# How Many Cycle Lengths Can a Unique Quotient Require?

Nuutti Kuosa

23 November 2026

hexagon:2611.00945v1

**Abstract**

Given two permutations $A$ and $B$, suppose there is just one finite dynamical system $X$, up to isomorphism, whose direct product with $A$ is $B$. How many different cycle lengths can $X$ have? For every sufficiently large $k$ we construct a divisor with $k$ cycle lengths and a target with exactly $k+1$. Their unique quotient has asymptotically $\frac{1}{2}k^2 \log_2 \log k$ distinct cycle lengths. Every one of these lengths is absent from the target. The divisor is fixed once $k$ is chosen. Its cycles have lengths $M/p$, where $M$ is the product of the first $k$ primes and $p$ runs over those primes. The proof chooses a set containing almost all the first $k$ primes. Each quotient cycle omits exactly two primes of this set. A congruence fixes the number of states assigned to each missing pair. Smaller primes then distinguish several forced cycle lengths within each pair.

## 1 A quotient whose lengths are absent from its product

Let $C_d$ denote one cycle of length $d$, and let $mC_d$ denote $m$ disjoint copies. A permutation is a sum of such cycles. Two permutations are isomorphic when each cycle length occurs with the same multiplicity in both. In the product $C_aC_b$ both coordinates advance one step at a time. A pair returns to its starting point after $\operatorname{lcm}(a,b)$ steps, so the $ab$ states form $\gcd(a,b)$ cycles of that length. Thus $C_2C_3 = C_6$: both factor lengths disappear from the product. In a division equation $AX \cong B$, we call the known factor $A$ the *divisor*. How many lengths can disappear when the quotient $X$ is uniquely determined?

Write $\operatorname{Len}(Y)$ for the set of distinct cycle lengths of $Y$. Let $P_k = \{p_1, \dots, p_k\}$ be the first $k$ primes in increasing order and set $M = \prod_{p \in P_k} p$. Our divisor has one cycle for each prime omitted from this product.

**Theorem 1.1** (A unique quotient with many cycle lengths)**.** *For every sufficiently large $k$ there is a permutation $B_k$ with exactly $k+1$ distinct cycle lengths such that, for*

$$A_k = \sum_{p \in P_k} C_{M/p},$$

*the equation $A_kX \cong B_k$ has exactly one solution $X_k$ up to isomorphism among all finite dynamical systems. As $k \to \infty$,*

$$|\operatorname{Len}(X_k)| \sim \tfrac{1}{2}k^2 \log_2 \log k, \qquad \operatorname{Len}(X_k) \cap \operatorname{Len}(B_k) = \varnothing.$$

A finite dynamical system here is a finite set with a map to itself. If the product is a permutation, a nonperiodic state in either factor would give a nonperiodic state in the product. Thus every candidate quotient in the theorem is already a sum of cycles. Throughout, log denotes the natural logarithm.

For a finite instance, take $k = 271$. The construction in Remark 3.4 gives a target with 272 distinct lengths whose unique quotient has 29 915. We begin with a much smaller example in which the quotient can be recovered by hand.

**Recovering a quotient by hand.** Take the first three primes, so that $M = 30$. Consider

$$(C_{15} + C_{10} + C_6)X = 8C_{15} + 7C_{10} + 5C_6 + 3C_{30}.$$

The divisor has 31 states and the target 310, so $X$ has ten. Each possible length $d$ in $X$ divides 30, since $d$ divides every product length to which its cycle contributes. Look at the eight 15-cycles. Only multiplication by $C_{15}$ can produce them. An odd $d$ contributes $d$ copies of $C_{15}$, one for each state of $C_d$, while an even $d$ produces 30-cycles. We have therefore counted eight states of odd period. The remaining two states force one $C_2$. In the same way the seven 10-cycles leave three states of period divisible by 3, and the five 6-cycles leave five states of period divisible by 5. They force one $C_3$ and one $C_5$. These cycles exhaust the ten states, and multiplication verifies the answer $X = C_2 + C_3 + C_5$.

**Can the same counts give two quotients?** Keep the known factor $A_3 = C_{15} + C_{10} + C_6$. Write $s_p$ for the number of states whose period is divisible by $p$. The two sums

$$3C_{10} + 2C_{15} \qquad \text{and} \qquad 6C_5 + C_{30}$$

each have 60 states and give $s_2 = s_3 = 30$ and $s_5 = 60$. Multiplying either by $A_3$ gives $30C_{15} + 30C_{10} + 37C_{30}$. Since $s_5 = 60$ already requires sixty states, these are two distinct minimum-size realisations of the same marginals.

**Where the large number of lengths comes from.** To force a quotient, we will choose counts $s_p$ for which exactly one sum $X_0$ attains the smallest size. Then we take $B = A_k X_0$. Any quotient of $B$ by $A_k$ must have the counts read from $B$ and the size $|B|/|A_k| = |X_0|$. It is therefore the same minimum-size sum $X_0$. Lemma 2.1 justifies this reading of the counts.

We say that a cycle *misses* $p$ if its length is not divisible by $p$. To force many lengths, choose a set $Q$ containing almost all the first $k$ primes. Arrange the counts so that every cycle in a smallest realisation omits exactly two primes of $Q$. There are about $k^2/2$ choices for the missing pair. A congruence determines the number of states assigned to each pair. A second set of primes, called markers, then forces about $\log_2 \log k$ different lengths for *each* pair. Eisenbrand and Shmonin [16] give an earlier example in which a unique sum uses every one of a chosen set of integer vectors. Here every length $d$ dividing the integer $M$ is available. The proof must explain why all competing choices fail. Section 3 gives the construction. Section 3.2 returns to the product and proves Theorem 1.1.

**Relation to division algorithms.** Doré et al. find a connected quotient in polynomial time when one exists [13]. For sums of cycles, Bridoux et al. can count all quotient classes in polynomial time for each fixed divisor [8]. Thus their result applies to each fixed $A_k$, with a polynomial degree that can depend on $k$. Does a single polynomial bound suffice to find a quotient when $k$ varies and the input permutations are given by their transition tables? Theorem 1.1 shows that the only answer can contain more than any fixed multiple of $k^2$ distinct lengths. The running time for this divisor family remains open.

Crespelle, Phan and Richard describe all cycle-sum quotients by recursion on the number of primes [9, Theorem 3.8]. In their notation our question asks how large $\mathrm{card}(X)$, the number of distinct lengths, can be when that recursion has a single quotient to describe [9, §2.1]. For unrestricted explicit divisors, deciding existence is NP-complete even for permutations [19, Theorem 1]. For background on the semiring of finite dynamical systems and the equations studied in it, see [12, 20, 21].

## 2 What the product tells us about the quotient

### Reading the target

For a finite set $P$ of primes let $M_P = \prod_{p\in P} p$. Consider a sum $X = \sum_{d|M_P} x_d C_d$, where $x_d$ is the number of $d$-cycles. Each cycle contributes $d$ states. The *marginal at $p$* counts the states on all cycles whose length is divisible by $p$. Thus

$$|X| = \sum_{d|M_P} d x_d, \qquad s_p(X) = \sum_{\substack{d|M_P \\ p|d}} d x_d \quad (p \in P).$$

Distinct primes may count the same state. For example, a single 6-cycle contributes six to each of $s_2$ and $s_3$, but only six to $|X|$. In the notation used below, *size* always counts states and *support size* always counts distinct cycle lengths.

**Lemma 2.1** (Reading the marginals from a product)**.** *Let $k \geq 2$ and $M = M_{P_k}$. Suppose the cycle lengths of the permutation $B$ belong to $\{M\} \cup \{M/p : p \in P_k\}$. Write $B(\ell)$ for the multiplicity of $C_\ell$ in $B$, taking a missing length to have multiplicity zero. Then $A_k X \cong B$ holds exactly for sums of cycles with lengths dividing $M$ that satisfy*

$$|X| = |B|/|A_k|, \qquad s_p(X) = |X| - B(M/p) \quad (p \in P_k).$$

*Proof.* A nonperiodic state $x$ would make $(a, x)$ nonperiodic for every state $a$ of $A_k$. Thus $X$ is a sum of cycles. Each length $d$ in $X$ divides the product length $\operatorname{lcm}(d, M/p)$, which in turn divides $M$. Since $M$ is squarefree, there are just two outcomes:

$$C_{M/p} C_d = \begin{cases} d\, C_{M/p}, & p \nmid d, \\ (d/p)\, C_M, & p \mid d. \end{cases}$$

Summing over the cycles of $X$ gives

$$A_k X = \sum_{p\in P_k} (|X| - s_p(X)) C_{M/p} + \left( \sum_{p\in P_k} \frac{s_p(X)}{p} \right) C_M. \tag{1}$$

The $k$ shorter lengths are distinct. Their coefficients give exactly the stated marginals. Conversely, once these coefficients agree, the equality of the two total sizes fixes the remaining coefficient at $M$. □

### Choosing a size that forces a quotient

For any nonnegative vector $s = (s_p)_{p\in P}$ with $p \mid s_p$, let

$$F(s) = \min\{|X| : s_p(X) = s_p \text{ for every } p \in P\},$$

where the allowed lengths divide $M_P$. A realisation always exists: $\sum_p (s_p/p) C_p$ has these marginals. We allow the empty sum for $s = 0$. An *optimum* is a realisation with $F(s)$ states. Removing a fixed point changes none of the marginals, so an optimum contains no fixed point.

