) \in p\mathbb{Z}_p\ . \tag{4.1}$$

(The reader should note the absence of the term $\Xi_N N!^k$ in comparison with (3.11) or (3.12).) In [6, Sec. 9.1], we have also shown that for any

$$D \le 1 + \max_{1\le\varepsilon\le\lfloor Na/p\rfloor} v_p(Nj+\varepsilon)$$

we have

$$\sum_{\ell=2}^{D}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right) \ge D-1, \tag{4.2}$$

a fact that we shall use later on.

We now embark on the proof of the lemma in its full form.

4.2. **Second part: the case $j = 0$.** In this case, we want to prove that, for $a \ge 2$, we have

$$B_{\mathbf{N}}(a)H_{\lfloor Na/p\rfloor} \in p\Xi_N N!^k\mathbb{Z}_p\ , \tag{4.3}$$

respectively that

$$B_{\mathbf{N}}(1)B_{\mathbf{N}}(K)H_{\lfloor N/p\rfloor} \in p\Xi_N N!^k\mathbb{Z}_p\ . \tag{4.4}$$

The reader should keep in mind that we still assume that $p/N \le a < p$, so that, in particular, $a > 0$.

If $p > N$, then our claim would follow from $B_{\mathbf{N}}(a)H_{\lfloor Na/p\rfloor} \in p\mathbb{Z}_p$, which is indeed true because of (4.1) with $j = 0$.

Now let $p \le N$. If $a = 1$, then use of (3.4) in the other direction yields

$$B_{\mathbf{N}}(1)B_{\mathbf{N}}(K)H_{\lfloor N/p\rfloor} \equiv pN!^kB_{\mathbf{N}}(K)H_N \mod pN!^kB_{\mathbf{N}}(K)\mathbb{Z}_p\ .$$

By the assumption that $K \ge 1$ and Lemma 2, this congruence implies

$$B_{\mathbf{N}}(1)B_{\mathbf{N}}(K)H_{\lfloor N/p\rfloor} \equiv pN!^kB_{\mathbf{N}}(K)H_N \mod pN!^{k+1}\mathbb{Z}_p\ . \tag{4.5}$$

It is easy to see that

$$v_p(N!) \ge \begin{cases} 1, & \text{if } \frac{N}{3} < p \le N,\\ 3, & \text{if } p \le \frac{N}{3}.\end{cases}$$

We have $v_p(H_N) = -1$ as long as $\frac{N}{3} < p \le N$ (for $p = 2$ this follows from Lemma 8), except if $p = 3$ and $N = 7$, in which case we have $v_3(H_7) = v_p(\frac{363}{140}) = 1$. From the definition of Ξ_N in (1.8) we see that (4.5) implies

$$B_{\mathbf{N}}(1)B_{\mathbf{N}}(K)H_{\lfloor N/p\rfloor} \equiv pN!^kB_{\mathbf{N}}(K)H_N \mod p\Xi_N N!^k\mathbb{Z}_p\ .$$

Again by the definition of Ξ_N, we have $H_N \in \Xi_N\mathbb{Z}_p$, and, thus,

$$B_{\mathbf{N}}(1)B_{\mathbf{N}}(K)H_{\lfloor N/p\rfloor} \in p\Xi_N N!^k\mathbb{Z}_p\ ,$$

which is exactly (4.4).

From now on, we assume $a \geq 2$. We bound the p-adic valuation of the expression on the left in (4.3) from below:

$$\begin{aligned} v_p\big(B_{\mathbf{N}}(a)H_{\lfloor Na/p\rfloor}\big) &= k\sum_{\ell=1}^{\infty}\left(\left\lfloor\frac{Na}{p^\ell}\right\rfloor - N\left\lfloor\frac{a}{p^\ell}\right\rfloor\right) + v_p(H_{\lfloor Na/p\rfloor}) \\ &\geq k\sum_{\ell=1}^{\infty}\left\lfloor\frac{Na}{p^\ell}\right\rfloor - \big\lfloor\log_p Na/p\big\rfloor \\ &\geq k\sum_{\ell=1}^{\infty}\left\lfloor\frac{2N}{p^\ell}\right\rfloor - \big\lfloor\log_p N\big\rfloor \\ &\geq \left\lfloor\frac{N}{p}\right\rfloor + k\sum_{\ell=1}^{\infty}\left\lfloor\frac{N}{p^\ell}\right\rfloor - \big\lfloor\log_p N\big\rfloor \qquad (4.6) \\ &\geq \lfloor N/p\rfloor + k\cdot v_p(N!) - \big\lfloor\log_p N\big\rfloor\,. \qquad (4.7) \end{aligned}$$

If $p = 2$, then we can continue the estimation (4.7) as

$$v_2\big(B_{\mathbf{N}}(a)H_{\lfloor Na/2\rfloor}\big) \geq 1 + k\cdot v_2(N!) - \lfloor\log_2 N\rfloor = v_2\big(2\Xi_N N!^k\big), \qquad (4.8)$$

where we used Lemma 8 and the definition of Ξ_N to obtain the equality. (In fact, at this point it was not necessary to consider the case $p = 2$ because $a < p$ and because we assumed $a \geq 2$. However, we shall re-use the present estimations later in the third part of the current proof, in a context where $a = 1$ is allowed.)

From now on let $p \geq 3$. We use the fact that

$$\lfloor x\rfloor \geq \big\lfloor\log_p x\big\rfloor + 2 \qquad (4.9)$$

for all $x \geq 2$ and primes $p \geq 3$. If we suppose that $v_p(H_N) \leq 0$, or that $p = 3$ and $N = 7$, then $v_p(\Xi_N) \leq 0$. (The reader should recall that Ξ_N has a separate definition for $N = 7$, Remarks 1(a) in the Introduction.) Hence, in the case that $N \geq 2p$, the estimation (4.7) can be continued as

$$v_p\big(B_{\mathbf{N}}(a)H_{\lfloor Na/p\rfloor}\big) \geq 2 + \big\lfloor\log_p N/p\big\rfloor + k\cdot v_p(N!) - \big\lfloor\log_p N\big\rfloor \geq v_p(pN!^k) \geq v_p(p\Xi_N N!^k),$$

implying (4.3) in this case. If $p \leq N < 2p$, in which case $v_p(H_N) = -1$, the estimation (4.7) can be continued as

$$v_p\big(B_{\mathbf{N}}(a)H_{\lfloor Na/p\rfloor}\big) \geq 1 + k\cdot v_p(N!) - 1 = v_p(p\Xi_N N!^k),$$

implying (4.3) in this case also.

Now let $v_p(H_N) > 0$, but not $p = 3$ and $N = 7$. (The latter case was already discussed.) We claim that under this condition we have

$$\lfloor N/p\rfloor - \big\lfloor\log_p N/p\big\rfloor \geq \begin{cases} 4, & \text{if } p = 5, \\ 5, & \text{if } p > 5, \text{ or if } p = 3 \text{ and } N = 22, \\ 6, & \text{if } v_p(H_N) > 2. \end{cases} \qquad (4.10)$$

Indeed, if $p = 3$, then Lemma 9 tells us that only the case $N = 22$ needs to be considered, in which case the above inequality is trivially true. If $p = 5$ then Lemma 10 says that N must be at least 20 because otherwise $v_5(H_N) \le 0$. The inequality (4.10) follows immediately. On the other hand, if $p = 11$ then it can be checked (by our table mentioned in Remarks 1(c) in the Introduction, for example) that N must be at least 77 because otherwise $v_{11}(H_N) \le 0$. If p is different from 3, 5, and 11, then, according to Lemma 11(3), we must have $N/p \ge 5$. In both of the latter cases, the inequality (4.10) follows immediately. Finally, if $v_p(H_N) > 2$, then, according to Lemmas 10 and 11(4), we must have $p > 5$ and $N/p \ge 6$, thus implying (4.10).

