J. COMBIN. MATH. COMBIN. COMPUT. 131 (2026) 495–514

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

The exponential rarity of vertical cancellable fractions

Mohsen Aliabadi*

ABSTRACT

A cancellable fraction is a displayed numerator–denominator pair in which deleting the same decimal digit from the numerator and denominator leaves the represented rational number unchanged. We study *vertical* cancellations, where the deleted digits occur in the same decimal position in the numerator and denominator. Fix a nonzero digit h. Let $C_h(n)$ denote the number of marked vertical cancellations of the digit h among numerator–denominator pairs (N, D), where N and D are both n-digit positive integers and $D < N$. Here *marked* means that the cancelled position is part of the data. We prove that

$$C_h(n) = O_h(n^2 10^{n-1}).$$

Since the number of such displayed pairs (N, D) is of order 10^{2n}, the proportion of vertically h-cancellable pairs is $O_h(n^2/10^n)$. Thus vertical h-cancellability is exponentially rare. We also give an explicit family showing that $C_h(n) \geq c_h 10^n$ for all sufficiently large n, and we include exact enumerations for small values of n. Finally, we formulate a uniqueness conjecture asserting that almost every marked vertical cancellation belongs to a pair that is cancellable in exactly one vertical way.

Keywords: cancellable fractions, illegal cancellation, decimal digits, enumerative combinatorics, lattice points, gcd sums

2020 Mathematics Subject Classification: Primary 11A63; Secondary 05A16, 11B99.

1. Introduction and conventions

Illegal cancellation is one of those mathematical jokes that refuses to stay only a joke. The familiar examples

$$\frac{16}{64} = \frac{1}{4} \qquad \text{and} \qquad \frac{49}{98} = \frac{4}{8},$$

* Corresponding author.

Received 04 Apr 2026; Revised 20 Jul 2026; Accepted 03 Aug 2026; Published Online 12 Aug 2026.

DOI: 10.61091/jcmcc131-24

© 2026 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

are obtained by deleting a common decimal digit from the numerator and denominator. Of course, such cancellation is not a valid algebraic operation in general. Nevertheless, some displayed fractions happen to survive it.

The classical problem goes back at least to Morley and Schub's problem in the *American Mathematical Monthly*. Schwartz later studied related examples under the title of illegal cancellation.

The existence of vertical cancellations is not new. In [1], explicit infinite families were constructed in which the numerator and denominator have equal decimal length and the cancelled digits occur in the same position. The geometry of cancellation positions, including the possible slopes of the cancellation line, was subsequently studied in [2]. The novel issue considered here is therefore not the existence of vertical cancellations, but their global enumeration among all displayed pairs of a prescribed length.

More precisely, I address the following question. Fix a nonzero digit h. Among all displayed pairs (N, D) for which N and D are n-digit positive integers and $D < N$, how many marked positions k allow the digit h to be deleted vertically from N and D without changing the represented rational number? Equivalently, what proportion of all such displayed pairs admit a vertical h-cancellation as $n \to \infty$?

This question differs in its order of quantification from the counting problem studied in [3]. There one fixes a reduced rational number L/M, together with the cancellable digit and its position, and counts cancellable representatives of that particular rational number. In particular, the formulas for $Z(L, M, n, H, k, k)$ and the associated limit results keep L, M, and the cancellation data fixed while the number of denominator digits tends to infinity. Here the represented rational number is not fixed: L/M varies over all reduced values arising from pairs of n-digit integers.

Although one may formally partition the present count according to the reduced value $L/M = N/D$, the fixed-rational results cannot simply be summed to obtain the present estimate. The family of reduced fractions being summed over grows with n, while the formulas in [3] depend on L and M through nonuniform divisibility, congruence, and floor conditions. Their limits are therefore pointwise in a fixed rational number and do not provide a uniform majorant over the n-dependent family needed here. In addition, their length parameter specifies the number of digits in the denominator, whereas the present problem requires both the numerator and denominator to have exactly n digits.

The approach taken here instead removes the represented rational number from the count. For each cancellation position, the cancellation identity is converted into a linear Diophantine equation in the digit blocks. Uniform lattice-point estimates and gcd-sum bounds then give

$$C_h(n) = O_h(n^2 10^{n-1}).$$

Since the total number of displayed pairs under consideration is of order 10^{2n}, this implies that vertical h-cancellability has proportion $O_h(n^2/10^n)$. Thus the principal contribution of the paper is a global enumeration and density theorem, rather than a new construction of vertical cancellations.

We work throughout in base 10. The units position is position 0, the tens position is position 1, and so on. An n-digit positive integer means an integer N satisfying

$$10^{n-1} \leq N < 10^n.$$

Throughout the paper, a fraction is counted as a displayed numerator–denominator pair (N, D), not as an equivalence class of rational numbers. Thus two different pairs may represent the same rational number but are counted separately.

We write $O_h(\cdot)$ to mean that the implied constant may depend on the fixed digit h, but not on n, k, B, X, or on any summation variables. Similarly, $A \ll_h B$ means $A = O_h(B)$. Although the constants may be made uniform in h because h ranges over the finite set $\{1, \dots, 9\}$, the subscript is retained to record the fixed-digit dependence and to facilitate comparison with base-q analogues. When summing over all nonzero decimal digits, I take the maximum of the nine implied constants.