For a target in Lemma 2.1, first check that $n = |B|/|A_k|$ is an integer and each $s_p = n - B(M/p)$ is nonnegative. Divisibility by $p$ then follows by reducing $n \sum_{a\in P_k} M/a = MB(M) + \sum_{a\in P_k} (M/a) B(M/a)$ modulo $p$, since only the terms indexed by $a = p$ survive. The lemma shows that a quotient exists exactly when $F(s) \leq n$: one can add $n - F(s)$ fixed points to an optimum. In particular, if $n = F(s)$, the quotients are precisely the optima. This last observation is how we will force uniqueness. An arbitrary quotient can be larger than an optimum. For $A_2 = C_2 + C_3$, for instance, $C_1 + C_2$ is a three-state quotient of $3C_2 + C_3 + C_6$ whose marginals already occur on the two-state system $C_2$.

**Remark** (An interpretation using group actions)**.** Let the cyclic group of order $M$ act on $X$ by sending its generator to one time step. The subgroup of order $p$ fixes the states of period dividing $M/p$, so its fixed-point count (its *mark*) is $|X| - s_p$. Together with $|X|$, the mark at the trivial subgroup, the product thus gives $k+1$ of the $2^k$ subgroup marks. The full collection determines each $x_d$ in order of increasing $d$. A $d$-cycle contributes all $d$ states after $m$ steps if $d \mid m$, and zero otherwise. The fixed-point count is therefore $\sum_{d|m} dx_d$ for each $m \mid M$. This is the finite case of the mark formula for cyclic actions [14, author version, (2-6-1) and (2-6-4)].

# 3 Forcing many different cycle lengths

## 3.1 A cycle for each pair of missing primes

Return to the ten-state quotient $C_2 + C_3 + C_5$. Its three lengths can also be read as the three pairs of primes they miss. For instance, missing 2 and 3 leaves $30/(2 \cdot 3) = 5$. With a set $Q$ of $q \geq 3$ primes there are $\binom{q}{2}$ possible pairs. We want several lengths for each pair. Multiplying its base length by different primes outside $Q$ gives such lengths. We call these primes *markers*. Can the marginals force every one of the resulting cycles?

Write $L = \prod_{a \in Q} a$. For a pair $e = \{a, b\}$ put $d_e = d_{ab} = L/(ab)$, and let $H = \sum_e d_e$, where $e$ ranges over all unordered pairs in $Q$. Choose a set $R = \{r_1 < \cdots < r_t\}$ of marker primes outside $Q$, with $t \geq 1$. Let $w_0 = 0$, $w_h = r_1 + \cdots + r_h$ and $w = w_t$. The condition $r_h > w_{h-1}$ for every $h$ says that each marker exceeds the sum of all smaller markers. Taking one cycle for each pair and each marker gives

$$X_0 = \sum_e \sum_{h=1}^{t} C_{d_e r_h}, \qquad n = |X_0| = wH.$$

In each group of cycles indexed by $e$ there are $wd_e$ states. The marker $r_h$ occurs on cycles with $r_h H$ states in total.

It helps to see both kinds of information in one example. Take $Q = \{11, 13, 17\}$ and $R = \{2, 3\}$, so $w = 5$ and $H = 41$. The three groups of $X_0$ are

| missing pair $e$ | $d_e$ | cycle lengths | states |
|---|---|---|---|
| $\{11, 13\}$ | 17 | 34, 51 | 85 |
| $\{11, 17\}$ | 13 | 26, 39 | 65 |
| $\{13, 17\}$ | 11 | 22, 33 | 55 |

We know the three group sizes because we built the table. To recover it we are given only the size, 205 states, and the marginals. For example, $85 + 65 = 150$ states miss 11, and $51 + 39 + 33 = 123$ states carry 3. Can these overlapping counts recover the table?

**Lemma 3.1** (Forcing the cycles of each missing pair)**.** *Take $P = Q \cup \{r_1, \ldots, r_t\}$ and $s = s(X_0)$. Suppose*

$$w \sum_{b \in Q \setminus \{a\}} \frac{1}{b} < 1 \qquad (a \in Q). \tag{2}$$

*Then $F(s) = n$, and $X_0$ is the only realisation of size $n$. It has $t\binom{q}{2}$ different cycle lengths.*

*Proof.* First we show that every cycle of a competing sum must miss exactly two primes. Once the cycles fall into pair groups, we can determine the size of each group and finally the cycles that fill it.

*Which primes can a cycle miss?* For $a \in Q$, let $b_a = n - s_a$ count the states of $X_0$ on cycles whose lengths are not divisible by $a$. By construction,

$$b_a = w \sum_{b \neq a} d_{ab} = \frac{L}{a}\, w \sum_{b \neq a} \frac{1}{b} < \frac{L}{a}.$$

Each state of $X_0$ misses two primes, so $\sum_{a\in Q} b_a = 2n$. Summing $b_a < L/a$ gives $2n < L\rho$, where $\rho = \sum_{a\in Q} 1/a$. The inequality for $a = \max Q$ in (2) gives $\rho < 1/w + 1/\max Q < 2$. Thus $n < L$.

Take any realisation $Y$ of these marginals with $N = |Y| \le n$. A cycle containing all primes of $Q$ would use at least $L > N$ states. A cycle missing only $a$ has length divisible by $L/a$. None of its states would be counted by $s_a$, but only $N - s_a \le n - s_a = b_a < L/a$ states of $Y$ are missing from this count. Thus each state of $Y$ is counted in at most $q-2$ of the marginals indexed by $Q$. In $X_0$ every state is counted exactly $q-2$ times. Hence

$$(q-2)n = \sum_{a\in Q} s_a \le (q-2)N.$$

Since $q \ge 3$, this forces $N = n$ and equality at each state. Every cycle of $Y$ therefore misses exactly two primes of $Q$.

*How large is each group?* A cycle missing exactly $a, b$ has length divisible by $d_{ab}$. Write $d_{ab}z_{ab}$ for the total size of this group in $Y$. Since $N = n$, exactly $n - s_a = b_a$ states of $Y$ miss $a$. Hence

$$\sum_{c\neq a} d_{ac}z_{ac} = b_a = \sum_{c\neq a} d_{ac}w. \tag{3}$$

This is where the classical one-row result applies. With $H_a = L/a$, the coefficients $d_{ac} = H_a/c$ are complementary products. A row total below $H_a$ has at most one nonnegative representation by these numbers [17, arXiv version, Proposition 1(a) and Example 12]. Remark A.2 gives the semigroup interpretation. Here the congruence can be seen directly.

In the example, the row for 11 reads $17z_{11,13} + 13z_{11,17} = 150$. Modulo 13 it gives $z_{11,13} \equiv 5 \pmod{13}$, while $17z_{11,13} \le 150$ leaves just $z_{11,13} = 5$.

In general, reduction of (3) modulo $b \neq a$ isolates $d_{ab}z_{ab}$ because every other coefficient is divisible by $b$, and $d_{ab}$ is invertible modulo $b$. Hence $z_{ab} \equiv w \pmod{b}$. Both $z_{ab}$ and $w$ lie in $[0, b)$, because multiplying either by $d_{ab}$ gives at most $b_a < L/a = b\,d_{ab}$. Thus $z_{ab} = w$. This reasoning also recovers a different prescribed value for each group whenever the corresponding row totals still satisfy $b_a < L/a$. We will use that observation again when every marginal is required to be positive.

*Which cycles can fill a group?* Return to the example. Each group now has size five in its own units. Its cycles bearing 3 use a multiple of three units. Six would exceed five, so their contribution is zero or three. The observed total $123 = 3(17 + 13 + 11)$ requires all three groups to contribute three. Each group thus contains its cycle marked 3. Two units remain in each group. The observed 2-marginal is $82 = 2(17 + 13 + 11)$, so every group must use both units on its cycle marked 2. Two unmarked copies would give too small a marginal.