The effect of the above improvement over (4.9) is that the estimation (4.7) may now be continued to result in

$$B_{\mathbf{N}}(a)H_{\lfloor Na/p\rfloor} \in \begin{cases} p^3 N!^k \mathbb{Z}_p & \text{if } p = 5, \\ p^5 N!^k \mathbb{Z}_p & \text{if } v_p(H_N) > 2, \\ p^4 N!^k \mathbb{Z}_p & \text{otherwise.} \end{cases}$$

This implies (4.3) since $v_5(H_N) \le 1$ for $N \ge 5$ according to Lemma 10.

Thus, (3.11) with $j = 0$ is established.

4.3. **Third part: the case $j > 0$.** Now let $j > 0$. If $p > N$, then (3.11) reduces to

$$B_{\mathbf{N}}(a + pj)\left(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\right) \in p\mathbb{Z}_p \ ,$$

which is again true because of (4.1).

Now let $p \le N$. The reader should keep in mind that we still assume that $p/N \le a < p$, so that, in particular, $a > 0$. In a similar way as we did for the expression on the left in (4.1), we bound the p-adic valuation of the expression on the left in (3.11) from below. For the sake of convenience, we write T_1 for $\max_{1\le\varepsilon\le\lfloor Na/p\rfloor} v_p(Nj+\varepsilon)$ and T_2 for $\bigl\lfloor \log_p(a + pj)\bigr\rfloor$. Since it is somewhat hidden where our assumption $j > 0$ enters the subsequent considerations, we point out to the reader that $j > 0$ implies that $T_2 \ge 1$; without this property the split of the sum over ℓ into subsums in the chain of inequalities below would be impossible. So, using the above notation, we have (the detailed explanations for the various steps are given immediately after the following chain of estimations)

$$\begin{aligned} &v_p\Bigl(B_{\mathbf{N}}(a + pj)\left(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\right)\Bigr) \\ &\quad = k\sum_{\ell=1}^{\infty}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right) + v_p\bigl(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\bigr) \\ &\quad = \left\lfloor\frac{N(a+pj)}{p}\right\rfloor - N\left\lfloor\frac{a+pj}{p}\right\rfloor + \sum_{\ell=2}^{\min\{1+T_1,T_2\}}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right) \\ &\qquad + \sum_{\ell=\min\{1+T_1,T_2\}+1}^{\infty}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right) \end{aligned}$$

$$+ (k-1)\sum_{\ell=1}^{\infty}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right) + v_p\big(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\big)$$

$$\geq \left\lfloor\frac{Na}{p}\right\rfloor - N\left\lfloor\frac{a}{p}\right\rfloor + \min\{1+T_1, T_2\} - 1 + k\sum_{\ell=T_2+1}^{\infty}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right)$$
$$+ v_p\big(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\big) \tag{4.11}$$

$$\geq \left\lfloor\frac{Na}{p}\right\rfloor + T_1 + v_p\big(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\big) + \min\{0, T_2 - T_1 - 1\}$$
$$+ k\sum_{\ell=\lfloor\log_p(a+pj)\rfloor+1}^{\infty}\left(\left\lfloor\frac{N(a+pj)}{p^\ell}\right\rfloor - N\left\lfloor\frac{a+pj}{p^\ell}\right\rfloor\right) \tag{4.12}$$

$$\geq \lfloor N/p\rfloor + \min\{0, T_2 - T_1 - 1\} + k\sum_{\ell=1}^{\infty}\left\lfloor\frac{N}{p^\ell}\cdot\frac{a+pj}{p^{\lfloor\log_p(a+pj)\rfloor}}\right\rfloor \tag{4.13}$$

$$\geq \lfloor N/p\rfloor + \big\lfloor\log_p(a+pj)\big\rfloor - \big\lfloor\log_p\big(Nj + \lfloor Na/p\rfloor\big)\big\rfloor - 1 + k\sum_{\ell=1}^{\infty}\left\lfloor\frac{N}{p^\ell}\cdot\frac{a+pj}{p^{\lfloor\log_p(a+pj)\rfloor}}\right\rfloor \tag{4.14}$$

$$\geq \lfloor N/p\rfloor + \big\lfloor\log_p j\big\rfloor - \big\lfloor\log_p\big(Nj + \lfloor Na/p\rfloor\big)\big\rfloor + k\sum_{\ell=1}^{\infty}\left\lfloor\frac{N}{p^\ell}\right\rfloor \tag{4.15}$$

$$\geq \lfloor N/p\rfloor + \big\lfloor\log_p j\big\rfloor - \big\lfloor\log_p N\big\rfloor - \left\lfloor\log_p\left(j + \frac{1}{N}\lfloor Na/p\rfloor\right)\right\rfloor - 1 + k\cdot v_p(N!) \tag{4.16}$$

$$\geq \lfloor N/p\rfloor - \big\lfloor\log_p N\big\rfloor - 1 + v_p(N!^k). \tag{4.17}$$

Here, we used (4.2) in order to get (4.11). To get (4.13), we used the inequalities

$$\left\lfloor\frac{Na}{p}\right\rfloor \geq \left\lfloor\frac{N}{p}\right\rfloor \tag{4.18}$$

and

$$T_1 + v_p\big(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}\big) \geq 0. \tag{4.19}$$

To get (4.14), we used that

$$T_2 - T_1 - 1 \geq \big\lfloor\log_p(a+pj)\big\rfloor - \big\lfloor\log_p\big(Nj + \lfloor Na/p\rfloor\big)\big\rfloor - 1$$

and

$$\big\lfloor\log_p(a+pj)\big\rfloor - \big\lfloor\log_p\big(Nj + \lfloor Na/p\rfloor\big)\big\rfloor - 1 = \big\lfloor\log_p j\big\rfloor - \big\lfloor\log_p\big(Nj + \lfloor Na/p\rfloor\big)\big\rfloor \leq 0,$$

so that

$$\min\{0, T_2 - T_1 - 1\} \geq \big\lfloor\log_p(a+pj)\big\rfloor - \big\lfloor\log_p\big(Nj + \lfloor Na/p\rfloor\big)\big\rfloor - 1. \tag{4.20}$$

Next, to get (4.15), we used

$$\left\lfloor \frac{N}{p^\ell} \cdot \frac{a+pj}{p^{\lfloor \log_p(a+pj) \rfloor}} \right\rfloor \geq \left\lfloor \frac{N}{p^\ell} \right\rfloor . \tag{4.21}$$

To get (4.16), we used

$$\big\lfloor \log_p \big(Nj + \lfloor Na/p \rfloor \big) \big\rfloor \leq \big\lfloor \log_p N \big\rfloor + \left\lfloor \log_p \left(j + \frac{1}{N} \lfloor Na/p \rfloor \right) \right\rfloor + 1. \tag{4.22}$$

Finally, we used $\frac{1}{N} \lfloor Na/p \rfloor < 1$ in order to get (4.17).

