

Deterministic and Las Vegas Algorithms for Sparse Nonnegative Convolution

Karl Bringmann · Nick Fischer · Vasileios Nakos

Saarland University and Max Planck Institute for Informatics,
Saarland Informatics Campus, Saarbrücken, Germany

Abstract

Computing the convolution $A \star B$ of two length- n integer vectors A, B is a core problem in several disciplines. It frequently comes up as a subroutine in various problem domains, e.g. in algorithms for Knapsack, k -SUM, All-Pairs Shortest Paths, and string pattern matching problems. For these applications it typically suffices to compute convolutions of *nonnegative* vectors. This problem can be classically solved in time $O(n \log n)$ using the Fast Fourier Transform.

However, in many applications the involved vectors are *sparse* and hence one could hope for *output-sensitive* algorithms to compute nonnegative convolutions. This question was raised by Muthukrishnan and solved by Cole and Hariharan (STOC '02) by a randomized algorithm running in near-linear time in the (unknown) output-size t . Chan and Lewenstein (STOC '15) presented a deterministic algorithm with a $2^{O(\sqrt{\log t \cdot \log \log n})}$ overhead in running time and the additional assumption that a small superset of the output is given; this assumption was later removed by Bringmann and Nakos (ICALP '21).

In this paper we present the first deterministic near-linear-time algorithm for computing sparse nonnegative convolutions. This immediately gives improved deterministic algorithms for the state-of-the-art of output-sensitive Subset Sum, block-mass pattern matching, N -fold Boolean convolution, and others, matching up to log-factors the fastest known randomized algorithms for these problems. Our algorithm is a blend of algebraic and combinatorial ideas and techniques.

Additionally, we provide two fast *Las Vegas* algorithms for computing sparse nonnegative convolutions. In particular, we present a simple $O(t \log^2 t)$ time algorithm, which is an accessible alternative to Cole and Hariharan's algorithm. Subsequently, we further refine this new algorithm to run in Las Vegas time $O(t \log t \cdot \log \log t)$, which matches the running time of the dense case apart from the $\log \log t$ factor.

Funding This work is part of the project TIPEA that has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 850979).

1 Introduction

The *convolution* of two integer vectors A, B is the vector $A \star B$ which is defined coordinate-wise by $(A \star B)_k = \sum_{i+j=k} A_i \cdot B_j$. Computing convolutions of integer vectors A, B is a fundamental computational primitive, which arises in several disciplines of science and engineering. It has been a vital component in fields like signal processing, deep learning (convolutional neural networks) and computer vision. Inside traditional algorithm design it is crucially used as a subroutine in k -SUM [15], Subset Sum [9, 30, 12, 14] and various string problems [21, 27, 18], to name a few.

The aforementioned applications of interest within theoretical computer science typically come in the form of *nonnegative convolution*, where the vectors A, B have nonnegative entries. In fact, for many applications it suffices to solve the simpler *Boolean convolution* problem—here, the vectors A, B have 0–1 entries and the task is to compute the vector $A \otimes B$ with entries $(A \otimes B)_k = \bigvee_{i+j=k} A_i \wedge B_j$. This problem is equivalent to the computation of sumsets $X + Y = \{x + y : x \in X, y \in Y\}$ and this interpretation shows up very often in k -SUM and Subset Sum algorithms.

Moreover, Boolean (or nonnegative) convolutions form an essential ingredient to the *partition-and-convolve* design paradigm. The typical task for a problem approachable by a partition-and-convolve algorithm is to check for solutions of *all prescribed sizes* k . The general idea is to partition the search space into (usually) two parts, each of which is solved recursively. In that way, a size- k solution to the original problem is split into two parts of sizes i, j such that $i + j = k$. Therefore, to check whether there exists a size- k solution to the original problem we recombine the recursive computations by a Boolean convolution. This approach is very flexible and can also be applied to other convolution-type problems; for instance, using nonnegative convolutions in place of Boolean convolutions corresponds to *counting* solutions of all prescribed sizes.