Measure the size of group $e$ in units of $d_e$. Its total is $w$. Cycles bearing the largest marker $r_t$ contribute a multiple of $r_t$. As $w = w_{t-1} + r_t < 2r_t$, that contribution is either 0 or $r_t$. Summing the maximal contributions over all groups gives $\sum_e d_e r_t = s_{r_t}$. Every group must therefore attain its maximum. It contains exactly one cycle of length $d_e r_t$. Removing this cycle from each group leaves normalised size $w_{t-1}$ in each and leaves all smaller marker marginals intact. Descending through the markers recovers every cycle of $X_0$. The missing pair and the marker can both be read from the prime factors of a length, so all $t\binom{q}{2}$ lengths differ. □

With no markers, taking one copy of $C_{d_e}$ for each pair already forces $\binom{q}{2}$ lengths under (2) with $w = 1$. The first two parts of the proof give each group normalised size one. The ten-state example has $Q = \{2, 3, 5\}$ and $w = 1$ in this unmarked case. Even the largest row sum, $1/2 + 1/3$ at $a = 5$, is below one, so these two steps already force $C_2 + C_3 + C_5$. Appendix A explains both the need for the strict marker inequality and the connection between one row and numerical semigroups.

### 3.2 The unique quotient for the first $k$ primes

*Proof of Theorem 1.1.* The lemma gives $t\binom{q}{2}$ lengths. To make the pair count large, we want $q \sim k$. To fit many markers into each pair, we need room for their sum $w$ in the row inequalities. Each row uses at most $\rho = \sum_{a \in Q} 1/a$ in reciprocal weights, so $w\rho < 1$ suffices. We can achieve both goals by reserving the smaller primes for markers.

Take $j_0 = \lceil k/\log\log k \rceil$ and $Q = \{p_{j_0}, \dots, p_k\}$. This discards only $o(k)$ primes from $Q$. Since $p_j \sim j \log j$ by the prime number theorem,

$$\rho = \sum_{a \in Q} \frac{1}{a} \le \frac{q}{p_{j_0}} = O\Big(\frac{\log\log k}{\log k}\Big).$$

Choose a budget $W = \lfloor \log k/(\log\log k)^2 \rfloor$ for the sum of the markers. Then $W \to \infty$ and $W\rho = O(1/\log\log k)$. Start with 2 and repeatedly take the least prime larger than the running sum. The first markers are $2, 3, 7, 13$, with sums $2, 5, 12, 25$. Stop before the next sum would exceed $W$. Each chosen prime is at most $W < p_{j_0}$ for large $k$, so the two prime sets are disjoint. In the infinite greedy sequence the least prime above $w_h$ is $(1 + o(1))w_h$ by the prime number theorem. Hence $w_{h+1}/w_h \to 2$. Averaging the logarithms of these ratios gives $\log w_h/h \to \log 2$. The stopping rule $w_t \le W < w_{t+1}$ now gives $t \sim \log_2 W \sim \log_2 \log k$.

Apply Lemma 3.1 to $Q \cup R$. Use the resulting $X_0$ unchanged over all of $P_k$. The unused primes $u$ have $s_u = 0$, so no cycle of a competing realisation can contain them. The lemma therefore still gives $X_0$ as the unique minimum, of size $n = wH$.

Set $B_k = A_k X_0$. Any quotient of $B_k$ by $A_k$ has size $n$ and, by Lemma 2.1, the marginals of $X_0$. It must therefore equal $X_0$ up to isomorphism. To count the lengths of $B_k$, use (1). For $a \in Q$, a pair containing $a$ gives a cycle missing $a$. For $r \in R$, a cycle carrying a different marker misses $r$, since $t \ge 2$ for large $k$. For an unused $u$ we have $n - s_u = n > 0$. Thus $n - s_p > 0$ for every $p$, and $\sum_p s_p/p > 0$ because some prime occurs. The target has exactly the $k + 1$ lengths $M$ and $M/p$.

Each length of $X_0$ omits two primes of $Q$; a target length omits at most one. The two length sets are disjoint. Finally

$$|\operatorname{Len}(X_0)| = t\binom{q}{2} \sim \tfrac{1}{2}k^2 \log_2 \log k.$$

□

### 3.3 Can every prime occur in the quotient?

The division theorem is now proved. Some primes in its construction have zero marginal, which immediately rules out any length divisible by them. Is that shortcut essential? We can require each prime to occur on some cycles and to be absent from others. Keeping the missing pairs will also let us compare this example with the upper bound in Corollary 5.2.

**Remark 3.2** (extending the prime set)**.** Suppose $Q \cup R \subseteq P_k$, where $R = \{r_1, \dots, r_t\}$. One way to incorporate the unused primes $U = P_k \setminus (Q \cup R)$ is to multiply every length by $D = \prod_{u \in U} u$. The resulting witness has size $Dn$, original prime marginals $Ds_p$, and marginal $Dn$ for every $u \in U$. If $U$ is nonempty, these last marginals force every realisation of size at most $Dn$ to have exactly $Dn$ states, all on cycles with lengths divisible by $D$. Dividing the lengths by $D$ recovers the original instance. Thus the lifted optimum is unique and its support is unchanged. For $U = \varnothing$ this is the identity. The cost of this simple construction is that the marginals for $u \in U$ are saturated.

**Theorem 3.3** (Positive marginals in every coordinate)**.** *For all sufficiently large $k$ there is a marginal vector $s$ on $P_k$ with entries below $M$ and a unique minimum-size realisation $X$. Each $p \in P_k$ occurs on some cycles of $X$ and is absent from others, so $0 < s_p < F(s)$. For some $Q \subseteq P_k$, every cycle of $X$ misses exactly two primes of $Q$. Moreover,*

$$|\operatorname{Len}(X)| \geq (\tfrac{1}{2} - o(1))k^2 \log_2 \log k.$$

*Proof.* Keep $Q, R, w$ and the parameter choices from the proof of Theorem 1.1. It remains to introduce the unused primes without saturating their marginals or losing uniqueness.

*Where do the unused primes go?* Let $U = P_k \setminus (Q \cup R)$. Its cardinality is $o(k)$ and all its primes are below $p_{j_0} = o(p_k)$. A positive proportion of the $k$ primes exceed $0.8p_k$, again by the prime number theorem. Assign to each $u \in U$ a pair $e(u)$ of such primes, with disjoint endpoints for different $u$. Large endpoints give a small base length $d_{e(u)}$. We will use this to make the entire $u$-marginal fit just one cycle. Set

$$X_1 = X_0 + \sum_{u \in U} C_{u d_{e(u)}}, \qquad n = |X_1|.$$

Let $z_e^0$ be the normalised size of group $e$ in $X_1$. It equals $w + u$ on a chosen pair $e(u)$ and $w$ elsewhere. Each row contains at most one chosen pair. Thus, writing $u_* = \max(U \cup \{0\})$,

$$\frac{ab_a}{L} = \sum_{b \neq a} \frac{z_{ab}^0}{b} \leq w\rho + \frac{u_*}{0.8p_k} = o(1), \qquad b_a = n - s_a(X_1).$$

In particular $b_a < L/a$ for every row and $n/L \leq (\rho/2) \max_a(ab_a/L) < 1$ for large $k$. By the first two parts of the lemma, any realisation of the marginals of $X_1$ of size at most $n$ has size exactly $n$. Every cycle in it misses a pair, and every group has its prescribed size $d_e z_e^0$.

*Recovering the additional cycles.* Let $d_{\min} = \min_f d_f$. Our choice of endpoints gives

$$d_{e(u)} \leq \frac{L}{(0.8p_k)^2} < \frac{2L}{p_k^2} \leq 2d_{\min}.$$

In a competing optimum, write $x_{f,m}$ for the multiplicity of length $d_f m$, where $m$ is a product of primes outside $Q$. The $u$-marginal of $X_1$ is $u d_{e(u)}$. Dividing by $u$ gives

$$d_{e(u)} = \sum_{\substack{f,m \\ u \mid m}} d_f (m/u) x_{f,m}.$$

Every positive term costs at least $d_{\min}$, so there is just one. Its integer coefficient $(m/u)x_{f,m}$ must also be one. Hence $m = u$, $x_{f,m} = 1$, and $d_f = d_{e(u)}$ forces $f = e(u)$. This identifies the cycle carrying each $u$, including its group. Removing all these cycles restores size $w$ in each group in normalised units. The marker argument from the lemma then recovers $X_0$.

Every prime of $Q$ is present on cycles from some pairs and missing from cycles of other pairs. Each marker occurs on its own cycles and is absent from cycles bearing another marker, since $t \geq 2$ for large $k$. Each unused prime was assigned to exactly one cycle. Thus $0 < s_p < n < L \leq M$ for every $p$. The divisibility $p \mid s_p$ holds by definition. Finally the unique optimum uses

$$t\binom{q}{2} + |U| = (\tfrac{1}{2} + o(1))k^2 \log_2 \log k$$

different lengths. The quadratic factor counts the pairs. The extra factor counts the markers that the marginals can distinguish within each pair. □

**Remark 3.4** (finite sizes)**.** The asymptotic factor grows slowly. One concrete instance with all marginals strictly below $F(s)$ uses the first 271 primes. Take $R = \{2\}$, $Q = \{p_{27}, \dots, p_{271}\}$, and assign the remaining 25 primes to disjoint pairs from the 50 largest primes of $Q$, pairing largest with smallest. The unique optimum has $\binom{245}{2} + 25 = 29\,915$ lengths. The single marker is absent from the 25 private cycles. Appendix A verifies the numerical conditions, gives a sufficient threshold for the asymptotic choice of parameters, and compares finite choices of markers.