If we now repeat the arguments after (4.7), then we see that the estimation (4.17) implies

$$v_p\Big(B_{\mathbf{N}}(a+pj) \left(H_{Nj+\lfloor Na/p \rfloor} - H_{Nj} \right) \Big) \geq v_p\big(\Xi_N N!^k \big). \tag{4.23}$$

This almost proves (3.11), our lower bound on the p-adic valuation of the number in (3.11) is just by 1 too low.

In order to establish that (3.11) is indeed correct, we assume by contradiction that all the inequalities in the estimations leading to (4.17) and finally to (4.23) are in fact equalities. In particular, the estimations in (4.18) hold with equality only if $a = 1$, which we shall assume henceforth.

If we examine the arguments after (4.7) that led us from (4.17) to (4.23), then we see that they prove in fact

$$v_p\Big(B_{\mathbf{N}}(a+pj) \left(H_{Nj+\lfloor Na/p \rfloor} - H_{Nj} \right) \Big) \geq 1 + v_p\big(\Xi_N N!^k \big) \tag{4.24}$$

except if $v_p(H_N) \leq 0$ and:

Case 1: $p = 2$ and $\lfloor N/2 \rfloor = 1$;
Case 2: $p \geq 3$ and $p \leq N < 2p$;
Case 3: $p = 3$ and $\lfloor N/3 \rfloor = 2$;

Indeed, if $N \geq 3p$, this is obvious while, if $2p \leq N < 3p$, we have $v_p(H_N) = -1$, except if $p = 3$ and $N = 7$, a case included in Case 3. Therefore, if we exclude the case where $p = 3$ and $N = 7$, then, if $2p \leq N < 3p$, we can continue (4.17) as

$$\begin{aligned} &v_p\Big(B_{\mathbf{N}}(a+pj) \left(H_{Nj+\lfloor Na/p \rfloor} - H_{Nj} \right) \Big) \\ &\qquad \geq 2 + \big\lfloor \log_p N/p \big\rfloor - \big\lfloor \log_p N \big\rfloor - 1 + v_p(N!^k) \geq v_p(N!^k) \geq v_p(p \Xi_N N!^k), \end{aligned}$$

where we used (4.9) in the first step. This is exactly (4.24).

We now show that (4.24) holds as well in Cases 1–3, thus completing the proof of (3.11).

Case 1. Let first $p = 2$ and $N = 2$. We then have

$$\begin{aligned} \min\{0, T_2 - T_1 - 1\} &= \min\{0, \lfloor \log_2(2j+1) \rfloor - v_2(2j+1) - 1\} \\ &= \min\{0, \lfloor \log_2(2j+1) \rfloor - 1\} = 0 > -1, \end{aligned}$$

in contradiction to having equality in (4.20).

On the other hand, if $p = 2$ and $N = 3$, we have

$$H_{Nj+\lfloor Na/p\rfloor} - H_{Nj} = H_{3j+1} - H_{3j} = \frac{1}{3j+1}.$$

If there holds equality in (4.20), then $Nj + \lfloor Na/p\rfloor = 3j + 1$ must be a power of 2, say $3j + 1 = 2^e$ or, equivalently, $j = (2^e - 1)/3$. It follows that

$$\left\lfloor \frac{N}{p} \cdot \frac{a+pj}{p^{\lfloor \log_p(a+pj)\rfloor}} \right\rfloor = \left\lfloor \frac{3}{2} \cdot \frac{1+2j}{2^{\lfloor \log_2(1+2j)\rfloor}} \right\rfloor = \left\lfloor \frac{3}{2} \cdot \frac{2^{e+1}+1}{3 \cdot 2^{e-1}} \right\rfloor = 2 > 1 = \left\lfloor \frac{3}{2} \right\rfloor = \left\lfloor \frac{N}{p} \right\rfloor,$$

in contradiction to having equality in (4.21) with $\ell = 1$.

Case 2. Our assumptions $p \geq 3$ and $p \leq N < 2p$ imply

$$H_{Nj+\lfloor Na/p\rfloor} - H_{Nj} = H_{Nj+1} - H_{Nj} = \frac{1}{Nj+1}.$$

Arguing as in the previous case, in order to have equality in (4.20), we must have $Nj+1 = f \cdot p^e$ for some positive integers e and f with $0 < f < p$. Thus, $j = (f \cdot p^e - 1)/N$ and $p < N$. (If $p = N$ then j would be non-integral.) It follows that

$$\left\lfloor \frac{N}{p} \cdot \frac{a+pj}{p^{\lfloor \log_p(a+pj)\rfloor}} \right\rfloor = \left\lfloor \frac{N}{p} \cdot \frac{f \cdot p^{e+1} + N - p}{N \cdot p^{\lfloor \log_p((f \cdot p^{e+1}+N-p)/N)\rfloor}} \right\rfloor. \tag{4.25}$$

If $f = 1$, then we obtain from (4.25) that

$$\left\lfloor \frac{N}{p} \cdot \frac{a+pj}{p^{\lfloor \log_p(a+pj)\rfloor}} \right\rfloor = \left\lfloor \frac{N}{p} \cdot \frac{p^{e+1}+N-p}{N \cdot p^{e-1}} \right\rfloor \geq \left\lfloor \frac{p^{e+1}+N-p}{p^e} \right\rfloor > 1 = \left\lfloor \frac{N}{p} \right\rfloor,$$

in contradiction with having equality in (4.21) with $\ell = 1$.

On the other hand, if $f \geq 2$, then we obtain from (4.25) that

$$\left\lfloor \frac{N}{p} \cdot \frac{a+pj}{p^{\lfloor \log_p(a+pj)\rfloor}} \right\rfloor \geq \left\lfloor \frac{f \cdot p^{e+1}+N-p}{p^{e+1}} \right\rfloor \geq f > 1 = \left\lfloor \frac{N}{p} \right\rfloor,$$

again in contradiction with having equality in (4.21) with $\ell = 1$.

Case 3. Our assumptions $p = 3$ and $\lfloor N/3\rfloor = 2$ imply

$$H_{Nj+\lfloor Na/p\rfloor} - H_{Nj} = H_{Nj+2} - H_{Nj} = \frac{1}{Nj+1} + \frac{1}{Nj+2}.$$

Similar to the previous cases, in order to have equality in (4.20), we must have $Nj+\varepsilon = f \cdot 3^e$ for some positive integers ε, e, f with $0 < \varepsilon, f < 3$. The arguments from Case 2 can now be repeated almost verbatim. We leave the details to the reader.

This completes the proof of the lemma.