Definition 1.1. Let $h \in \{0, 1, \dots, 9\}$, and let $k \geq 0$. We say that an integer $N \geq 0$ has digit h in position k if N can be written uniquely in the form

$$N = a10^{k+1} + h10^k + b, \qquad a \geq 0, \qquad 0 \leq b < 10^k.$$

Deleting the digit h in position k gives the integer

$$N' = a10^k + b.$$

The post-deletion expression is interpreted as an integer. Thus, if deleting the digit produces a displayed string beginning with one or more zeros, those leading zeros are discarded.

Here a is the full block of digits to the left of position k, not necessarily a single digit, and b is the full block of digits to the right of position k.

Definition 1.2. Fix $h \in \{1, \dots, 9\}$, and let $n \geq 1$. A triple (N, D, k) is called a *marked vertical h-cancellation of length n* if the following conditions hold:

(a) N and D are n-digit positive integers;

(b) $D < N$;

(c) $0 \leq k \leq n - 1$;

(d) both N and D have digit h in position k;

(e) after deleting these two occurrences of h, one obtains integers N' and D' with $D' > 0$ and
$$\frac{N}{D} = \frac{N'}{D'}.$$

The word *marked* means that the position k is part of the data.

Let $C_h(n, k)$ denote the number of marked vertical h-cancellations of length n with cancelled digit in position k. Define

$$C_h(n) = \sum_{k=0}^{n-1} C_h(n, k).$$

Remark 1.3. For every $h \in \{1, \dots, 9\}$, one has

$$C_h(1) = C_h(2) = 0.$$

When $n = 1$, the only digit position is both the leading and the units position. The endpoint argument used in the proof of Theorem 3.1 shows that no admissible cancellation exists. When $n = 2$, both possible cancellation positions are endpoint positions, so the same argument again gives no admissible cancellation.

Remark 1.4. The restriction $h \neq 0$ avoids degeneracies caused by zeros, especially trailing zeros. For example, deleting a terminal zero may simply divide both numerator and denominator by 10. Zero cancellations for fixed rational numbers were considered in earlier work, but the global counting model considered here has additional boundary effects and is better treated separately.

Our main result is the following.

Theorem 1.5. *Fix $h \in \{1, \dots, 9\}$. Then*

$$C_h(n) = O_h(n^2 10^{n-1}).$$

Consequently, among all pairs (N, D) of n-digit positive integers with $D < N$, the proportion that admit at least one vertical h-cancellation is

$$O_h\left(\frac{n^2}{10^n}\right).$$

In particular, this proportion tends to zero exponentially fast as $n \to \infty$.

By summing over the nine nonzero decimal digits, the same density estimate holds for pairs admitting a vertical cancellation of any nonzero digit. We also prove the following lower bound.

Theorem 1.6. *Fix $h \in \{1, \dots, 9\}$. There exists a constant $c_h > 0$ such that, for all sufficiently large n,*

$$C_h(n) \geq c_h 10^n.$$

Thus

$$c_h 10^n \leq C_h(n) \leq C'_h n^2 10^{n-1},$$

for suitable constants $c_h, C'_h > 0$ and all sufficiently large n.

The proof of the upper bound is elementary, but the main point is to recast the cancellation condition as a family of linear Diophantine equations and then estimate the resulting gcd sums. The lower bound is given by an explicit construction. The gap between these bounds remains open; we discuss this and related questions in the final section.

2. The cancellation equation and counting tools

We begin with the basic algebraic reduction. Similar digit decompositions appear in earlier work on cancellable representatives of a fixed rational number [1, 3]; the form below is tailored to the present lattice-point count over all n-digit pairs.

Lemma 2.1. *Let $h \in \{1, \ldots, 9\}$, let $k \geq 0$, and put $B = 10^k$. Suppose*

$$N = a10^{k+1} + h10^k + b, \qquad D = c10^{k+1} + h10^k + d,$$

where $a, c \geq 0$ and $0 \leq b, d < B$. After deleting the digit h in position k, one obtains

$$N' = aB + b, \qquad D' = cB + d.$$

Assume $D' > 0$. Then

$$\frac{N}{D} = \frac{N'}{D'},$$

if and only if

$$(9a + h)d - (9c + h)b = h(a - c)10^k. \tag{1}$$

Proof. Put $B = 10^k$. Then

$$N = (10a + h)B + b, \qquad D = (10c + h)B + d,$$

and

$$N' = aB + b, \qquad D' = cB + d.$$

Since $D > 0$ and $D' > 0$, the equality $N/D = N'/D'$ is equivalent to

$$ND' = DN'.$$

Thus

$$((10a + h)B + b)(cB + d) = ((10c + h)B + d)(aB + b).$$

Expanding both sides gives

$$(10ac + hc)B^2 + (10ad + hd + bc)B + bd,$$

on the left and

$$(10ac + ha)B^2 + (10bc + hb + ad)B + bd,$$

on the right. Cancelling common terms gives

$$h(c - a)B^2 + (9ad - 9bc + h(d - b))B = 0.$$

Since $B > 0$, division by B gives

$$h(c - a)B + 9ad - 9bc + h(d - b) = 0.$$

Equivalently,

$$(9a + h)d - (9c + h)b = h(a - c)B.$$

Since $B = 10^k$, this is the desired equation. □

We shall use two elementary estimates for solutions of linear equations in rectangular boxes.