# 4 How large could the support be in general?

Our construction gives a unique minimum-size realisation. Other marginals can have several such realisations. How many lengths does one really need if one may choose among them?

For the primes $2, 3, 5, 7$, consider $s = (20, 9, 20, 14)$. Both

$$3C_3 + 2C_7 + 2C_{10} \qquad \text{and} \qquad C_3 + 4C_5 + C_6 + C_{14}$$

have these marginals and have 43 states. This is the minimum. A cycle whose length contains two of $3, 5, 7$ would contribute at least their product to both marginals, exceeding the smaller one. The states counted by these three primes are therefore disjoint and already number $9 + 20 + 14 = 43$. We need at least three distinct lengths, one for each prime. The first sum achieves three and the second uses four.

Write $r(s)$ for the fewest lengths among realisations of size $F(s)$. For the worst case over $P_k$, call $s$ *admissible* when $p \mid s_p$ and $0 \le s_p < M$ for every $p$. The bound $s_p < M$ makes the following maximum finite. Lemma 2.1 also applies to larger marginals. In symbols,

$$r(s) = \min_{\substack{s(X)=s \\ |X|=F(s)}} |\operatorname{Len}(X)|, \qquad m(k) = \max_{s \text{ admissible}} r(s).$$

Theorem 3.3 supplies the lower bound below. Exchanging cycles in an optimum will give the upper bound.

$$m(k) \ge (\tfrac{1}{2} - o(1))k^2 \log_2 \log k, \qquad m(k) = O(k^2 \log k).$$

Section 5 sharpens the upper bound for missing pairs.

**Lemma 4.1** (Distinct subsets give distinct marginal vectors)**.** *Fix any finite nonempty prime set $P$. Let $X$ minimise the number of distinct lengths among the minimum-size realisations of $s$. For each of its $r$ distinct lengths, take one copy of that cycle. The $2^r$ subsets of these cycles have different marginal vectors. In particular, with $\sigma$ the sum-of-divisors function,*

$$r \le \sum_{p \in P} \log_2(1 + \min\{s_p/p,\, \sigma(M_P/p)\}) = O(|P| \log M_P).$$

*Proof.* Any subset is a minimum-size realisation of its own marginals. Replacing it by a smaller realisation would improve $X$. Suppose two distinct subsets have the same marginals. Their sizes must then agree. Cancel the common cycles. Each side still has a cycle, since each remaining length contains a prime. Replace one side by the other as many times as the multiplicities in $X$ allow. Each replacement preserves size and marginals and uses only lengths already present. When one length is exhausted we have an optimum with fewer distinct lengths, a contradiction.

Divide the $p$-coordinate of a subset sum by $p$. It is an integer between 0 and $s_p/p$. It is also at most $\sigma(M_P/p)$: each possible contribution $d/p$ is a different divisor of $M_P/p$, used at most once. Counting the possible coordinate vectors gives the displayed bound. For $P = P_k$, $\log M = \sum_{j \le k} \log p_j \sim k \log k$ by the prime number theorem, giving the asserted upper bound on $m(k)$. □

The subset-sum exchange appears in Eisenbrand and Shmonin [16, Lemma 3]. Their Corollary 5(i) gives a sparse optimum by adding a coordinate for the objective value. Aliev et al. [3, proof of Theorem 1] use optimality directly: when both signs of an exchange are feasible, its cost must be zero. Here the divisor structure gives the additional coordinate bound $\sigma(M_P/p)$, since the subsets use each length at most once.

In integer programming, each nonempty subset $S$ of the primes gives a column. Its cycle has length $d_S = \prod_{p\in S} p$ and contributes $d_S$ states to every marginal indexed by $S$. Including size as the last coordinate gives the column $d_S(\mathbf{1}_S, 1)$, where $\mathbf{1}_S$ is the incidence vector of $S$. All $2^{|P|} - 1$ nontrivial lengths are available. The minimum number of columns used to reach a target $b$ is often written $m_0(b)$. Its maximum over targets is an integer Carathéodory rank [4, 1, 2]. Our targets have the particular form $(s, F(s))$. In numerical semigroups the analogous count of distinct generators is called the 0-length [10]. General bounds for integer matrices appear in Aliev et al. [3] and Berndt et al. [7]. Jansen et al. [18] force many distinct configurations in bin packing. In our setting the entire matrix is prescribed by the primes before the marginals are chosen.

### Which lengths would a search have to consider?

A different count arises when we use the lengths in the input and the size of the quotient to list possible lengths $d$. In the support test of Bridoux et al., $d$ is a candidate when it fits in the quotient and its products with the divisor's cycles all have lengths present in the target. The candidates absent from the target form the *principal support*. Specifying the multiplicity of each of these lengths determines at most one quotient [8, Lemma 11].

For the family of Theorem 1.1, write $n = |X_k|$. The candidate set is exactly $\{d : d \mid M,\ d \le n\}$. Indeed, for $d \mid M$, each $\operatorname{lcm}(M/p, d)$ is either $M/p$ or $M$, and both occur in $B_k$. Conversely any candidate divides one of these product lengths and thus divides $M$. For large $k$ the shortest target length $M/p_k$ exceeds $n$. To see this, recall that the construction takes $Q = \{p_{j_0}, \dots, p_k\}$ and $n < L = \prod_{p\in Q} p$. Since $j_0 = \lceil k/\log\log k \rceil$ and $p_k \sim k \log k$, we have $M/L \ge 2^{j_0-1} > p_k$. Thus every candidate belongs to the principal support. Fix a missing pair $a, b \in Q$. Its marked cycle of length $r_1 L/(ab)$ fits in the quotient, so every divisor of $L/(ab)$ fits as well. This gives at least $2^{|Q|-2}$ candidates out of at most $2^k$ divisors of $M$.

In short, $2^{\Theta(k)}$ lengths pass this test, and the only quotient uses $\sim \frac{1}{2}k^2 \log_2 \log k$ of them. In terms of the number of input states $N = |A_k| + |B_k|$, the latter count is $(\log N)^{2+o(1)}$. Indeed, $N = |A_k|(1 + n)$, $M/2 \le |A_k| \le kM$ and $n < M$, so $\log N = \Theta(\log M) = \Theta(k \log k)$. Even the candidate count is $N^{o(1)}$. Finding compatible multiplicities is a separate algorithmic question.

## 5 How sharp is the missing-pair construction?

In the construction of Theorem 3.3, each length misses a pair of primes from a common set $Q$. Could one force substantially more lengths by adding more markers to each pair? Under the same inequalities on the missing mass, the answer is no (Corollary 5.2). To see why, group the lengths by their missing pair. Members of a group share a large divisor. After removing it, we can record a subset of lengths by a few small sums. If two subsets give the same sums, an exchange reduces the support.

### Exchanges within a partition of the lengths

We give the argument for any partition. For a group $g$, choose a positive common divisor $D_g$ of its lengths and put

$$w_g = \sum_{d \in g} \frac{d}{D_g}.$$

Each distinct length occurs once in this sum. For example, if a group contains $3C_6 + 2C_{30}$, taking $D_g = 6$ gives $w_g = 1 + 5 = 6$, while its actual number of states divided by $D_g$ is $3 + 2 \cdot 5 = 13$. The first quantity bounds the sum of one copy of each selected length after division by $D_g$. The second is the normalised mass that an exchange must preserve.

**Theorem 5.1** (Grouping criterion)**.** *Let $X$ be a minimum-size realisation of nonzero marginals $s$, and write $n = F(s)$. Partition* $\mathrm{Len}(X)$ *into $G$ nonempty groups, with common divisors $D_g$ as above, and set $W = \sum_g w_g$. There is an optimum $X'$ with* $\mathrm{Len}(X') \subseteq \mathrm{Len}(X)$ *such that for every real $T \geq 2$,*

$$|\mathrm{Len}(X')| \leq G \log_2\!\left(1 + \frac{W}{G}\right) + \pi(T) \log_2(n+1) + \sqrt{2}\,\frac{W}{\sqrt{T}}.$$

*Here $\pi(T)$ is the number of primes at most $T$.*

The first term accounts for one sum per group with $w_g \leq T$. The second allows one marginal for each prime up to $T$. The last bounds all the lengths in the remaining groups.