Classically, the convolution of length- n vectors can be computed in deterministic time $O(n \log n)$ using the Fast Fourier Transform (FFT). It is widely conjectured that this algorithm is optimal but the evidence is scarce [2, 1] and this remains an important open problem. Furthermore, it is known that nonnegative convolution, general convolution (where entries can also be negative or complex) and the computation of Discrete Fourier transforms (DFT) are computationally equivalent, as each one can be reduced to the other.¹

The situation is less clear for *sparse convolutions* in all regards. Here the goal is to achieve *output-sensitive* algorithms where we analyze the running time in terms of t , the combined number of nonzero entries in A, B and $A \star B$ (in the Boolean and nonnegative cases, t is dominated by the number of nonzero entries in $A \star B$). The need for such a primitive appears in many situations, e.g. [18, 25, 15, 12, 13], as one may often be interested in algorithms that run in time proportional to the actual complexity of the output rather than the space the output points live in. These type of problems have been investigated by different communities, including fine-grained complexity [15] and string algorithms [18], computer algebra [38] and compressed sensing [26, 22], and they are very closely related to the famous sparse recovery problem, see e.g. [23, 22]. Subsequently, we review the most relevant literature on sparse convolutions.

Randomized Algorithms for Sparse Convolution. A large body of work addresses this problem [34, 18, 37, 33, 42, 5, 15, 38, 35, 24, 11]. The first breakthrough was a randomized Las Vegas algorithm for sparse nonnegative convolution in time $O(t \log^2 n)$, obtained by Cole and Hariharan [18]. Subsequent work improved upon this result in two directions: On the one hand, the nonnegativity assumption can be removed by a Monte Carlo algorithm with the same running time $O(t \log^2 n)$ [35], or with bit-complexity $\tilde{O}(t \log n)$ [24]. On the other hand, there exists an improved Monte Carlo algorithm for sparse nonnegative convolution in time $O(t \log t + \text{polylog } n)$ [11], which is optimal assuming that the dense problem requires FFT time $\Theta(n \log n)$. Most of these algorithms rely on a hashing-based approach.

Another more algebraic avenue to sparse convolution algorithms is via polynomial evaluation and interpolation. At the heart of this approach lies an old algorithm called *Prony's method* [19] which allows to efficiently interpolate a sparse polynomial. Unfortunately, this algorithm involves heavy algebraic computations (see [38] for a detailed survey) and the currently only known way to achieve near-linear running time in t uses randomization [38]. Thus, near-linear-time sparse convolution algorithms resulting from this approach are randomized as well. We compare our work against Prony's method in more detail in Section 1.2.

Note that in many applications of interest a sparse convolution algorithm is called many times with varying output size and thus it is desirable to have deterministic (or Las Vegas) algorithms for performing such a task. Additionally, the fastest known algorithm in the dense case (FFT) is deterministic, and thus it is natural to wonder to what extent randomization is necessary in the sparse case.

¹ For dense vectors, the nonnegativity assumption can be removed by appropriately increasing all entries. For the equivalence of computing convolutions and DFTs we remark that it is standard to express convolutions using DFT and inverse DFT, and the reverse direction is known as well [8] (assuming complex exponentials can be evaluated in constant time).

Deterministic Algorithms for Sparse Convolution. The first nontrivial deterministic result for sparse nonnegative convolution is a data structure that, after preprocessing one of the vectors in time $\Theta(t^2)$, computes the convolution with any given query vector in near-linear time $O(t \log^3 t)$ [4]. Later, Chan and Lewenstein [15] devised a deterministic algorithm running in time $t \cdot 2^{O(\sqrt{\log t \log \log n})}$, without preprocessing. Their algorithm is limited in the sense that it expects as an additional input the support (i.e., the set of nonzero coordinates) of $A \star B$. This assumption can be removed as shown by Bringmann and Nakos [13]; in Section 1.2 we provide more details. In summary, the state-of-the-art deterministic algorithms for computing sparse nonnegative convolutions either require heavy precomputations or fail to achieve near-linear time $O(t \text{ polylog } n)$. Our driving question is therefore:

*Can sparse nonnegative convolutions be computed
in deterministic time $O(t \text{ polylog } n)$?*

1.1 Our Results

Our main result is an affirmative answer to our driving question.