Lemma 2.2. *Let A, C be positive integers, let $T \in \mathbb{Z}$, and let $B \geq 1$. The number of integer pairs (x, y) satisfying*

$$0 \leq x, y < B \qquad \text{and} \qquad Ay - Cx = T,$$

is at most

$$1 + \frac{B \gcd(A, C)}{\max\{A, C\}}.$$

Proof. Let $g = \gcd(A, C)$. If $g \nmid T$, then there are no integer solutions. Assume $g \mid T$, and choose one integer solution (x_0, y_0). The full integer solution set is

$$x = x_0 + \frac{A}{g}t, \qquad y = y_0 + \frac{C}{g}t, \qquad t \in \mathbb{Z}.$$

The condition $0 \leq x < B$ restricts t to an interval of length at most Bg/A, while the condition $0 \leq y < B$ restricts t to an interval of length at most Bg/C. Therefore the two conditions together restrict t to an interval of length at most

$$\frac{Bg}{\max\{A, C\}}.$$

An interval of length L contains at most $1+L$ integers. Hence the number of admissible solutions is at most

$$1 + \frac{B \gcd(A, C)}{\max\{A, C\}}.$$

□

Lemma 2.3. *Let $P, Q \in \mathbb{Z}$ be not both zero, let $T \in \mathbb{Z}$, and let $X \geq 1$. The number of integer pairs (x, y) satisfying*

$$0 \leq x, y < X \qquad \text{and} \qquad Px - Qy = T,$$

is at most

$$1 + \frac{X \gcd(|P|, |Q|)}{\max\{|P|, |Q|\}},$$

where $\gcd(|P|, 0) = |P|$ *and* $\gcd(0, |Q|) = |Q|$.

Proof. If $Q = 0$, then the equation is $Px = T$. It has either no solution for x, or exactly one value of x, and then at most X choices for y. In this case the asserted upper bound is $1 + X$. The case $P = 0$ is identical.

Now assume $P, Q \neq 0$. Let

$$g = \gcd(|P|, |Q|).$$

If $g \nmid T$, then there are no integer solutions. Otherwise choose one integer solution (x_0, y_0). The full integer solution set is

$$x = x_0 + \frac{Q}{g}t, \qquad y = y_0 + \frac{P}{g}t, \qquad t \in \mathbb{Z}.$$

The condition $0 \le x < X$ restricts t to an interval of length at most $Xg/|Q|$, and the condition $0 \le y < X$ restricts t to an interval of length at most $Xg/|P|$. Hence the two conditions together restrict t to an interval of length at most

$$\frac{Xg}{\max\{|P|, |Q|\}}.$$

Again, an interval of length L contains at most $1 + L$ integers. This gives the desired bound. □

We also need two gcd-sum estimates.

Lemma 2.4. *There is an absolute constant $K > 0$ such that, for all $X \ge 2$,*

$$\sum_{1 \le u,v \le X} \frac{\gcd(u,v)}{\max\{u,v\}} \le KX \log X.$$

Consequently, for each fixed $h \in \{1, \ldots, 9\}$,

$$\sum_{0 \le a,c < X} \frac{\gcd(9a+h, 9c+h)}{\max\{9a+h, 9c+h\}} = O_h(X \log(2X)),$$

for all $X \ge 1$.

Proof. We first prove the unshifted estimate. We use the identity

$$\gcd(u,v) = \sum_{\substack{r \mid u \\ r \mid v}} \varphi(r),$$

where φ is Euler's totient function. Therefore

$$\sum_{1 \le u,v \le X} \frac{\gcd(u,v)}{\max\{u,v\}} = \sum_{r \le X} \varphi(r) \sum_{\substack{1 \le u,v \le X \\ r \mid u,\ r \mid v}} \frac{1}{\max\{u,v\}}.$$

Writing $u = ru'$ and $v = rv'$, the inner sum is

$$\frac{1}{r} \sum_{1 \le u',v' \le X/r} \frac{1}{\max\{u',v'\}}.$$

For every real $Y \ge 1$,

$$\begin{aligned}\sum_{1 \le u',v' \le Y} \frac{1}{\max\{u',v'\}} &\le \sum_{m \le Y} \frac{\#\{(u',v') : 1 \le u',v' \le Y,\ \max(u',v') = m\}}{m} \\ &\le \sum_{m \le Y} \frac{2m-1}{m} \le 2Y.\end{aligned}$$

Hence

$$\sum_{1 \le u,v \le X} \frac{\gcd(u,v)}{\max\{u,v\}} \le 2X \sum_{r \le X} \frac{\varphi(r)}{r^2}.$$

Since $\varphi(r) \leq r$, we have

$$\sum_{r \leq X} \frac{\varphi(r)}{r^2} \leq \sum_{r \leq X} \frac{1}{r} = O(\log X).$$

This proves the first estimate.

For the shifted estimate, note that $9a + h$ and $9c + h$ are positive integers at most $9X + h$. Therefore the shifted sum is bounded above by

$$\sum_{1 \leq u, v \leq 9X+h} \frac{\gcd(u, v)}{\max\{u, v\}}.$$

The first estimate gives $O_h(X \log(2X))$, with the bounded range $1 \leq X < 2$ absorbed into the implied constant. □

Lemma 2.5. *Fix $h \in \{1, \ldots, 9\}$, let $k \geq 0$, and put $B = 10^k$. Then*

$$9d - hB \neq 0 \qquad \text{for every } 0 \leq d < B.$$

Consequently, for all $0 \leq b, d < B$, the quantity

$$\max\{|9d - hB|, |9b - hB|\},$$

is nonzero.