5. Proof of Lemma 5

The claim is trivially true if p divides m. We may therefore assume that p does not divide m for the rest of the proof. Let us write $m = a + pj$, with $0 < a < p$. Then comparison with (3.11) shows that we are in a very similar situation here. Indeed, we may derive (3.17) from Lemma 3. In order to see this, we observe that

$$\begin{aligned} H_{Nmp^s} - H_{N\lfloor m/p\rfloor p^{s+1}} &= \sum_{\varepsilon=1}^{Nap^s} \frac{1}{Njp^{s+1}+\varepsilon} \\ &= \sum_{\varepsilon=1}^{\lfloor Na/p\rfloor} \frac{1}{Njp^{s+1}+\varepsilon p^{s+1}} + \sum_{\substack{\varepsilon=1\\ p^{s+1}\nmid\varepsilon}}^{Nap^s} \frac{1}{Njp^{s+1}+\varepsilon} \\ &= \frac{1}{p^{s+1}}(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}) + \sum_{\substack{\varepsilon=1\\ p^{s+1}\nmid\varepsilon}}^{Nap^s} \frac{1}{Njp^{s+1}+\varepsilon}. \end{aligned}$$

Because of $v_p(x+y) \geq \min\{v_p(x), v_p(y)\}$, this implies

$$v_p(H_{Nmp^s} - H_{N\lfloor m/p\rfloor p^{s+1}}) \geq \min\{-1-s+v_p(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj}), -s\}.$$

It follows that

$$\begin{aligned} &v_p\Big(B_{\mathbf{N}}(m)\big(H_{Nmp^s} - H_{N\lfloor m/p\rfloor p^{s+1}}\big)\Big) \\ &\qquad \geq -1-s+\min\left\{v_p\big(B_{\mathbf{N}}(a+pj)(H_{Nj+\lfloor Na/p\rfloor} - H_{Nj})\big), 1+v_p\big(B_{\mathbf{N}}(a+pj)\big)\right\}. \quad (5.1) \end{aligned}$$

If $v_p(H_N) \leq 0$ or $p > N$, then use of Lemmas 2 and 3 (cf. (3.11)) implies (3.17) immediately since $v_p(\Xi_N) \leq 0$ in this case. On the other hand, if $v_p(H_N) > 0$ and $p \leq N$, a combination of Corollary 1 with Lemma 11(2) yields that $v_p\big(B_{\mathbf{N}}(a+pj)\big) \geq 3 + v_p(N!^k)$ as long as $p \neq 3$ and $m \geq 2$. If this is used in (5.1) together with Lemma 3, we obtain again (3.17) under the above restriction on p and m. If $p = 3$, $N = 7$, and $m \geq 2$, another use of Corollary 1 yields $v_p\big(B_{\mathbf{N}}(a+pj)\big) \geq 2 + v_p(N!^k)$, which, together with (5.1) and Lemma 3, implies (3.17) also in this case (recall Remarks 1(a)). If $p = 3$, $N \neq 7$, and $m \geq 2$, then either $N = 22$, in which case $v_p\big(B_{\mathbf{N}}(a+pj)\big) \geq 7 + v_p(N!^k)$ according to Corollary 1, or $v_p(H_N) \leq 0$ according to Lemma 9. In both cases, the claimed congruence (3.17) follows immediately if one combines this with (5.1).

In the remaining case $m = 1$, we compute directly:

$$B_{\mathbf{N}}(1)H_{Np^s} = N!^k\left(\frac{1}{p^s}H_N + \frac{1}{p^{s-1}}\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{Np}\frac{1}{\varepsilon} + \frac{1}{p^{s-2}}\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{Np^2}\frac{1}{\varepsilon} + \sum_{\substack{\varepsilon=1\\ p^{s-2}\nmid\varepsilon}}^{Np^s}\frac{1}{\varepsilon}\right).$$

The last sum over ε is clearly an element of $p^{-s+3}\mathbb{Z}_p$. Moreover, if $p \geq 5$, Lemma 12 implies that the next-to-last expression between parentheses is an element of $p^{-s+4}\mathbb{Z}_p \subset p^{-s+3}\mathbb{Z}_p$, and that the second expression between parentheses is an element of $p^{-s+3}\mathbb{Z}_p$. If $p = 3$, it

can be checked directly that the expression between parentheses is an element of $3^{-s+2}\mathbb{Z}_3$. Putting everything together, we conclude that

$$B_{\mathbf{N}}(1)H_{Np^s} \in p^{-s}N!^k\Xi_N\mathbb{Z}_p,$$

which finishes the proof of (3.17).

6. Auxiliary lemmas

In this section, we prove the auxiliary results necessary for the proof of Theorem 1, of which the outline was given in Section 3, with the proofs of Lemmas 3 and 5 given in Sections 4 and 5, respectively. These are, on the one hand, improvements of Lemma 2 (see Lemmas 6, 7 and Corollary 1), and, on the other hand, assertions addressing specific p-adic properties of harmonic numbers (see Lemmas 8–14 and Corollary 2). Some of the results of this section are also referred to in the next section.

We begin by two lemmas improving on Lemma 2.

Lemma 6. *For all positive integers N, a, and primes p with $2 \le a < p$, we have*

$$v_p(B_{\mathbf{N}}(a)) \ge \left\lfloor \frac{N}{p} \right\rfloor + v_p(N!^k),$$

where $\mathbf{N} = (N, N, \dots, N)$ (with k occurrences of N).

Proof. This was implicitly proved by the estimations leading to (4.6). □

Lemma 7. *For all positive integers N, a, j, and primes p with $1 \le a < p$, we have*

$$v_p(B_{\mathbf{N}}(a+jp)) \ge \left\lfloor \frac{N}{p} \right\rfloor + \min\{1+T_1, T_2\} - 1 + v_p(N!^k),$$

where $\mathbf{N} = (N, N, \dots, N)$ (with k occurrences of N), and where

$$T_1 = \max_{1\le\varepsilon\le\lfloor Na/p\rfloor} v_p(Nj+\varepsilon) \quad \text{and} \quad T_2 = \left\lfloor \log_p(a+pj) \right\rfloor.$$

Proof. This is seen by going through the estimations leading to (4.17), without employing (4.19) and (4.20). □

As a corollary to Lemmas 6 and 7, we obtain the following succinct p-adic estimation for $B_{\mathbf{N}}(m)$, which is needed in the proofs of (3.8), (3.9), and (3.17).

Corollary 1. *Let N and m be positive integers and p be a prime such that m is at least 2 and not divisible by p. Then we have*

$$v_p(B_{\mathbf{N}}(m)) \ge \left\lfloor \frac{N}{p} \right\rfloor + v_p(N!^k),$$

where $\mathbf{N} = (N, N, \dots, N)$ (with k occurrences of N).

The next three lemmas of this section provide elementary information on the p-adic valuation of harmonic numbers for $p = 2, 3, 5$ which is needed in the proof of Lemma 11 and is also referred to frequently at other places. (For example, Lemma 8 was used in the proof of (4.8).) The proofs are not difficult (cf. [2]) and are therefore omitted.

Lemma 8. *For all positive integers N, we have $v_2(H_N) = -\lfloor \log_2 N \rfloor$.*

Lemma 9. *We have $v_3(H_2) = v_3(H_7) = v_3(H_{22}) = 1$. For positive integers $N \notin \{2, 7, 22\}$, we have $v_3(H_N) \leq 0$.*

Lemma 10. *We have $v_5(H_4) = 2$ and $v_5(H_{20}) = v_5(H_{24}) = 1$. For positive integers $N \notin \{4, 20, 24\}$, we have $v_5(H_N) \leq 0$.*

Next, we record some properties of integers N and primes p for which $v_p(H_N) > 0$. These are needed throughout Section 3.