*Proof.* Keep the original set of available lengths, the prescribed marginals and the number of states in each group. Among all realisations with these data, choose one, $Y$, with the fewest distinct lengths. Its size is still $n$, so $Y$ is an optimum. We will use the same $Y$ for every $T$.

First consider a group with $w_g > T$. If $Y$ uses $h$ lengths in this group, their quotients $d/D_g$ are distinct positive integers, so $h(h+1)/2 \leq w_g$. The total number of lengths used in these groups is at most $\sum_{w_g > T} \sqrt{2w_g} \leq \sqrt{2}\, W/\sqrt{T}$.

It remains to count the lengths in the groups with $w_g \leq T$. Take a subset $E$ of these lengths, one copy of each. Record the normalised sum $\sum_{d \in E \cap g} d/D_g$ for each such group $g$ and the marginal $\sum_{d \in E:\, p \mid d} d$ for each prime $p \leq T$. These data determine the subset's other marginals as well. Indeed, if $p > T$ divides a length $d = D_g m$ in one of the groups, then $m \leq w_g \leq T$, so $p \mid D_g$. Every state counted in that group contributes to the $p$-marginal. The recorded group sum determines that contribution.

If two different subsets had the same data, their difference would give an exchange with coefficients in $\{-1, 0, 1\}$ preserving every group's mass and every marginal. Since all lengths are positive and the total mass is preserved, the exchange has both signs. Repeat it until the first length being subtracted disappears. This produces a realisation with fewer lengths among the same choices, contradicting the definition of $Y$.

If these groups contain $h_0$ lengths of $Y$, the $2^{h_0}$ subsets therefore give different records. There are at most

$$(n+1)^{\pi(T)} \prod_{g:\, w_g \leq T} (1 + w_g)$$

records. Take base-two logarithms and use concavity to bound $\sum_g \log_2(1 + w_g) \leq G \log_2(1 + W/G)$. Adding the count for the groups with $w_g > T$ proves the theorem. □

### The scale of the lower bound

**Corollary 5.2** (The missing-pair class)**.** *Work with the first $k$ primes $P_k$. Suppose some optimum $X$ of size $n = F(s)$ has the following property. There is a set $Q \subseteq P_k$, with $q = |Q| \geq 3$ and $L = \prod_{a\in Q} a$, such that each cycle length of $X$ misses exactly two primes of $Q$, and $n - s_a < L/a$ for every $a \in Q$. Then the marginals $s$ have an optimum with at most*

$$\binom{k}{2} \log_2\left(1 + \frac{p_k}{k-1}\right) + O(k^2) = \tfrac{1}{2}k^2 \log_2 \log k + O(k^2)$$

*distinct lengths. The largest minimum support over the admissible instances in this class is $(\frac{1}{2} + o(1))k^2 \log_2 \log k$.*

*Proof.* The zero vector has the empty optimum, so take $s \neq 0$. For the pair $e = \{a, b\}$ use the common divisor $d_e = L/(ab)$, and write $d_e z_e$ for the number of states in the group. Put $z_e = 0$ for an empty group. The states missing $a$ belong to exactly the groups $\{a, b\}$. Counting them gives

$$\sum_{b\in Q\setminus\{a\}} \frac{z_{\{a,b\}}}{b} = \frac{n - s_a}{L/a} < 1.$$

Since $b \leq p_k$, the sum of the $z_{\{a,b\}}$ in each row is less than $p_k$. Summing the rows counts each pair twice. It follows that $W \leq \sum_e z_e < qp_k/2$, since $w_e$ counts each used length once and $z_e$ counts it with its multiplicity. Spread over all $\binom{k}{2}$ possible pairs, assigning zero to unused pairs, this gives average normalised mass below $p_k/(k-1) = O(\log k)$. The logarithm in the grouping bound is where $\log \log k$ enters.

There are at most $\binom{k}{2}$ nonempty groups. The function $G \log(1 + W/G)$ increases in both arguments (differentiate in $G$ to get $\log(1 + x) - x/(1 + x) \geq 0$, where $x = W/G$). These bounds on $W$ and $G$ give the displayed main term in Theorem 5.1. For the error terms take $T = (\log(k + 2))^2$. Every state misses two primes of $Q$, so $2n = \sum_{a\in Q}(n - s_a) < L \sum_{a\in Q} 1/a$. In particular $\log(n+1) = O(k \log k)$. Thus $\pi(T) \log_2(n+1) = o(k^2)$, and $W/\sqrt{T} = O(k^2)$ since $p_k = O(k \log k)$. Finally $p_k \sim k \log k$ gives the stated asymptotic upper bound. The unique optimum of Theorem 3.3 belongs to this class, including the inequalities on each $n - s_a$. It supplies the matching lower bound. □

### Grouping by the large prime factors

Another natural choice is to put two lengths in the same group if they have the same prime factors above a threshold $y$. Once the group mass is known, every prime above $y$ is accounted for.

**Proposition 5.3** (Large-prime patterns)**.** *Let $P$ be a finite set of primes, $M_P = \prod_{p\in P} p$, and $n = F(s)$. For $y \geq 2$ put*

$$D_y(d) = \prod_{\substack{p\mid d\\ p>y}} p, \qquad K_y = \prod_{\substack{p\in P\\ p\leq y}} p, \qquad H = \min\{n, \sigma(M_P) - 1\},$$

*where $\sigma$ denotes the sum of positive divisors. If some optimum has $G_y$ distinct values of $D_y(d)$ on its lengths, then the same marginals have an optimum with at most*

$$G_y \log_2(1 + \sigma(K_y)) + \pi(y) \log_2(1 + H)$$

*distinct lengths. In particular, for $P = P_k$ and $y = \log \log k$, the condition $G_y = O(k^2)$ gives $O(k^2 \log \log k)$ lengths.*

*Proof.* Group the lengths by $D_y(d)$. In any group the normalised lengths $d/D_y(d)$ are distinct divisors of $K_y$, so the sum of any subset lies between 0 and $\sigma(K_y)$. The subset has at most $n$ states and uses each divisor $d > 1$ of $M_P$ at most once. Hence its marginal at each $p \le y$ lies between 0 and $H$. For $p > y$, either every length of a given group is divisible by $p$, or none is. The exchange argument of Theorem 5.1 therefore counts all the lengths using these group sums and small-prime marginals alone.

We have $\log \sigma(K_y) \le \sum_{p \le y} \log(1+p) = O(y)$. For the first $k$ primes, $\log(1+H) \le \log \sigma(M_{P_k}) = \sum_{p \in P_k} \log(1+p) = O(k \log k)$, uniformly in the marginal sizes. Taking $y = \log\log k$ for sufficiently large $k$ gives $\pi(y)\log_2(1+H) = o(k^2)$. This proves the final assertion. □

For example, an optimum has $G_y = O(k^2)$ if every cycle misses at most two primes above $y$. The same conclusion holds for missing triples if every pair occurs in only $O(1)$ distinct triples. Neither example needs a bound on $n - s_a$.

The distinction between the two ways of grouping matters already in Theorem 3.3. Take $y = \log\log k$. All primes in $Q$ exceed $y$. For every marker $r_h > y$ the large-prime part of $d_e r_h$ is the full length $d_e r_h$. The superincreasing markers leave at most $1 + \log_2 y$ markers at most $y$, giving $\binom{q}{2}(t - O(\log y)) = (\frac{1}{2} - o(1))k^2 \log_2 \log k$ different large-prime parts. Grouping by the missing pair alone uses just $\binom{q}{2}$ groups and has $W = O(k^2 \log k)$.

## How many lengths can one pair carry?

**Proposition 5.4** (One pair)**.** *In Corollary 5.2, fix one witnessing optimum $X$ and the normalised masses $z_e$ of its pair groups. Consider the realisations with these marginals and masses in which each length misses exactly two primes of $Q$. Every such realisation with smallest total support uses*

$$O\left(1 + \sqrt{\frac{z_e}{\log(2+z_e)}}\right)$$

*distinct lengths in each group $e$. Examples attain this order in one group.*

The restriction to these lengths fixes the total size at $\sum_e d_e z_e = F(s)$. Allowing other lengths could add states outside the groups while preserving the same $z_e$.

*Proof.* Choose a support-minimising member $Y$ of the stated family. Such a member exists because the family contains $X$ and has fixed total size. In one group write $d = d_e m$ and $z = z_e$. The sum of the positive integers $m$, with multiplicities, is $z$, which settles the claim for bounded $z$. For large $z$, split the lengths at a threshold $U$. Each $m > U$ uses more than $U$ units of mass, so there are at most $z/U$ such lengths. For $m \le U$ we use the prime marginals to count subsets. This gives at most $\log_2(1+z) + O(U/\log(2+U))$ lengths, as we show below. To balance the two terms that depend on $U$, take $U = \sqrt{z\log(2+z)}$.