Proof. Suppose that $9d = hB = h10^k$ for some $0 \leq d < B$. Since $\gcd(9, 10^k) = 1$, we get $9 \mid h$. As $h \in \{1, \ldots, 9\}$, this forces $h = 9$. Then $9d = 9B$, so $d = B$, contradicting $d < B$. This proves the assertion. The final statement follows by applying the same argument to both variables. □

Lemma 2.6. *Fix $h \in \{1, \ldots, 9\}$, let $k \geq 0$, and put $B = 10^k$. Then*

$$\sum_{0 \leq b, d < B} \frac{\gcd(|9d - hB|, |9b - hB|)}{\max\{|9d - hB|, |9b - hB|\}} = O(B \log(2B)),$$

with an absolute implied constant.

Proof. For $0 \leq d < B$, the integer $9d - hB$ lies in the interval

$$[-hB, (9 - h)B - 9].$$

In particular,

$$|9d - hB| \leq 9B.$$

The same bound holds for $|9b - hB|$. By Lemma 2.5, the denominator in the displayed sum is never zero.

The map
$$d \longmapsto |9d - hB|,$$
from $\{0, 1, \ldots, B-1\}$ to $\{1, 2, \ldots, 9B\}$ has multiplicity at most 2. Indeed, if
$$|9d_1 - hB| = |9d_2 - hB|,$$
then either $9d_1 - hB = 9d_2 - hB$, giving $d_1 = d_2$, or
$$9d_1 - hB = -(9d_2 - hB),$$
which determines d_2 uniquely from d_1. The same multiplicity bound holds for the corresponding map in the variable b. Therefore the induced map on ordered pairs (b, d) has multiplicity at most 4. Hence the sum in the statement is at most
$$4 \sum_{1 \leq u,v \leq 9B} \frac{\gcd(u, v)}{\max\{u, v\}}.$$
By Lemma 2.4, this is $O(B \log(2B))$, with an absolute implied constant. □

3. The upper bound: exponential rarity

We now prove the main counting estimate.

Theorem 3.1. *Fix $h \in \{1, \ldots, 9\}$. Then*
$$C_h(n) = O_h(n^2 10^{n-1}).$$

Proof. We estimate $C_h(n, k)$ for each position k.

First suppose $k = 0$. Then $b = d = 0$. If $c = 0$, then after deleting the units digit from D one obtains $D' = 0$, so no marked cancellation is valid. If $c > 0$, then Lemma 2.1 applies and gives
$$h(a - c) = 0.$$
Since $h \neq 0$, this implies $a = c$, and hence $N = D$, contradicting $D < N$. Thus
$$C_h(n, 0) = 0.$$
Next suppose $k = n - 1$, so the cancelled digit is the leading digit. Since $k = n - 1$, the decomposition has no digits to the left of the cancelled digit, so $a = c = 0$. If $d = 0$, then after deleting the leading digit from D one obtains $D' = 0$, so no marked cancellation is valid. If $d > 0$, then Lemma 2.1 applies and gives
$$h(d - b) = 0.$$
Again $h \neq 0$, so $b = d$, and therefore $N = D$, contradicting $D < N$. Thus
$$C_h(n, n-1) = 0.$$

It remains to treat $1 \leq k \leq n-2$. Put

$$B = 10^k, \qquad X = 10^{n-k-1}.$$

Every marked vertical h-cancellation in position k has the form

$$N = a10^{k+1} + h10^k + b, \qquad D = c10^{k+1} + h10^k + d,$$

with

$$0 \leq b, d < B.$$

Here a and c are the entire blocks of digits to the left of the cancelled digit. Since N and D are n-digit numbers and $1 \leq k \leq n-2$, the exact restrictions on these left blocks are

$$10^{n-k-2} \leq a, c < 10^{n-k-1} = X.$$

In particular, $c \geq 10^{n-k-2} \geq 1$, and hence

$$D' = cB + d > 0,$$

automatically in the exact interior range.

For the purpose of obtaining an upper bound, I enlarge the left-block range to

$$0 \leq a, c < X.$$

After this enlargement, I deliberately omit both the condition $D < N$ and the restriction $D' > 0$. Removing these conditions can only increase the number of solutions, so the resulting count remains a valid upper bound.

By Lemma 2.1, every valid cancellation satisfies

$$(9a + h)d - (9c + h)b = h(a - c)B. \tag{2}$$

We first fix a, c. Then (2) is a linear equation in b, d. Applying Lemma 2.2 with

$$A = 9a + h, \qquad C = 9c + h, \qquad T = h(a - c)B,$$

we find that the number of possible pairs (b, d) is at most

$$1 + \frac{B \gcd(9a + h, 9c + h)}{\max\{9a + h, 9c + h\}}.$$

Therefore, by Lemma 2.4,

$$\begin{aligned} C_h(n, k) &\leq \sum_{0 \leq a, c < X} \left(1 + \frac{B \gcd(9a + h, 9c + h)}{\max\{9a + h, 9c + h\}}\right) \\ &\ll_h X^2 + BX \log(2X). \end{aligned}$$