Lemma 11. *Let p be a prime, and let N be an integer with $p \leq N$. Then the following assertions hold true:*

(1) *If $v_p(H_N) > 0$ then $N \geq 2p$.*
(2) *If $v_p(H_N) > 0$ and $p \neq 3$ then $N \geq 3p$.*
(3) *If $v_p(H_N) > 0$ and $p \notin \{3, 5, 11\}$ then $N \geq 5p$.*
(4) *If $v_p(H_N) > 2$ then $N \geq 6p$.*

Proof. We have

$$H_N = \frac{1}{p} H_{\lfloor N/p \rfloor} + \sum_{\substack{\varepsilon=1 \\ p \nmid \varepsilon}}^{N} \frac{1}{\varepsilon}.$$

Since the sum over ε is in $\mathbb{Z}_p$, in order to have $v_p(H_N) > 0$ we must have $v_p(H_{\lfloor N/p \rfloor}) > 0$. Clearly, $v_p(H_1) = 0$ so that (1) follows. If $\lfloor N/p \rfloor < 3$ and $p \neq 3$, then $v_p(H_{\lfloor N/p \rfloor})$ cannot be positive since $H_1 = 1$ and $H_2 = \frac{3}{2}$. This implies (2). If $\lfloor N/p \rfloor < 5$ and $p \notin \{3, 5, 11\}$, then, again, $v_p(H_{\lfloor N/p \rfloor})$ cannot be positive since, as we already noted, $H_1 = 1$ and $H_2 = \frac{3}{2}$, and since $H_3 = \frac{11}{6}$ and $H_4 = \frac{25}{12}$. This yields (3).

To see (4), we observe that, owing to Lemmas 8–10, we may assume that $p \notin \{2, 3, 5\}$. Furthermore, if $p = 11$ then, according to our table referred to in Remarks 1(c) in the Introduction (see also [2]), we have $N \geq 848$. Similarly, if $p = 137$ then, according to our table, we have $N > 500000$. The claim can now be established in the style of the proofs of (1)–(3) upon observing that $H_5 = \frac{137}{60}$. □

We turn to a slight generalisation of Wolstenholme's theorem on harmonic numbers. (We refer the reader to [4, Chapter VII] for information on this theorem, which corresponds to the case $r = 1$ in the lemma below.)

Lemma 12. *For all primes $p \geq 5$ and positive integers r, we have*

$$v_p(H_{rp-1} - H_{rp-p}) \geq 2.$$

Proof. By simple rearrangement, we have

$$H_{rp-1} - H_{rp-p} = \sum_{\varepsilon=1}^{p-1} \frac{1}{rp-p+\varepsilon} = \sum_{\varepsilon=1}^{(p-1)/2} \left(\frac{1}{rp-p+\varepsilon} + \frac{1}{rp-\varepsilon} \right)$$
$$= p(2r-1) \sum_{\varepsilon=1}^{(p-1)/2} \frac{1}{(rp-p+\varepsilon)(rp-\varepsilon)}.$$

It therefore suffices to consider the last sum over ε in $\mathbb{Z}/p\mathbb{Z}$ (with $1/\alpha$ being interpreted as the multiplicative inverse of α in $\mathbb{Z}/p\mathbb{Z}$) and show that it is congruent to 0 modulo p. If we reduce this sum modulo p, then we are left with

$$- \sum_{\varepsilon=1}^{(p-1)/2} \frac{1}{\varepsilon^2},$$

which is, up to the sign, the sum of all quadratic residues in $\mathbb{Z}/p\mathbb{Z}$, that is, equivalently,

$$- \sum_{\varepsilon=1}^{(p-1)/2} \varepsilon^2 = \frac{p(p-1)(p+1)}{24}.$$

Clearly, this is divisible by p for all primes $p \geq 5$. □

As a corollary, we obtain strengthenings of (3.4) that we need in the proof of (3.8) and (3.7).

Corollary 2. *For all primes $p \geq 5$ and positive integers J divisible by p, we have*

$$pH_J \equiv H_{J/p} \mod p^3\mathbb{Z}_p.$$

Moreover, for all positive integers J divisible by 3, we have

$$3H_J \equiv H_{J/3} \mod 3^2\mathbb{Z}_3.$$

Proof. By simple rearrangement, we have

$$pH_J - H_{J/p} = p \sum_{r=1}^{J/p} (H_{rp-1} - H_{rp-p}).$$

Due to Lemma 12, the p-adic valuation of this expression is at least 3 if $p \geq 5$. If $p = 3$, it is easily seen directly that the 3-adic valuation of this expression is at least 2. □

Further strengthenings of (3.4), needed in the proof of (3.9), are given in the final two lemmas of this section.

Lemma 13. *For all primes $p \geq 5$ and positive integers J divisible by p^2, we have*

$$pH_J \equiv H_{J/p} \mod p^5\mathbb{Z}_p.$$

Proof. Again, by simple rearrangement, we have

$$\begin{aligned} pH_J - H_{J/p} &= p\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{J-1}\frac{1}{\varepsilon} = p\sum_{r=1}^{J/p^2}\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{p^2-1}\frac{1}{rp^2-p^2+\varepsilon} \\ &= p^3\sum_{r=1}^{J/p^2}(2r-1)\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{(p^2-1)/2}\frac{1}{(rp^2-p^2+\varepsilon)(rp^2-\varepsilon)}. \end{aligned}$$

It therefore suffices to consider the last sum over ε in $\mathbb{Z}/p^2\mathbb{Z}$ and show that it is congruent to 0 modulo p^2. If we reduce this sum modulo p^2, then we are left with

$$-\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{(p^2-1)/2}\frac{1}{\varepsilon^2},$$

which is, up to the sign, the sum of all quadratic residues in $\mathbb{Z}/p^2\mathbb{Z}$, that is, equivalently,

$$-\sum_{\substack{\varepsilon=1\\ p\nmid\varepsilon}}^{(p^2-1)/2}\varepsilon^2 = -\sum_{\varepsilon=1}^{(p^2-1)/2}\varepsilon^2 + \sum_{\varepsilon=1}^{(p-1)/2}(p\varepsilon)^2 = -\frac{p^2(p^2-1)(p^2+1)}{24} + p^2\frac{p(p-1)(p+1)}{24}.$$

Clearly, this is divisible by p^2 for all primes $p \geq 5$. □

Lemma 14. *For all primes $p \geq 5$ and positive integers N, we have*

$$pH_{pN} \equiv H_N \mod p^4\mathbb{Z}_p \tag{6.1}$$

if and only if p is a Wolstenholme prime or p divides N.