Let $\mathcal{M}$ be the distinct used values $m \le U$. To distinguish subsets $\mathcal{S} \subseteq \mathcal{M}$, record

$$\sum_{m \in \mathcal{S}} m, \qquad \left( \sum_{\substack{m \in \mathcal{S} \\ p \mid m}} \frac{m}{p} \right)_{p \le U,\ p \text{ prime}} .$$

The first coordinate preserves the group mass. For a prime $p \mid d_e$ this already preserves its marginal. For $p \nmid d_e$ and $p \le U$, the corresponding coordinate times $pd_e$ gives the marginal of the chosen subset. A prime larger than $U$ cannot divide any $m$ in this subset. Equal records would therefore allow us to exchange copies in this group up to the first disappearing length, as in Theorem 5.1. This would reduce the total support of $Y$.

There are at most $1+z$ choices of the first coordinate. For each $p \le U$, the distinct possible values $m/p$ belong to $\{1,\dots,j_p\}$, where $j_p = \lfloor U/p \rfloor$. Their subset sums are at most $j_p(j_p+1)/2$. Taking logarithms gives

$$|\mathcal{M}| \le \log_2(1+z) + 2\sum_{p\le U} \log_2(1+U/p).$$

For $U \ge 2$ the last sum, up to the choice of logarithm base, is bounded by

$$\pi(U)\log 2 + \sum_{p\le U} \log(U/p) = \pi(U)\log 2 + \int_2^U \frac{\pi(t)}{t}\,dt = O\Big(\frac{U}{\log(2+U)}\Big).$$

The integral identity follows by summing $\int_p^U dt/t$. For the estimate, split at $\sqrt{U}$ and use $\pi(t) = O(t/\log t)$. Adding $z/U$ for the large values of $m$ proves the bound in every group of the same $Y$.

*Sharpness.* The scale in the bound suggests using many distinct primes of comparable size. Let $R$ be the primes in $(x, 2x]$, for $x \ge 3$, and put $w = 2\sum_{r\in R} r$. Choose distinct primes $a, b, c > 4w$ and take $Q = \{a, b, c\}$. Work over an initial prime set containing $\{2\} \cup R \cup Q$, with all unused marginals zero. Set

$$X_* = \sum_{r\in R} C_{2ra} + C_b + C_c, \qquad n = aw + b + c.$$

Its $Q$-marginals are $aw, b, c$, respectively. A length containing two primes of $Q$ would already exceed one of these marginals. Thus every realisation has size at least their sum $n$. The witness attains equality, forcing every state of every optimum to be counted by exactly one prime of $Q$. The three normalised group masses are $w, 1, 1$, and the missing masses satisfy $b + c < bc,\ aw + c < ac,\ aw + b < ab$. The marginals are admissible since $n < abc$.

The unit groups force $C_b$ and $C_c$. As $s_2 = aw$, every state in the remaining group lies on an even cycle. For each $r \in R$ we have $s_r = 2ra$, and each cycle contributing to $s_r$ has length divisible by $2ra$. This forces exactly one copy of $C_{2ra}$. These cycles exhaust the remaining group, proving uniqueness. Its $|R|$ distinct lengths have normalised mass $w$. The prime number theorem gives $|R| = \Theta(x/\log x)$ and $w = \Theta(x^2/\log x)$, so $|R| = \Theta(\sqrt{w/\log w})$. The corresponding estimates for sums of primes are classical [6, Theorem 1.4]. □

## 6 What forces a large support?

A cycle of length $ij$, for distinct primes $i, j$, contributes $ij$ states to both marginals. We could obtain the same marginals with $j$ copies of $C_i$ and $i$ copies of $C_j$, at a cost of $2ij$ states. This gives two ways to vary the optimisation problem. How many lengths suffice if we drop the size requirement? What if we keep the prescribed size but allow negative multiplicities? If signed copies of $C_1$ are available, their coefficient can adjust the size by any integer while leaving every marginal unchanged. We ask how short a fixed list of lengths greater than one will suffice.

**Proposition 6.1.** *Let $P = P_k$, $k \ge 4$, and let $s$ be admissible. Write $S = \sum_{p\in P} s_p$.*

*(a) For every integer $n$ there is a formal sum with signed integer multiplicities, marginals $s$ and total size $n$, supported on $C_{p_1}, \dots, C_{p_k}, C_6, C_{35}$.*

*(b) With nonnegative multiplicities the sum $\sum_p (s_p/p)C_p$ realises $s$ on at most $k$ lengths and has size $S$.*

*Proof.* The prime-length sum in (b) is our starting point for (a). The trade above preserves the marginals and reduces the size by $ij$ each time we introduce one $ij$-cycle. Since $\gcd(6, 35) = 1$,

choose integers $z, w$ with $6z + 35w = S - n$. Perform the trade for $2, 3$ with coefficient $z$ and for $5, 7$ with coefficient $w$. This leaves multiplicities

$$s_2/2 - 3z, \quad s_3/3 - 2z, \quad s_5/5 - 7w, \quad s_7/7 - 5w$$

on the first four prime lengths and $s_p/p$ on the others. The total size is $S - 6z - 35w = n$. □

**Why $k + 2$ fixed nontrivial lengths are needed when $k \geq 4$.** Choose $k + 1$ nontrivial lengths $d_j$, with prime sets $S_j$. Divide each marginal equation for $p$ by $p$ and keep the size equation. The integer matrix $V$ for these $k + 1$ equations has

$$\det V = \frac{\prod_j d_j}{M} \det \begin{pmatrix} \mathbf{1}_{S_1} & \cdots & \mathbf{1}_{S_{k+1}} \\ 1 & \cdots & 1 \end{pmatrix}.$$

If $V$ is nonsingular, every prime appears. Since there are $k + 1$ nonempty sets, some prime appears twice. It divides $\prod_j d_j/M$, so $|\det V| > 1$. Thus in every case the columns span a proper sublattice of $\mathbb{Z}^{k+1}$. To cover all targets one would already have to represent the normalised unit vectors $(\mathbf{e}_i, 0)$ and $(\mathbf{0}, 1)$, whose marginals are admissible for $k \geq 2$. Together with Proposition 6.1, this proves the optimality of $k + 2$ for a selection fixed in advance when $k \geq 4$.

**Fewer primes and fixed points.** For $k = 3$, all six singleton and pair lengths suffice because $\gcd(6, 10, 15) = 1$. The target $s = (0, 0, 0), n = 1$ requires six lengths even if they may be chosen after seeing the target. Indeed $S - n = -1 = 6x_6 + 10x_{10} + 15x_{15} + 60x_{30}$, so reduction modulo $5, 3, 2$ forces each of $x_6, x_{10}, x_{15}$ to be nonzero. In $n - s_3 - s_5$ all surviving terms other than $2x_2$ are divisible by 15, so $2x_2 \equiv 1 \pmod{15}$. The other two singletons are forced by $3x_3 \equiv 1 \pmod{10}$ and $5x_5 \equiv 1 \pmod{6}$. In the order $2, 3, 5, 6, 10, 15, 30$ a witness is $(-7, 7, -1, -1, 2, -1, 0)$. For $k = 2$, the only available pair is 6, so the signed size equation has a solution precisely when $n \equiv S \pmod{6}$. If signed fixed points are allowed, $x_p = s_p/p$ and $x_1 = n - S$ give a solution for every $k \geq 1$.

There is also a comparison with general lattice bounds. For $k \geq 3$ all full-rank minors of the unnormalised matrix with columns $d_S(\mathbf{1}_S, 1)$ are divisible by $M$, since the entries in row $p$ are divisible by $p$. Taking the $k$ singleton columns and one of the pair columns $6, 10, 15$ gives minors with absolute values $6M, 10M, 15M$, so the greatest common divisor of all full-rank minors is $M$. Applying Theorem 1 of [1] with the minor $6M$ gives a signed support bound of $k + 3$. The explicit construction above gives $k + 2$ for $k \geq 4$. For comparison, see also the lattice bounds of Dubey and Liu [15]. For general nonnegative matrices the lower bounds of Berndt et al. [7] can reach order $m \log \|A\|$ for $m$ rows. The sharper estimate in Corollary 5.2 uses two properties of one optimum of size $n$. Every cycle length misses exactly two primes of the same set $Q$, and the missing mass in each row satisfies $n - s_a < L/a$ for $a \in Q$, where $L = \prod_{a \in Q} a$.