Since $BX = 10^{n-1}$, this gives

$$C_h(n, k) \ll_h X^2 + 10^{n-1} \log(2X). \tag{3}$$

We obtain a complementary estimate by fixing b, d. Rewriting (2), we get

$$(9d - hB)a - (9b - hB)c = h(b - d). \tag{4}$$

For fixed b, d, this is a linear equation in a, c. By Lemma 2.5, the two coefficients $9d-hB$ and $9b-hB$ are not both zero, indeed each is nonzero. Hence Lemma 2.3 applies and gives at most

$$1 + \frac{X \gcd(|9d - hB|, |9b - hB|)}{\max\{|9d - hB|, |9b - hB|\}},$$

solutions (a, c) with $0 \leq a, c < X$. Summing over $0 \leq b, d < B$ and using Lemma 2.6, we obtain

$$C_h(n, k) \ll B^2 + XB \log(2B) = B^2 + 10^{n-1} \log(2B).$$

Thus

$$C_h(n, k) \ll_h B^2 + 10^{n-1} \log(2B). \tag{5}$$

Combining (3) and (5), we get

$$C_h(n, k) \ll_h \min\{X^2, B^2\} + 10^{n-1}(\log(2X) + \log(2B)).$$

Since

$$\min\{X^2, B^2\} \leq XB = 10^{n-1}$$

and

$$\log(2X) + \log(2B) \ll n,$$

we conclude that

$$C_h(n, k) \ll_h n10^{n-1},$$

uniformly for $1 \leq k \leq n-2$. Summing over the at most n possible positions k yields

$$C_h(n) = \sum_{k=0}^{n-1} C_h(n, k) \ll_h n^2 10^{n-1}.$$

This proves the theorem. □

The density statement follows immediately.

Corollary 3.2. *Fix $h \in \{1, \ldots, 9\}$. Among all pairs (N, D) of n-digit positive integers with $D < N$, the proportion admitting at least one vertical h-cancellation is*

$$O_h\left(\frac{n^2}{10^n}\right).$$

In particular, vertically h-cancellable pairs have density zero, indeed exponentially fast.

Proof. There are

$$M_n = 9 \cdot 10^{n-1},$$

positive n-digit integers. Therefore the number of pairs (N, D) with $D < N$ and both N, D n-digit is

$$\binom{M_n}{2} \asymp 10^{2n}.$$

Let $U_h(n)$ denote the number of such pairs (N, D) admitting at least one vertical h-cancellation. Since $C_h(n)$ counts marked cancellations, each pair counted by $U_h(n)$ contributes at least one to $C_h(n)$. Hence

$$U_h(n) \leq C_h(n).$$

By Theorem 3.1,

$$\frac{U_h(n)}{\binom{M_n}{2}} \leq \frac{C_h(n)}{\binom{M_n}{2}} = O_h\left(\frac{n^2 10^{n-1}}{10^{2n}}\right) = O_h\left(\frac{n^2}{10^{n+1}}\right) = O_h\left(\frac{n^2}{10^n}\right).$$

This proves the claim. □

Corollary 3.3. *Let $U_{\neq 0}(n)$ denote the number of pairs (N, D) of n-digit positive integers with $D < N$ that admit at least one vertical cancellation of some nonzero digit $h \in \{1, \ldots, 9\}$. Then*

$$U_{\neq 0}(n) = O(n^2 10^{n-1}).$$

Consequently, the proportion of such pairs among all displayed pairs of n-digit positive integers with $D < N$ is

$$O\left(\frac{n^2}{10^n}\right).$$

In particular, the proportion of displayed pairs admitting a vertical cancellation of any nonzero digit tends to zero exponentially fast.

Proof. Every pair counted by $U_{\neq 0}(n)$ admits a vertical cancellation of at least one digit $h \in \{1, \ldots, 9\}$. Therefore

$$U_{\neq 0}(n) \leq \sum_{h=1}^{9} C_h(n).$$

By Theorem 3.1, for each $h \in \{1, \ldots, 9\}$,

$$C_h(n) = O_h(n^2 10^{n-1}).$$

Since h ranges over a finite set, the implied constants may be replaced by their maximum. Hence

$$U_{\neq 0}(n) = O(n^2 10^{n-1}).$$

The number of pairs (N, D) with $D < N$ and both N and D n-digit positive integers is of order 10^{2n}. Dividing by this quantity gives

$$O\left(\frac{n^2}{10^{n+1}}\right),$$

and hence also

$$O\left(\frac{n^2}{10^n}\right).$$

□

4. A lower bound and numerical evidence

The preceding upper bound shows that vertical cancellations are rare among all displayed pairs, but it does not by itself show that many such cancellations exist. We now give a simple explicit construction that supplies exponentially many examples.

Proposition 4.1. *Fix $h \in \{1, \dots, 9\}$. There exists a constant $c_h > 0$ such that, for all sufficiently large n,*

$$C_h(n) \geq c_h 10^n.$$

Proof. Let

$$k = \left\lceil \frac{n}{2} \right\rceil, \qquad B = 10^k, \qquad X = 10^{n-k-1}.$$

For all sufficiently large n, we have $1 \leq k \leq n-2$. Set

$$H = \frac{h(B-1)}{9}.$$

Since $B = 10^k \equiv 1 \pmod 9$, the integer H is well defined. Also $0 \leq H < B$ because $1 \leq h \leq 9$.