Proof. Using a rearrangement in the spirit of Lemma 12, we obtain

$$pH_{pN} - H_N = p\sum_{r=1}^{N}\sum_{\varepsilon=1}^{p-1}\frac{1}{rp-p+\varepsilon}.$$

We consider the sum over r in $\mathbb{Z}/p^3\mathbb{Z}$. This leads to

$$\begin{aligned} \sum_{r=1}^{N}\sum_{\varepsilon=1}^{p-1}(rp-p+\varepsilon)^{-1} &\equiv \sum_{r=1}^{N}\sum_{\varepsilon=1}^{p-1}\varepsilon^{-1}\big(1+p(r-1)\varepsilon^{-1}\big)^{-1} \\ &\equiv \sum_{r=1}^{N}\sum_{\varepsilon=1}^{p-1}\big(\varepsilon^{-1}-p(r-1)\varepsilon^{-2}+p^2(r-1)^2\varepsilon^{-3}\big) \\ &\equiv NH_{p-1} - p\binom{N}{2}H_{p-1}^{(2)} + p^2\frac{N(N-1)(2N-1)}{6}H_{p-1}^{(3)} \mod \mathbb{Z}/p^3\mathbb{Z}, \end{aligned} \tag{6.2}$$

where $H_m^{(\alpha)}$ denotes the higher harmonic number defined by $H_m^{(\alpha)} = \sum_{n=1}^{m} \frac{1}{n^\alpha}$. By a rearrangement analogous to the one in the proof of Lemma 12, one sees that $v_p(H_{p-1}^{(3)}) \ge 1$, whence we may disregard the last term in the last line of (6.2). As it turns out, H_{p-1} and $H_{p-1}^{(2)}$ are directly related modulo $\mathbb{Z}/p^3\mathbb{Z}$. Namely, we have

$$\begin{aligned} H_{p-1} &\equiv \sum_{\varepsilon=1}^{p-1} \varepsilon^{-1} \equiv 2^{-1} \sum_{\varepsilon=1}^{p-1} \bigl(\varepsilon^{-1} + (p-\varepsilon)^{-1}\bigr) \equiv p2^{-1} \sum_{\varepsilon=1}^{p-1} \bigl(\varepsilon(p-\varepsilon)\bigr)^{-1} \\ &\equiv -p2^{-1} \sum_{\varepsilon=1}^{p-1} \varepsilon^{-2}\bigl(1 + p\varepsilon^{-1}\bigr) \equiv -p2^{-1} \sum_{\varepsilon=1}^{p-1} \bigl(\varepsilon^{-2} + p\varepsilon^{-3}\bigr) \\ &\equiv -p2^{-1} H_{p-1}^{(2)} - p^2 2^{-1} H_{p-1}^{(3)} \mod \mathbb{Z}/p^3\mathbb{Z}. \end{aligned} \tag{6.3}$$

Again, we may disregard the last term. If we substitute this congruence in (6.2), then we obtain

$$\sum_{r=1}^{N} \sum_{\varepsilon=1}^{p-1} (rp - p + \varepsilon)^{-1} \equiv N^2 H_{p-1} \mod \mathbb{Z}/p^3\mathbb{Z}.$$

Hence, the sum is congruent to zero modulo $\mathbb{Z}/p^3\mathbb{Z}$ if and only if N is divisible by p (recall Lemma 12 with $r = 1$) or $v_p(H_{p-1}) \ge 3$, that is, if p is a Wolstenholme prime. This establishes the assertion of the lemma. □

7. The sequence $(t_N)_{N\ge1}$

The purpose of this section is to report on the evidence for our conjecture that the largest integer t_N such that $q_{N,(N)}(z)^{1/t_N} \in \mathbb{Z}[[z]]$ is given by $t_N = \Xi_N N!$, that is, that Theorem 1 with $k = 1$ is optimal. We assume $k = 1$ throughout this section.

We first prove that there cannot be any prime number p larger than N which divides t_N.

Proposition 1. *Let p be a prime number and N a positive integer with $p > N$. Then there exists a positive integer $a < p$ such that*

$$v_p\bigl(B_N(a)H_{Na}\bigr) = 0, \tag{7.1}$$

where $B_N(m) = \frac{(Nm)!}{m!^N}$. In particular, p does not divide t_N.

Proof. If $N = 1$, we choose $a = 1$ to obtain $B_1(1)H_1 = 1$. On the other hand, if $N > 1$, we choose a to be the least integer such that $aN \ge p$. Since then $a < p$, we have

$$v_p\bigl(B_N(a)H_{Na}\bigr) = \sum_{\ell=1}^{\infty} \left(\left\lfloor \frac{Na}{p^\ell} \right\rfloor - N \left\lfloor \frac{a}{p^\ell} \right\rfloor \right) + v_p(H_{Na}) = 1 - 1 = 0.$$

To see that p cannot divide t_N, we observe that we have $C(a) = -pB_N(a)H_{Na}$ (for the sum $C(\cdot)$ defined in (3.2) in the case $k = 1$) because $a < p$. Hence, the assertion (7.1)

can be reformulated as $v_p\big(C(a)\big) = 1$. Since, because of $p > N$, we have $v_p(N!) = 0$ and $v_p(\Xi_N) = 0$, it follows that

$$C(a) \notin p^2 \Xi_N N! \, \mathbb{Z}_p.$$

This means that one cannot increase the exponent of p in (3.3) (with $k = 1$) in our special case, and thus p cannot divide t_N. □

So, if we hope to improve Theorem 1 with $k = 1$, then it must be by increasing exponents of prime numbers $p \le N$ in (1.8). It can be checked directly that the exponent of 3 cannot be increased if $N = 7$. (The reader should recall Remarks 1(a) in the Introduction.) According to Remarks 1(b), an improvement is therefore only possible if $v_p(H_N) > 2$ for some $p \le N$. Lemmas 8–10 in Section 6 tell that this does not happen with $p = 2, 3, 5$, so that the exponents of $2, 3, 5$ cannot be improved. (The same conclusion can also be drawn from [2] for many other prime numbers, but so far not for 83, for example.)

We already discussed in Remarks 1(c) whether there are any primes p and integers N such that $p \le N$ and $v_p(H_N) \ge 3$. We recall that, so far, only five examples are known, four of them involving $p = 11$.

The final result of this section shows that, if $v_p(H_N) = 3$, the exponent of p in the definition of Ξ_N in (1.8) cannot be increased so that Theorem 1 would still hold. (The reader should recall Remarks 1(b).)

Proposition 2. *Let p be a prime number and N a positive integer with $p \le N$ and $v_p(H_N) = 3$. If p is not a Wolstenholme prime and p does not divide N, then $q_{N,(N)}(z)^{1/p\Xi_N N!} \notin \mathbb{Z}[[z]]$.*

Proof. We assume that p is not a Wolstenholme prime and that p does not divide N. In particular, this implies $\xi(p, N) = 0$ and thus also $v_p(\Xi_N) = 2$. By Lemmas 8–10, we can furthermore assume that $p \ge 7$.

Going back to the outline of the proof of Theorem 1 in Section 3, we claim that

$$C(p) = \big(B_N(1)H_N - B_N(p)pH_{Np}\big) \notin p^4 N! \, \mathbb{Z}_p, \tag{7.2}$$

where $B_N(m) = \frac{(Nm)!}{m!^N}$. (The claim here is the non-membership relation; the equality holds by the definition of $C(\cdot)$ in (3.3).) This would imply that $C(p) \notin p^2 \Xi_N N! \, \mathbb{Z}_p$, and thus, by Lemma 1 (recall the outline of the proof of Theorem 1 in Section 3), that $q_{N,(N)}(z)^{1/p\Xi_N N!} \notin \mathbb{Z}[[z]]$, as desired.