# 7 What happens with four to eight primes?

For the first four primes, every admissible vector has an optimum using at most twelve lengths, and some vectors require all twelve. For five through eight primes, each row of Table 1 certifies one vector whose unique optimum needs the stated number of lengths. The corresponding maxima remain unknown. The computational supplement contains the data and two checks. From its directory, run `python -B verify.py` to verify the five examples and `python -B all_four.py` (with NumPy) to reproduce the four-prime maximum.

| $k$ | Distinct lengths | States in the unique optimum | Consequence |
|---|---|---|---|
| 4 | 12 | 246 | $m(4) = 12$ |
| 5 | 21 | 2 051 | $m(5) \geq 21$ |
| 6 | 30 | 20 185 | $m(6) \geq 30$ |
| 7 | 43 | 343 463 | $m(7) \geq 43$ |
| 8 | 58 | 6 526 232 | $m(8) \geq 58$ |

Table 1: Each row describes the unique minimum-size cycle sum for one admissible marginal vector. Only the four-prime search covered all admissible vectors.

**Why the four-prime maximum is exact.** Here $M = 210$. The coordinates $s_p/p$ range from 0 to $M/p - 1$, giving $105 \cdot 70 \cdot 42 \cdot 30 = 9\,261\,000$ vectors. A dynamic programme processes the fifteen nontrivial divisors of 210 one at a time. At each vector it minimises $16|X| + |\operatorname{Len}(X)|$. The factor 16 makes one state more expensive than all fifteen support entries together. The first copy of length $d$ costs $16d+1$, and each further copy costs $16d$. For exactly 98 vectors, the minimum support among their optima is twelve. The four-prime row uses $s = (160, 117, 165, 175)$.

**How to certify a single row.** Suppose a proposed cycle sum has $n$ states and contains no fixed point. Count all $n$-state realisations of its marginals, this time allowing fixed points. A count of one proves both optimality and uniqueness. Indeed, any smaller realisation could be padded to size $n$ with fixed points, giving a different realisation at size $n$.

For $k = 5$ and 6 the marginal vectors in the table, in increasing prime order, are respectively

$$(1008, 1332, 1110, 1218, 1947), \qquad (10242, 19101, 15275, 11963, 15136, 15678).$$

For these two rows and the displayed four-prime row, an integer checker branches on cycle multiplicities. At a branch it can discard a residual vector if an integer linear functional takes a value there that the remaining columns cannot realise. It checks each such exclusion, checks that every remaining multiplicity was tried and counts the surviving solutions. The three search trees contain respectively 17, 95 and 893 nodes in order of increasing $k$. Each has exactly one solution.

**Passing from six primes to seven and eight.** There is a shorter verification for the two largest rows. Start with the certified six-prime optimum $X$. Multiply each of its lengths by $q = 17$, keeping each multiplicity, and adjoin one copy of each of thirteen lengths coprime to 17, totalling 318 states. The new support is $30 + 13 = 43$. At the next step take $q = 19$ and adjoin fifteen lengths totalling 435 states, giving $43 + 15 = 58$. The file `data/lifts.json` in the supplement lists the added lengths for both steps.

Why can no different optimum arise? Write $H$ for the added cycles at one step and $t = |H|$. In any competing realisation padded to the proposed size, the new $q$-marginal fixes the number of states on $q$-divisible cycles. The remaining cycles therefore form a system $Y$ of size $t$. For every old prime $p$, the cycles divisible by both $p$ and $q$ contribute a multiple of $pq$ to the $p$-marginal. Consequently

$$s_p(Y) \equiv s_p(H) \pmod{pq}.$$

There are respectively 486 and 2 640 vectors in $[0, t]^{k-1}$ satisfying these congruences. Counting all realisations across those vectors gives just one, namely $H$. After removing it and dividing the remaining lengths by $q$, we recover the preceding optimum by uniqueness. The check allows fixed points in $Y$, so it also excludes smaller solutions by the padding argument.

The count in this last step uses a recursion on the old primes. For a chosen prime $p$, split the cycles into those with $p \nmid d$ and those with $p \mid d$. The first part has size $t - s_p$. Dividing the lengths of the second part by $p$ gives size $s_p/p$. Enumerate the ways to distribute the other marginals between these two parts and recurse. Each cycle sum determines its split uniquely. At the end of the recursion only fixed points remain. All counts use integer arithmetic.

For comparison, the marked-pair construction of Lemma 3.1 also gives small examples directly. With $Q = \{23, 29, 31, 37\}$ and $R = \{2, 3\}$ its unique optimum has 12 lengths and 26 750 states. With $Q = \{37, 41, 43, 47\}$ and $R = \{2, 3, 7\}$ the counts are 18 and 126 696. With $Q = \{11, 13, 17\}$ and $R = \{2, 3\}$, multiplying all lengths by $5 \cdot 7$ as in Remark 3.2 gives 6 lengths and 7 175 states.

# 8 Open problems

**Open question.** Is $m(k) = \Theta(k^2 \log\log k)$ for the first $k$ primes?

The lower bound comes from lengths missing exactly two primes of $Q$. Corollary 5.2 gives the matching leading constant within the stated pair class. One way to reach the same order for arbitrary admissible marginals would be to find an optimum whose lengths admit the grouping of Theorem 5.1 with $G = O(k^2)$ and $W = O(k^2(\log k)^C)$ for some fixed $C > 0$. Taking $T = \max\{2, (\log k)^{2C}\}$ would then give the desired upper bound. The example in Theorem 3.3 explains why simply grouping by the primes larger than $\log\log k$ need not suffice.

Can cycles missing three primes force more lengths than the pair class? For pairs, reducing row $a$ modulo $b$ leaves just the term for $\{a, b\}$. For triples, every group $\{a, b, c\}$ can contribute a term to that same congruence, one for each possible third prime $c$. Understanding that ambiguity could also decide whether $m(k)/(k^2 \log_2 \log k)$ has a limit.

There is a separate algorithmic question. Given the marginals $s$, how quickly can one compute $F(s)$? The examples proving the lower bound have large support, yet the proof recovers their optima from congruences and marker marginals. A bound on how many lengths must occur leaves room for an efficient way to find them.

# A Further details on the pair construction

## A boundary case with nine forced lengths

**Remark A.1** (why the marker inequality is strict)**.** When a marker equals the sum of earlier markers, the last part of the proof can fail. Take $Q = \{31, 41, 59, 229\}$ and markers $2, 3, 5$. The row inequalities still hold. The six bases $d_e$ split into two triples of equal sum, since

$$59(31 + 41 + 229) = 31 \cdot 41 + 31 \cdot 229 + 41 \cdot 229 = 17\,759.$$

In each group of one triple replace the three normalised lengths $2, 3, 5$ by $5, 5$. In each group of the other use $2, 2, 3, 3$. Every group still has size ten in its own units, and the equality of the two sums preserves each marker marginal. The size bound in the lemma used only the row inequalities, so this common size 355 180 is still optimal. The original witness has 18 distinct lengths. The exchange produces an optimum with nine, the minimum for these marginals, as we now verify.

Each of the six pair groups has normalised size ten. If a group bearing the marker 3 had only one length, its normalised length would both divide ten and be divisible by three. Hence every group bearing 3 uses at least two lengths. A support of at most eight would put all of the 3-marginal in at most two groups.

The six bases sum to 35 518. The total 3-marginal divided by three is therefore $35\,518 = \sum_e h_e d_e$, where $h_e$ is the normalised mass bearing the marker in group $e$, divided by three. A positive $h_e$ can only be $1, 2, 3$. Of the six bases, three are divisible by 59. For each of the other three bases, the possible residues of its positive contribution modulo 59 form one of the sets

$$\{8, 16, 24\}, \qquad \{19, 38, 57\}, \qquad \{32, 5, 37\}.$$

The target is divisible by 59. No set contains zero and no two entries from different sets sum to zero modulo 59. A sum supported on at most two groups would therefore use only bases divisible by 59. After dividing by 59, we would have to make 602 with at most two of $229, 41, 31$, using coefficients at most three. With no 229 the sum is at most 216. With one or two copies of 229 it is at most $2 \cdot 229 + 3 \cdot 41 = 581$. Three copies already exceed 602. Thus at least three groups bear the marker 3, and at least nine lengths occur. The exchange in Remark A.1 attains nine.

### What a single row determines

**Remark A.2** (the arithmetic of one row)**.** The congruence in (3) has a classical interpretation. After we know that every cycle misses a pair, row $a$ has precisely the generators $H_a/c$ for $c \in Q \setminus \{a\}$, with $H_a = L/a$. For the additive semigroup generated by these numbers, $H_a$ is the unique Betti element. A row total $u$ in the semigroup admits more than one factorisation exactly when $u - H_a$ belongs to the semigroup [17, arXiv version, Proposition 1(a) and Example 12]. Here $u = b_a < H_a$. Related complementary-product formulas appear in [5, Theorems 3.1–3.2] and [11, Proposition 5.5]. Aliev et al. use signed complementary products in a support lower bound [3, proof of Theorem 2]. In Lemma 3.1, the size and the marginal counts must first force every cycle to miss exactly two primes of $Q$. The row calculation then fixes the mass of each pair. It still leaves open which lengths share that mass. The size restriction matters even in the example with $Q = \{5, 7, 11\}$. The marginals of $C_{10} + C_{14} + C_{22}$ also occur in $23C_2 + 2C_5 + 2C_7 + 2C_{11}$, whose size is 92 instead of 46.