Let

$$I = \{10^{n-k-2}, 10^{n-k-2}+1, \dots, X-1\}.$$

For all sufficiently large n, the choice of $k = \lceil n/2 \rceil$ gives $H \geq X-1$. Indeed, if $n = 2m$, then $k = m$ and $X = 10^{m-1}$, while

$$H \geq \frac{10^m - 1}{9} \geq X - 1;$$

if $n = 2m+1$, then $k = m+1$ and $X = 10^{m-1}$, while

$$H \geq \frac{10^{m+1} - 1}{9} \geq X - 1.$$

Choose any pair $a, c \in I$ with $a > c$, and define

$$b = H - a, \qquad d = H - c.$$

Then $0 \leq b, d < B$. Also $c > 0$, so the denominator after deletion satisfies $D' = cB + d > 0$. Moreover a and c lie in the exact n-digit left-block range, namely

$$10^{n-k-2} \leq a, c < 10^{n-k-1},$$

so the numbers

$$N = a10^{k+1} + h10^k + b, \qquad D = c10^{k+1} + h10^k + d,$$

are both n-digit positive integers with digit h in position k.

We verify the cancellation equation. Since $9H = h(B-1)$, we have

$$9b - hB = 9(H-a) - hB = -(9a+h),$$

and similarly

$$9d - hB = -(9c+h).$$

Equivalently,

$$\begin{aligned}(9a+h)d - (9c+h)b &= (9a+h)(H-c) - (9c+h)(H-a)\\ &= (9H+h)(a-c)\\ &= hB(a-c).\end{aligned}$$

Thus Lemma 2.1 shows that deleting the two digits h in position k leaves the represented rational number unchanged.

Finally,

$$N - D = (a-c)10^{k+1} + (b-d) = (a-c)10^{k+1} - (a-c) = (a-c)(10^{k+1}-1) > 0,$$

so $D < N$. Therefore every pair $a, c \in I$ with $a > c$ gives a marked vertical h-cancellation.

The number of such choices is

$$\binom{|I|}{2}, \qquad |I| = X - 10^{n-k-2} = \frac{9}{10}X.$$

For the chosen k, one has $X^2 \asymp 10^n$, with an absolute implied constant depending only on the parity of n. Hence

$$\binom{|I|}{2} \gg 10^n.$$

This proves the claim. □

Combining Proposition 4.1 with Theorem 3.1 gives Theorem 1.6.

Remark 4.2. The lower-bound construction is not meant to be sharp. It shows that the order of magnitude is at least 10^n, while the upper bound is $O_h(n^2 10^{n-1})$. The remaining gap is polynomial in n.

The following table gives exact values of $C_h(n)$ for small n. The supplementary files `VCC.py` and `VCC.csv` contain, respectively, the source code and the position-wise values $C_h(n,k)$ used to obtain these totals. For each $h \in \{1, \ldots, 9\}$, each $n \in \{2, \ldots, 7\}$, and each interior position $1 \le k \le n-2$, the program uses the exact ranges

$$10^{n-k-2} \le a, c < 10^{n-k-1}, \qquad 0 \le b, d < 10^k.$$

For a solution of the cancellation Eq. (1), the condition $N > D$ is equivalent to $a > c$. Indeed, if $a = c$, then (1) forces $b = d$ and hence $N = D$, while the bound $|b - d| < 10^k$ shows that the sign of $N - D$ agrees with the sign of $a - c$ whenever $a \neq c$.

The computation does not traverse all $O(10^{2n})$ displayed numerator–denominator pairs. For each position, it loops over the smaller of the (a, c)-box and the (b, d)-box. Once two variables are fixed, the remaining two satisfy a linear Diophantine equation. One solution is found by the extended Euclidean algorithm, and the admissible solutions are counted by intersecting the resulting arithmetic progression with the exact digit-block bounds.

The computation was performed in Python 3.13.5 using only the Python standard library and arbitrary-precision integer arithmetic. The program verifies that $C_h(n, 0) = C_h(n, n-1) = 0$, that the position-wise values sum to the reported totals, and that the totals agree with Table 1. As an independent check, for $n \leq 4$ each position-wise count is also compared with a direct traversal of the exact digit-block ranges.

Table 1. Exact values of $C_h(n)$ for $2 \leq n \leq 7$ and $1 \leq h \leq 9$.

n	$C_1(n)$	$C_2(n)$	$C_3(n)$	$C_4(n)$	$C_5(n)$	$C_6(n)$	$C_7(n)$	$C_8(n)$	$C_9(n)$
2	0	0	0	0	0	0	0	0	0
3	1	2	21	6	11	19	22	28	72
4	159	154	462	129	143	421	150	147	1517
5	2387	2318	8578	2622	3126	8753	4339	5006	31246
6	47003	44800	136482	42323	43310	133577	44676	45111	483462
7	622583	609175	2000457	633934	682651	2017787	810603	883894	7372622

The values in Table 1 are exact small-n enumerations and are not sufficient to distinguish reliably among growth of order 10^n, $n10^n$, $n^2 10^{n-1}$, or other polynomial corrections to exponential growth. The table also shows a marked dependence on the cancelled digit h: the values for $h = 3$, 6, and especially 9 are substantially larger than those for most other digits. Any asymptotic interpretation of these differences remains conjectural.

5. Multiple cancellations and further questions

Theorem 3.1 shows that vertical cancellations are exponentially rare. A natural next question is whether a typical vertically cancellable pair has only one cancellation.

For a pair (N, D) of n-digit positive integers with $D < N$, let

$$r_h(N, D) = \#\{k : (N, D, k) \text{ is a marked vertical } h\text{-cancellation}\}.$$

Then

$$C_h(n) = \sum_{(N,D)} r_h(N, D),$$

where the sum is over all pairs (N, D) of n-digit positive integers with $D < N$.