To establish (7.2), we consider

$$\begin{aligned} &H_N(B_N(1) - B_N(p)) \\ &\quad = H_N N! \bigg(1 - \frac{1}{(p-1)!^N}\big(1 \cdot 2 \cdots (p-1)\big)\big((p+1) \cdot (p+2) \cdots (2p-1)\big) \cdots \\ &\qquad\qquad \times \cdots \big((pN - p + 1) \cdot (pN - p + 2) \cdots (pN - 1)\big)\bigg). \end{aligned}$$

Using $v_p(H_N) = 3$ and Wilson's theorem, we deduce

$$H_N(B_N(1) - B_N(p)) \in p^4 N! \, \mathbb{Z}_p. \tag{7.3}$$

However, by Lemma 14 and the fact that $v_p\big(B_N(p)\big) = v_p\big(B_N(1)\big) = v_p(N!)$, we obtain

$$B_N(1)H_N - B_N(p)pH_{pN} \not\equiv H_N(B_N(1) - B_N(p)) \mod p^4 N!\,\mathbb{Z}_p.$$

Together with (7.3), this yields (7.2). □

To summarise the discussion of this section: if the conjecture in Remarks 1(c) that no prime p and integer N exists with $v_p(H_N) \geq 4$ should be true, then Theorem 1 with $k = 1$ is sharp, that is, the sequence $(t_N)_{N\geq 1}$ is given by $t_N = \Xi_N N!$.

8. Sketch of the proof of Theorem 2

In this section we discuss the proof of Theorem 2. Since it is completely analogous to the proof of Theorem 1 (see Section 3), we content ourselves with pointing out the differences. At the end of the section, we present analogues of Propositions 1 and 2, addressing the question of optimality of Theorem 2 with $k = 1$.

First of all, by (1.6), we have

$$\big(z^{-1}q_{\mathbf{N}}(z)\big)^{1/kN} = \exp(\widetilde{G}_{\mathbf{N}}(z)/F_{\mathbf{N}}(z)),$$

where $F_{\mathbf{N}}(z)$ is the series from the Introduction and

$$\widetilde{G}_{\mathbf{N}}(z) := \sum_{m=1}^{\infty} \frac{(Nm)!^k}{m!^{kN}}\big(H_{Nm} - H_m\big)\, z^m.$$

We must adapt the proof of Theorem 1, as outlined in Section 3. Writing as before $B_{\mathbf{N}}(m) = \frac{(Nm)!^k}{m!^{kN}}$, we must consider the sum

$$\begin{aligned}\widetilde{C}(a + Kp) &:= \sum_{j=0}^{K} B_{\mathbf{N}}(a + jp)B_{\mathbf{N}}(K - j)\big((H_{N(K-j)} - H_{K-j}) - p(H_{Na+Njp} - H_{a+jp})\big)\\ &= \sum_{j=0}^{K} B_{\mathbf{N}}(a + jp)B_{\mathbf{N}}(K - j)(H_{N(K-j)} - pH_{Na+Njp})\\ &\qquad - \sum_{j=0}^{K} B_{\mathbf{N}}(a + jp)B_{\mathbf{N}}(K - j)(H_{K-j} - pH_{a+jp}) \quad (8.1)\end{aligned}$$

and show that it is in $p\Omega_N N!^k\mathbb{Z}_p$ for all primes p, and for all non-negative integers K, a, and j with $0 \leq a < p$. The special cases $K = a = 0$, respectively $K = 0$ and $a = 1$, are equally simple to be handled directly here. We leave their verification to the reader and assume $a + Kp \geq 2$ from now on.

If we now go through the outline of the proof of Theorem 1 in Section 3, we see that the crucial steps are the congruence (3.6), and Lemmas 3–5. The analogue of (3.6) in our

context is the assertion that

$$\widetilde{C}(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)\big((H_{N(K-j)}-H_{K-j})-(H_{\lfloor Na/p\rfloor+Nj}-H_j)\big) \mod p\Omega_N N!^k\mathbb{Z}_p \quad (8.2)$$

for any non-negative integers m, K, a, and j with $0 \le a < p$. This follows easily from the congruence (3.4) and Lemma 2 as long as $v_p(H_N-1) \le 0$. If $v_p(H_N-1) > 0$, then we use Lemma 18 below to conclude that $N/p \ge 4$, which together with Lemma 2 and Corollary 1 in Section 6 yields

$$\widetilde{C}(a+Kp) \equiv \sum_{j=0}^{K} B_{\mathbf{N}}(a+jp)B_{\mathbf{N}}(K-j)\big((H_{N(K-j)}-H_{K-j})-(H_{\lfloor Na/p\rfloor+Nj}-H_j)\big) \mod p^5 N!^k\mathbb{Z}_p \quad (8.3)$$

as long as $a+Kp \ge 2$. This is more than we actually need to establish (8.2) in this case also.

The analogue of Lemma 3 in our context is

$$B_{\mathbf{N}}(a+pj)\left(H_{Nj+\lfloor Na/p\rfloor}-H_{Nj}\right) \in p\Omega_N N!^k\mathbb{Z}_p\ , \quad (8.4)$$

which we must prove for any prime p, and non-negative integers a, j with $0 \le a < p$. We may use Lemma 4 as it stands, but instead of Lemma 5 we should prove

$$B_{\mathbf{N}}(m)\big((H_{Nmp^s}-H_{mp^s})-(H_{N\lfloor m/p\rfloor p^{s+1}}-H_{\lfloor m/p\rfloor p^{s+1}})\big) \in p^{-s}\Omega_N N!^k\mathbb{Z}_p\ . \quad (8.5)$$

Both (8.4) and (8.5) can be proved in a very similar way as we proved Lemmas 3 and 5 in Sections 4 and 5, respectively. Indeed, as long as $v_p(H_N-1) \le 0$, there is nothing to prove, since in this case, by the definitions of Ω_N and Ξ_N, Lemma 3 implies (8.4) immediately, as well as Lemma 5 implies (8.5).

If, on the other hand, $v_p(H_N-1) > 0$, then we need substitutes for Lemmas 8–11, which were used to accomplish the proofs of Lemmas 3 and 5 in the case that $v_p(H_N) > 0$.

Lemma 15. *For all positive integers $N \ge 2$, we have $v_2(H_N-1) = -\lfloor \log_2 N \rfloor$.*

Lemma 16. *We have $v_3(H_{66}-1) = v_3(H_{68}-1) = 1$. For positive integers $N \notin \{1, 66, 68\}$, we have $v_3(H_N-1) \le 0$.*

Lemma 17. *We have $v_5(H_3-1) = v_5(H_{21}-1) = v_5(H_{23}-1) = 1$. For positive integers $N \notin \{1, 3, 21, 23\}$, we have $v_5(H_N-1) \le 0$.*

Lemma 18. *Let p be a prime, and let N be an integer with $N \ge 2$ and $p \le N$. Then the following assertions hold true:*

(1) *If $v_p(H_N-1) > 0$ then $N \ge 4p$.*
(2) *If $v_p(H_N-1) > 0$ and $p \ne 5$ then $N \ge 6p$.*

These results can be proved in exactly the same way as Lemmas 8–11, respectively. In comparison to Lemma 11, the statement of Lemma 18 is in fact simpler, so that complications that arose in Section 3 (such as (3.7), for example) do not arise here.

Second, we need a substitute for Lemma 14.