Theorem 1 (Deterministic). *There is a deterministic algorithm to compute the convolution of two nonnegative vectors $A, B \in \mathbb{N}^n$ in time $O(t \text{ polylog}(n\Delta))$, where $t = \|A \star B\|_0$ denotes the number of nonzero entries in $A \star B$ and $\Delta = \|A \star B\|_\infty$ is the maximum entry size.*

This result improves the previously best known time $t \cdot 2^{O(\sqrt{\log t \cdot \log \log n})}$ obtained by [15, 13].

As a corollary we can efficiently derandomize known algorithms for several problems which use sparse nonnegative convolution as a subroutine. For all these applications, we can simply replace the former randomized algorithms with our deterministic one in a black-box manner. Of course, for the same derandomization we could alternatively use the $t \cdot 2^{O(\sqrt{\log t \log \log n})} = t \cdot n^{o(1)}$ -time algorithm from [15, 13] and therefore our contribution can alternatively be seen as improving the best deterministic time from $T^{1+o(1)}$ to $\tilde{O}(T)$. Specifically, we obtain improvements for the following problems:

- **Output-Sensitive Subset Sum:** Given a set X of integers and a threshold τ , compute the set S of all numbers less than τ which can be expressed as a subset sum of X . The best-known randomized algorithm runs in time $\tilde{O}(|S|^{4/3})$ [12], and it can be derandomized in same running time.
- **N -fold Boolean Convolution:** Given N Boolean vectors A_1, \dots, A_N , compute the Boolean convolution $A_1 \otimes \dots \otimes A_N$ (with or without wrap-around) in input- plus output-sensitive time. It was recently shown that this problem can be solved in randomized near-linear time $O(t \text{ polylog } n)$ [13]. Our derandomization achieves the same running time. This yields a new deterministic near-linear-time algorithm for Modular Subset Sum which is rather different than the known ones [6], as discussed in [13].
- **Block-Mass Pattern Matching:** Given a length- n text T and a length- m pattern P over the alphabet \mathbb{N} , the task is to output all possible indices $0 \leq k_0 \leq \dots \leq k_m \leq n$ such that $P_i = \sum_{k_i \leq j < k_{i+1}} T_j$ for all positions $i \in [m]$. Building on the data structure from [4], this problem is known to be solvable in deterministic time $\tilde{O}(n+m)$ after preprocessing the text in time $O(n^2)$ [3]. The preprocessing time was later reduced to $O(n^{1+\epsilon})$, for any $\epsilon > 0$ [15]. We entirely remove the necessity to precompute and thereby reduce the total running time to $\tilde{O}(n+m)$.
- **3SUM in Special Cases:** In a breakthrough paper, Chan and Lewenstein [15] use sophisticated techniques to obtain randomized and deterministic subquadratic algorithms for a variety of problems related to 3-SUM, such as bounded monotone two-dimensional 3SUM, bounded monotone (min, +)-convolution, clustered integer 3-SUM, etc. The precise running time of their deterministic algorithm for these problems is $O(n^{1.864})$. Here we remove an $o(1)$ overhead in the exponent which is invisible due to rounding the constant in the exponent.

In addition to our new deterministic algorithm, we improve the state-of-the-art *Las Vegas* algorithms for sparse nonnegative convolutions in two regards: *simplicity* and *efficiency*. In fact, to the best of our knowledge the only known Las Vegas algorithm is due to Cole and Hariharan [18]; all randomized algorithms published later have only Monte Carlo guarantees [37, 5, 38, 35, 24, 11]. The expected running time of [18] is $O(t \log^2 n)$ and moreover, they prove the additional guarantee that their algorithm terminates in time $O(t \log^2 n)$ with high probability $1 - \frac{1}{n}$. However, the algorithm is very complicated, involves various string problems as subtasks and the precomputation of a large prime number. We provide an accessible alternative with the same theoretical guarantees; the simplest version can be summarized in 13 lines of pseudocode (Algorithm 1 on Page 6).