For $j \geq 1$, define

$$A_{h,j}(n) = \#\{(N, D) : r_h(N, D) = j\}.$$

In particular, set

$$E_h(n) = A_{h,1}(n), \qquad M_h(n) = \sum_{j\geq 2} A_{h,j}(n),$$

and

$$R_h(n) = \sum_{j\geq 2} jA_{h,j}(n) = \sum_{\substack{(N,D) \\ r_h(N,D)\geq 2}} r_h(N,D).$$

Thus $E_h(n)$ counts pairs with exactly one vertical h-cancellation, $M_h(n)$ counts pairs with at least two such cancellations, and $R_h(n)$ counts the marked cancellations contributed by multiply cancellable pairs. Consequently,

$$C_h(n) = E_h(n) + R_h(n).$$

Conjecture 5.1. *Fix $h \in \{1, \ldots, 9\}$. Then*

$$R_h(n) = o(C_h(n)),$$

as $n \to \infty$. Equivalently,

$$E_h(n) \sim C_h(n).$$

In words, almost every marked vertical h-cancellation belongs to a pair that is cancellable in exactly one vertical way.

Different cancellation positions give different digit-block decompositions of the same pair (N, D). The resulting Diophantine constraints are therefore coupled, and no probabilistic independence between them is asserted here. Since endpoint cancellations are impossible, one has $r_h(N, D) \leq n - 2$ for $n \geq 2$. It follows that

$$R_h(n) \leq (n-2)M_h(n).$$

In particular, a sufficient condition for Conjecture 5.1 is

$$M_h(n) = o\left(\frac{C_h(n)}{n}\right).$$

This distinction is important because $C_h(n)$ and $R_h(n)$ count marked cancellations, whereas $M_h(n)$ counts numerator–denominator pairs.

The supplementary computational output gives the complete distribution of $r_h(N, D)$ for every $2 \leq n \leq 7$ and $1 \leq h \leq 9$. Tables 2 and 3 record the corresponding values of $M_h(n)$ and $R_h(n)$.

Table 2. The number $M_h(n)$ of pairs admitting at least two vertical h-cancellations.

n	$M_1(n)$	$M_2(n)$	$M_3(n)$	$M_4(n)$	$M_5(n)$	$M_6(n)$	$M_7(n)$	$M_8(n)$	$M_9(n)$
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	1	2	21	6	11	19	22	28	72
5	158	152	449	123	132	407	128	119	1610
6	2237	2174	8544	2509	3016	8643	4240	4907	34605
7	45137	42976	137295	40201	40929	132470	41316	40914	547416

Table 3. The number $R_h(n)$ of marked cancellations contributed by pairs admitting at least two vertical h-cancellations

n	$R_1(n)$	$R_2(n)$	$R_3(n)$	$R_4(n)$	$R_5(n)$	$R_6(n)$	$R_7(n)$	$R_8(n)$	$R_9(n)$
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	2	4	42	12	22	38	44	56	144
5	317	306	919	252	275	833	278	266	3292
6	4633	4502	17566	5147	6175	17717	8630	9961	71057
7	92679	88290	284064	83056	85050	274340	87073	86930	1136997

For $n = 2, 3$, no pair has more than one vertical h-cancellation. For $n = 4$, every pair counted by $M_h(n)$ has exactly two cancellation positions. For $5 \le n \le 7$, higher multiplicities occur, and the largest observed value of $r_h(N, D)$ is $n-2$, the total number of interior positions. Throughout the computed range, one has

$$R_h(n) < \frac{1}{2}C_h(n).$$

Thus more than half of the marked cancellations in every computed case arise from pairs having exactly one cancellation position. The largest observed ratio is

$$\frac{R_8(4)}{C_8(4)} = \frac{56}{147} < 0.381.$$

These computations provide preliminary evidence concerning Conjecture 5.1, but the range is too short to support an asymptotic conclusion.

The upper-bound argument also extends to every fixed base.

Theorem 5.2. *Fix an integer $q \ge 2$ and a nonzero base-q digit $h \in \{1, \dots, q-1\}$. Let $C_{q,h}(n)$ denote the number of marked vertical cancellations of the digit h among pairs (N, D) such that N and D are n-digit positive integers in base q and $D < N$. Then*

$$C_{q,h}(n) = O_{q,h}(n^2 q^{n-1}).$$

Consequently, the proportion of such pairs admitting at least one vertical cancellation of the digit h is

$$O_{q,h}\left(\frac{n^2}{q^n}\right).$$

Moreover, the number of pairs admitting a vertical cancellation of at least one nonzero base-q digit is $O_q(n^2 q^{n-1})$, and their proportion is $O_q(n^2/q^n)$.

Proof. Let $C_{q,h}(n, k)$ denote the number of marked vertical cancellations in position k, and put $B = q^k$. If the digit h occurs in position k, write

$$N = aq^{k+1} + hq^k + b, \qquad D = cq^{k+1} + hq^k + d,$$

where $a, c \geq 0$ and $0 \leq b, d < B$. After deleting the two occurrences of h, one obtains

$$N' = aB + b, \qquad D' = cB + d.$$

Cross-multiplication shows that the cancellation condition is equivalent to

$$((q-1)a + h)d - ((q-1)c + h)b = h(a-c)B.$$

As in the decimal case, the endpoint positions contribute nothing. It therefore remains to consider $1 \leq k \leq n-2$. Put $X = q^{n-k-1}$. For an upper bound, enlarge the exact left-block ranges to $0 \leq a, c < X$ and omit the conditions $D < N$ and $D' > 0$.