Lemma 19. *For all primes $p \geq 5$ and positive integers N, we have*

$$p(H_{pN} - H_p) \equiv H_N - 1 \mod p^4\mathbb{Z}_p \tag{8.6}$$

if and only if p is a Wolstenholme prime or $N \equiv \pm 1 \bmod p$.

Proof. From the proof of Lemma 14, we know that

$$pH_{pN} - H_N \equiv pN^2 H_{p-1} \mod \mathbb{Z}/p^4\mathbb{Z}.$$

As a consequence, we obtain

$$p(H_{pN} - H_p) - (H_N - 1) \equiv p(N^2 - 1)H_{p-1} \mod \mathbb{Z}/p^4\mathbb{Z}.$$

The assertion of the lemma follows now immediately. □

Finally, the computation for $m = 1$ in Section 5 must be replaced by

$$B_{\mathbf{N}}(1)(H_{Np^s} - H_{p^s}) = N!^k \left(\frac{1}{p^s}(H_N - 1) + \frac{1}{p^{s-1}} \sum_{\substack{\varepsilon=p+1 \\ p\nmid\varepsilon}}^{Np} \frac{1}{\varepsilon} + \frac{1}{p^{s-2}} \sum_{\substack{\varepsilon=p^2+1 \\ p\nmid\varepsilon}}^{Np^2} \frac{1}{\varepsilon} + \sum_{\substack{\varepsilon=p^s+1 \\ p^{s-2}\nmid\varepsilon}}^{Np^s} \frac{1}{\varepsilon} \right),$$

with the conclusion that

$$B_{\mathbf{N}}(1)(H_{Np^s} - H_{p^s}) \in p^{-s} N!^k \Omega_N \mathbb{Z}_p$$

being found in a completely analogous manner.

Altogether, this leads to a proof of Theorem 2.

9. The Dwork–Kontsevich sequence

In this section, we address the question of optimality of Theorem 2 when $k = 1$, that is, whether, given that $k = 1$, the largest integer u_N such that $\big(z^{-1}q_{(N)}(z)\big)^{\frac{1}{Nu_N}} \in \mathbb{Z}[[z]]$ is given by $\Omega_N N!$. Let us write $\widetilde{q}_{(N)}(z)$ for $\big(z^{-1}q_{(N)}(z)\big)^{1/N}$ with k *being fixed to* 1. The first proposition shows that there cannot be any prime number p larger than N which divides u_N. We omit the proof since it is completely analogous to the proof of Proposition 1 in Section 7.

Proposition 3. *Let p be a prime number and N a positive integer with $p > N \geq 2$. Then there exists a positive integer $a < p$ such that*

$$v_p\big(B_N(a)(H_{Na} - H_a)\big) = 0. \tag{9.1}$$

In particular, p does not divide u_N.

So, if we hope to improve Theorem 2 with $k = 1$, then it must be by increasing exponents of prime numbers $p \le N$ in (1.13). According to Remarks 2(b) in the Introduction, an improvement is therefore only possible if $v_p(H_N - 1) > 2$ for some $p \le N$.

The final result of this section shows that, if $v_p(H_N - 1) = 3$ (for which, however, so far no examples are known; see Remarks 2(c)), the exponent of p in the definition of Ω_N in (1.13) cannot be increased so that Theorem 2 with $k = 1$ would still hold.

Proposition 4. *Let p be a prime number and N a positive integer with $p \le N$ and $v_p(H_N) = 3$. If p is not a Wolstenholme prime and $N \not\equiv \pm 1 \bmod p$, then $\widetilde{q}_{(N)}(z)^{\frac{1}{p\Omega_N N!}} \notin \mathbb{Z}[[z]]$.*

Again, the proof is completely analogous to the proof of Proposition 2 in Section 7, which we therefore omit.

So, if the conjecture in Remarks 2(c) that no prime p and integer N exists with $v_p(H_N - 1) \ge 4$ should be true, then Theorem 2 with $k = 1$ is optimal, that is, the Dwork–Kontsevich sequence $(u_N)_{N\ge1}$ is given by $u_N = \Omega_N N!$.

Acknowledgements

The authors are extremely grateful to Alessio Corti and Catriona Maclean for illuminating discussions concerning the geometric side of our work, and to David Boyd for helpful information on the p-adic behaviour of the harmonic numbers H_N and for communicating to us the value (1.9) from his files from 1994. They also thank the referees for an extremely careful reading of the original manuscript.

Bibliography

[1] V. V. Batyrev and D. van Straten, *Generalized hypergeometric functions and rational curves on Calabi–Yau complete intersections in toric varieties*, Comm. Math. Phys. **168** (1995), 493–533.

[2] D. W. Boyd, *A p-adic study of the partial sums of the harmonic series*, Experimental Math. **3** (1994), 287–302.

[3] B. Dwork, *On p-adic differential equations IV: generalized hypergeometric functions as p-adic analytic functions in one variable*, Ann. Sci. É.N.S. (4) **6**, no. 3 (1973), 295–316.

[4] G. H. Hardy ans E. M. Wright, *An introduction to the theory of number theory*, 5th edition, Oxford University Press, 1979.

[5] N. Heninger, E. M. Rains and N. J. A. Sloane, *On the integrality of nth roots of generating functions*, J. Combin. Theory Ser. A **113** (2006), 1732–1745.

[6] C. Krattenthaler and T. Rivoal, *On the integrality of the Taylor coefficients of mirror maps*, Duke Math. J. (to appear).

[7] B. H. Lian and S.-T. Yau, *Arithmetic properties of mirror map and quantum coupling*, Comm. Math. Phys. **176**, no. 1 (1996), 163–191.

[8] B. H. Lian and S.-T. Yau, *Mirror maps, modular relations and hypergeometric series I*, appeared as *Integrality of certain exponential series*, in: Lectures in Algebra and Geometry, Proceedings of the International Conference on Algebra and Geometry, Taipei, 1995, M.-C. Kang (ed.), Int. Press, Cambridge, MA, 1998, pp. 215–227.

[9] B. H. Lian and S.-T. Yau, *The nth root of the mirror map*, in: Calabi–Yau varieties and mirror symmetry, Proceedings of the Workshop on Arithmetic, Geometry and Physics around Calabi–Yau Varieties

and Mirror Symmetry, Toronto, ON, 2001, N. Yui and J. D. Lewis (eds.), Fields Inst. Commun., **38**, Amer. Math. Soc., Providence, RI, 2003, pp. 195–199.
[10] The On-line encyclopedia of integers sequences, `http://www.research.att.com/~njas/sequences/`.
[11] M. Yoshida, *Fuchsian differential equations*, Aspects of Mathematics **11**, Vieweg, 1987.
[12] W. Zudilin, *Integrality of power expansions related to hypergeometric series*, Mathematical Notes **71**.5 (2002), 604–616.

C. Krattenthaler, Fakultät für Mathematik, Universität Wien, Nordbergstrasse 15, A-1090 Vienna, Austria. WWW: `http://www.mat.univie.ac.at/~kratt`.

T. Rivoal, Institut Fourier, CNRS UMR 5582, Université Grenoble 1, 100 rue des Maths, BP 74, 38402 Saint-Martin d'Hères cedex, France.
WWW: `http://www-fourier.ujf-grenoble.fr/~rivoal`.