Theorem 2 (Simple Las Vegas). *Given nonnegative vectors $A, B \in \mathbf{N}^n$, there exists an algorithm to compute their convolution $A \star B$ in expected time $O(t \log^2 t)$, where $t = \|A \star B\|_0$. Moreover, with probability $1 - \delta$ the running time is bounded by $O(t \log^2(t/\delta))$.*

In comparison to Cole and Hariharan’s algorithm, our algorithm runs slightly faster in expectation (at least if $t \ll n$) and achieves the same high-probability guarantee (indeed, by setting $\delta = \frac{1}{n}$ the running time is bounded by $O(t \log^2 n)$ with probability at least $1 - \frac{1}{n}$). We further show how to reduce the expected running time, achieving optimality up to a log log factor.

Theorem 3 (Fast Las Vegas). *Given nonnegative vectors $A, B \in \mathbf{N}^n$, there exists an algorithm to compute their convolution $A \star B$ in expected time $O(t \log t \log \log t)$, where $t = \|A \star B\|_0$.*

Assuming that FFT-time $O(n \log n)$ is best-possible for computing dense convolutions, the best-possible algorithm for computing sparse convolutions requires time $\Omega(t \log t)$. Hence, our algorithm is likely optimal, up to the log log t factor.

1.2 Technical Overview

In this section we briefly outline the ideas behind Theorems 1 to 3.

1.2.1 Deterministic Algorithm

The key machinery powering our deterministic algorithm (Theorem 1) is a basic result from structured linear algebra which can be viewed as efficiently evaluating and interpolating *sparse* polynomials—under certain conditions. We present the algorithm by first explaining that key part (Part 1) and the assumptions it requires. In Parts 2 and 3 we then remove these assumptions by appropriate precomputations.

Part 1: Evaluation & Interpolation. The high-level approach follows the typical evaluation and interpolation pattern. Any vector V can be viewed as a polynomial $V(X) = \sum_{i=0}^{n-1} V_i X^i$. In that analogy, computing the convolution $A \star B$ of two vectors A, B corresponds to computing the product of their respective polynomials $A(X) \cdot B(X)$. The idea is to:

- 1** Evaluate $A(X)$ and $B(X)$ at some carefully chosen points $\omega^0, \omega^1, \dots, \omega^{t-1}$,
- 2** Compute the product $A(\omega^i) \cdot B(\omega^i)$ for all $i = 0, \dots, t-1$,
- 3** Interpolate the (hopefully unique) polynomial $C(X)$ with evaluations $C(\omega^i) = A(\omega^i) \cdot B(\omega^i)$.

In this way, we have reduced the task to the evaluation and interpolation of sparse polynomials. Let us start with the evaluation problem: Note that computing $V(\omega^0), \dots, V(\omega^{t-1})$ is equivalent to computing the following matrix-vector product:

$$\begin{bmatrix} V(\omega^0) \\ V(\omega^1) \\ \vdots \\ V(\omega^{t-1}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \omega^{x_1} & \omega^{x_2} & \cdots & \omega^{x_t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{x_1(t-1)} & \omega^{x_2(t-1)} & \cdots & \omega^{x_t(t-1)} \end{bmatrix} \begin{bmatrix} V_{x_1} \\ V_{x_2} \\ \vdots \\ V_{x_t} \end{bmatrix}. \quad (*)$$

This matrix has a very special form: It is the transpose of a Vandermonde matrix. It is known that performing linear algebra operations (such as computing matrix-vector products, or solving linear systems) with transposed Vandermonde matrices can be implemented in $O(t \log^2 t)$ field operations [28, 31, 36]. Thus, by viewing the integer vector V as a vector over some appropriately large finite field, we can evaluate $V(\omega^0), \dots, V(\omega^{t-1})$ in near-linear time. (We remark that this algorithm is numerically unstable, so using complex arithmetic is not an option.)