First fix a, c. The number of pairs (b, d) with $0 \leq b, d < B$ satisfying the cancellation equation is at most

$$1 + \frac{B \gcd((q-1)a + h, (q-1)c + h)}{\max\{(q-1)a + h, (q-1)c + h\}}.$$

The gcd-sum estimate used in Lemma 2.4 gives

$$\sum_{0 \leq a, c < X} \frac{\gcd((q-1)a + h, (q-1)c + h)}{\max\{(q-1)a + h, (q-1)c + h\}} \ll_{q,h} X \log(2X).$$

Hence

$$C_{q,h}(n, k) \ll_{q,h} X^2 + BX \log(2X).$$

Since $BX = q^{n-1}$, this yields

$$C_{q,h}(n, k) \ll_{q,h} X^2 + q^{n-1} \log(2X).$$

For the complementary estimate, rewrite the cancellation equation as

$$((q-1)d - hB)a - ((q-1)b - hB)c = h(b-d).$$

The two coefficients in this equation are nonzero. Indeed, suppose that $(q-1)d = hq^k$. Since $\gcd(q-1, q^k) = 1$, it follows that $q - 1 \mid h$. Since $1 \leq h \leq q-1$, this forces $h = q-1$. The equation then gives $d = q^k = B$, contradicting $d < B$.

The map

$$d \longmapsto |(q-1)d - hB|,$$

has multiplicity at most 2, and its values lie between 1 and $(q-1)B$. The same argument used in Lemma 2.6 therefore gives

$$\sum_{0 \leq b, d < B} \frac{\gcd(|(q-1)d - hB|, |(q-1)b - hB|)}{\max\{|(q-1)d - hB|, |(q-1)b - hB|\}} \ll_q B \log(2B).$$

Applying the rectangular-box estimate now yields

$$C_{q,h}(n, k) \ll_{q,h} B^2 + q^{n-1} \log(2B).$$

Combining the two estimates gives

$$C_{q,h}(n, k) \ll_{q,h} \min\{X^2, B^2\} + q^{n-1}(\log(2X) + \log(2B)).$$

Since $\min\{X^2, B^2\} \le XB = q^{n-1}$ and $\log(2X) + \log(2B) \ll_q n$, it follows that

$$C_{q,h}(n,k) \ll_{q,h} nq^{n-1},$$

uniformly over the interior positions. Summing over at most n positions gives

$$C_{q,h}(n) \ll_{q,h} n^2 q^{n-1}.$$

There are $(q-1)q^{n-1}$ positive n-digit integers in base q. Therefore the number of pairs (N, D) with $D < N$ is of order q^{2n}. Dividing the preceding estimate by this quantity gives

$$O_{q,h}\left(\frac{n^2}{q^n}\right).$$

Finally, summing over the $q-1$ nonzero digits and taking the maximum of the finitely many implied constants proves the all-nonzero-digits assertion. □

The results above leave several natural problems.

Problem 5.3. *Determine the true order of magnitude of $C_h(n)$. Is it true that*

$$C_h(n) \asymp_h n10^n?$$

More strongly, does there exist a constant $\kappa_h > 0$ such that

$$C_h(n) \sim \kappa_h n10^n?$$

Problem 5.4. *Prove or disprove Conjecture 5.1. That is, are almost all marked vertical h-cancellations supported on pairs that are cancellable in exactly one vertical way?*

Problem 5.5. *Develop an analogous counting theory for cancellable blocks rather than single digits. More precisely, fix a finite word W in the alphabet $\{0, 1, \ldots, 9\}$, allow W to occur as a contiguous block of decimal digits, and count pairs (N, D) for which the same occurrence of W can be deleted vertically from the numerator and denominator without changing the represented rational number. One should specify separately whether leading zeros in the resulting numerator or denominator are allowed.*

Problem 5.6. *Treat the digit $h = 0$ in the present global counting model. Zero cancellations for representatives of a fixed rational number were studied in [1, 3], but the global n-digit pair-counting problem has additional phenomena arising from trailing zeros and should be separated from the nonzero-digit case.*

Data availability statement

The source code used to generate Table 1, together with the position-wise values $C_h(n, k)$ and the resulting totals $C_h(n)$, is provided in the supplementary files `VCC.py` and `VCC.csv`. A short description of that computation is provided in `VCC.txt`. The source code and complete multiplicity distributions used for Tables 2 and 3 are provided in `CMCs.cpp` and `MCS.csv`. The computations use exact integer arithmetic, and no external datasets were used.

References

[1] M. Aliabadi, J. Manheim, E. Nategh, and H. Shahmohamad. Cancellable numbers. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 102:45–54, 2017.

[2] J. Manheim and H. Shahmohamad. On slopes of cancellable numbers. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 110:3–7, 2019.

[3] J. Manheim and H. Shahmohamad. Some generalizations and limit theorems on cancellable numbers. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 114:361–371, 2020.

Mohsen Aliabadi
Department of Mathematics, Clayton State University
2000 Clayton State Boulevard, Morrow, GA 30260, USA
E-mail maliabadi@clayton.edu