Eisenbrand and Shmonin [16, introduction] illustrate how a single representation can require many generators. In their example each of several coordinates must represent $2^m - 1$ with powers of two. This takes at least $m$ summands per coordinate, and a final coordinate fixes the total number of summands at the sum of these minima. Thus each coordinate uses each power $2^i$, $0 \le i < m$, once. Our size counts states. Once the pair masses are fixed, each group can contribute at most $d_e r_t$ states to the largest marker. The observed marginal is the sum of these maxima, so every group contributes its marked cycle. Applying the same argument to the remaining mass recovers all the smaller markers.

### How large must the primes be?

For the instance with $k = 271$ from Remark 3.4, the primes of $Q$ run from 103 to 1741, every private prime is at most 101, and every endpoint of a chosen pair is at least 1399. For any assignment of the private primes to these pairs, the normalised row totals satisfy

$$\max_a \frac{ab_a}{L} \le 2 \sum_{p \in Q} \frac{1}{p} - \frac{2}{1741} + \frac{101}{1399} < 0.997.$$

The rational sum can be evaluated directly. It also gives $n/L \le (\rho/2) \max_a (ab_a/L) < 1$. Finally $1399 > 0.8 \cdot 1741$ gives $d_{e(u)} < 2d_{\min}$, as in the proof of Theorem 3.3.

For comparison, one may optimise the simpler construction in Remark 3.2, in which the unused primes have saturated marginals. For each suffix $Q = \{p_{j_0}, \dots, p_k\}$ with $q \ge 3$, the largest reciprocal sum in (2) occurs in the row $a = p_k$. For each $t \ge 1$ the greedy marker

sequence has the smallest possible sum of $t$ markers. This follows inductively since both the next admissible prime and the running sum increase with the previous sum. We also allow the unmarked construction with $t = 0$ and $w = 1$. The support to maximise over feasible choices is consequently $\max(1, t)\binom{q}{2}$.

| $k$ | First index $j_0$ | Markers $t$ | Largest support in this family |
|---:|---:|---:|---:|
| 100 | 5 | 0 | 4 560 |
| 1 000 | 243 | 2 | 573 806 |
| 10 000 | 1 528 | 2 | 71 783 256 |
| 100 000 | 9 681 | 2 | 8 157 612 080 |
| 1 000 000 | 61 701 | 2 | 880 405 951 700 |

The two markers in the last four examples are $2, 3$. These are maxima over the suffix construction described above.

The asymptotic choice of parameters stops producing markers altogether when the budget $W$ falls below two. For example, $W = 1$ for every integer $k$ from 84 to 896 677. A sufficient threshold for all the estimates used in Theorem 3.3 is

$$k \geq k_0 = 11 \left\lceil \frac{\exp(\exp(11))}{11} \right\rceil .$$

Here are the estimates. Put $x = \log k$ and $y = \log x$. Then $y \geq 11$, $x > 59874$, and $y/x \leq 11/59874$. Using $m \log m < p_m < m(\log m + \log\log m)$ for $m \geq 6$ [22, (3.12)–(3.13)] gives

$$\rho \leq \frac{y}{x - \log y} < 0.000185, \qquad W\rho \leq \frac{1}{y(1 - \log y/x)} < 0.091, \qquad \frac{\max U}{p_k} < \frac{1 + y/x}{y} < 0.091.$$

The smallest of the largest $2|U|$ endpoints has index $m > 9k/11$, so $p_m/p_k > 9(x-1)/(11(x+y)) > 0.817$. The same estimates give $W < p_{j_0}$ and $W/p_k < 1/(ky^2)$. In particular, every row ratio is less than $0.091 + 0.091/0.8 < 0.207$, $n < L$, and $d_{e(u)} < 2d_{\min}$. At $k = k_0$, the mean value theorem gives $0 \leq y - 11 < 11/(\exp(\exp(11))e^{11})$. Hence $j_0 = k_0/11$ and $W = 494$. The greedy markers at this threshold are $2, 3, 7, 13, 29, 59, 127, 241$. Their sum is 481, and the next candidate, 487, exceeds the remaining budget.

## References


[1] I. Aliev, G. Averkov, J. A. De Loera and T. Oertel. Sparse representation of vectors in lattices and semigroups. *Math. Program.* 192 (2022), 519–546.

[2] I. Aliev, M. Celaya and M. Henk. Sparsity and proximity transference in integer programming. *Math. Program.* 216 (2026), 29–48.

[3] I. Aliev, J. A. De Loera, F. Eisenbrand, T. Oertel and R. Weismantel. The support of integer optimal solutions. *SIAM J. Optim.* 28 (2018), 2152–2157.

[4] I. Aliev, J. A. De Loera, T. Oertel and C. O'Neill. Sparse solutions of linear Diophantine equations. *SIAM J. Appl. Algebra Geom.* 1 (2017), 239–253.

[5] J. Autry, P. Graves, J. Loucks, C. O'Neill, V. Ponomarenko and S. Yih. Square-free divisor complexes of certain numerical semigroup elements. *Involve* 14 (2021), 1–9.

[6] C. Axler. On the sum of the first $n$ prime numbers. *J. Théor. Nombres Bordeaux* 31(2) (2019), 293–311.

[7] S. Berndt, H. Brinkop, K. Jansen, M. Mnich and T. Stamm. New support size bounds for integer programming, applied to makespan minimization on uniformly related machines. *ISAAC 2023*, LIPIcs 283, 13:1–13:18.

[8] F. Bridoux, C. Crespelle, T. H. D. Phan and A. Richard. Dividing sum of cycles in the semiring of functional digraphs. *Natural Computing* 25 (2026), article 11. Also arXiv:2504.11943v2.

[9] C. Crespelle, T. H. D. Phan and A. Richard. Tensor product of functional graphs: algorithms for factorization and polynomial root-finding restricted to the cycle part. HAL hal-05392325v1 (2025). Also SSRN 5316210.

[10] S. Cyrusian, A. Domat, C. O'Neill, V. Ponomarenko, E. Ren and M. Ward. On numerical semigroup elements and the $\ell_0$ and $\ell_\infty$ norms of their factorizations. In *The Ideal Theory and Arithmetic of Rings, Monoids, and Semigroups*, *Contemp. Math.* 836 (2026), 95–108.

[11] S. Cyrusian and N. Kaplan. Ordinarization numbers of numerical semigroups. *Communications in Algebra*, published online 11 February 2026. Also arXiv:2506.10222v2.

[12] A. Dennunzio, V. Dorigatti, E. Formenti, L. Manzoni and A. E. Porreca. Polynomial equations over finite, discrete-time dynamical systems. *Cellular Automata (ACRI 2018)*, LNCS 11115, 298–306.

[13] F. Doré, K. Perrot, A. E. Porreca, S. Riva and M. Rolland. Infinite trees for division and roots over finite discrete-time dynamical systems. *Natural Computing* 25 (2026), article 16.

[14] A. W. M. Dress and C. Siebeneicher. The Burnside ring of the infinite cyclic group and its relations to the necklace algebra, $\lambda$-rings, and the universal ring of Witt vectors. *Adv. Math.* 78 (1989), 1–41. Author version.

[15] Y. Dubey and S. Liu. On the smallest support size of integer solutions to linear equations. *Math. Program.* 214 (2025), 789–800.

[16] F. Eisenbrand and G. Shmonin. Carathéodory bounds for integer cones. *Oper. Res. Lett.* 34 (2006), 564–568.

[17] P. A. García-Sánchez, I. Ojeda and J. C. Rosales. Affine semigroups having a unique Betti element. *J. Algebra Appl.* 12 (2013), 1250177. Also arXiv:1203.4138v2.

[18] K. Jansen, F. Ohnesorge, L. Pirotton and M. Tutas. The support of bin packing is exponential. arXiv:2512.02758v2 (2026). Earlier version by K. Jansen, L. Pirotton and M. Tutas in *ESA 2025*, LIPIcs 351, 48:1–48:16.

[19] N. Kuosa. Division of Explicit Finite Dynamical Systems Is NP-Complete. Preprint, 2026, `hexagon:2611.00886`.

[20] É. Naquin and M. Gadouleau. Factorisation in the semiring of finite dynamical systems. *Theoret. Comput. Sci.* 998 (2024), 114509.

[21] A. E. Porreca and M. Rolland. Injective and pseudo-injective polynomial equations: From permutations to dynamical systems. arXiv:2604.04065v2 (2026).

[22] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois J. Math.* 6 (1962), 64–94.