To perform the inverse task of interpolating the coefficients V_{x_1}, \dots, V_{x_t} given the evaluations $V(\omega^0), \dots, V(\omega^{t-1})$, we view $(*)$ as a system of linear equations with indeterminates V_{x_1}, \dots, V_{x_t} . As mentioned before this problem can be also solved in time $O(t \log^2 t)$. This nearly yields the algorithm, however, there are two major obstacles. First, in order to obtain a unique solution, the equation system should be nonsingular. It is easy to see that this is equivalent to the condition that $\omega^{x_1}, \dots, \omega^{x_t}$ are pairwise distinct. A reasonable way to achieve this is to let ω be a finite field element with multiplicative order at least $n \geq \deg(V)$. In Part 2 we explain how to obtain such an element. Second, in order to write down the equation system we have to know the indices x_1, \dots, x_t , i.e., the support of V . Concretely, in the algorithm we call the sparse interpolation problem for $V = A \star B$ and we therefore need to know $\text{supp}(A \star B)$ in advance. In Part 3 we discuss a recursive “scaling trick” to precompute a small superset of $\text{supp}(A \star B)$.

Part 2: Finding Large-Order Elements. In this part we care about finding an element ω with multiplicative order at least n in a finite field of size $\gg n$. There is a simple randomized algorithm: Pick a random element. Unfortunately, the best-known *deterministic* algorithms for finding a large-order element in a given prime field \mathbf{F}_p require time polynomial in p [16]. Thus it seems intractable to work over a finite field \mathbf{F}_p with $p \geq n$ as originally intended.

Fortunately, in a finite field $\mathbf{F}_q = \mathbf{F}_{p^m}$ with prime power order, it is possible to find large-order elements in time $\text{poly}(p, m)$ [17, 40, 41]. Specifically, setting $p, m = \text{polylog}(n)$ we can find an element ω with order at least n in time $\text{polylog}(n)$ [17]. Working over a finite field with small characteristic $p \leq \text{polylog}(n)$ has another drawback though: We cannot recover the entries of the vector $A \star B$ (which can have size up to n , even if A and B are bit-vectors to begin with). We remedy this problem by computing the convolution $A \star B$ over *several* finite fields $\mathbf{F}_{q_1}, \mathbf{F}_{q_2}, \dots$, and use the Chinese Remainder Theorem to identify the correct integer solution afterwards.

Part 3: Recursively Computing the Support. We finally discuss how to precompute the support $\text{supp}(A \star B)$. In fact, it suffices to compute a superset $T \supseteq \text{supp}(A \star B)$ with small size $|T| \leq O(t)$. We exploit a trick which was first applied to the context of convolutions in [13]; see also [12, 10]. Construct smaller vectors A', B' of length $\frac{n}{2}$ by $A'_i = A_i + A_{i+n/2}$ and $B'_j = B_j + B_{j+n/2}$ (that is, we fold A, B in half). We can recursively compute the convolution $C' = A' \star B'$. Then we extract T as

$$T = \left\{ k, k + \frac{n}{2}, k + n : k \in \text{supp}(C') \right\}.$$

This choice is correct: Clearly $|T| \leq 3t$, and it is easy to verify that T is indeed a superset of $\text{supp}(A \star B)$. The recursion only reaches depth $\log n$, and thus incurs a logarithmic overhead in the running time.

By combining Parts 2 and 3 we overcome both obstacles outlined in Part 1, and solve sparse nonnegative convolution in deterministic time $O(t \text{polylog}(n\Delta))$.

Comparison to Prony’s Method. Note that our main contribution can also be viewed as a deterministic near-linear-time algorithm to interpolate a univariate sparse polynomial *with nonnegative coefficients*. The classical approach to the sparse interpolation problem is by Prony’s method—an old algorithm first discovered by Prony in 1795 [19], and rediscovered later by Ben-Or and Tiwari [7]; see [38] for a detailed survey. Prony’s method involves heavy algebraic computations such as finding the minimal solution to a linear recurrence,

Algorithm 1**Input:** Nonnegative vectors $A, B \in \mathbf{N}^n$ **Output:** $C = A \star B$

```

1  for  $m \leftarrow 1, 2, 4, \dots, \infty$  do
2      repeat  $2 \log m$  times
3          Sample a linear hash function  $h : [n] \rightarrow [m]$ 
4          Compute  $X \leftarrow h(A) \star_m h(B)$ 
5          Compute  $Y \leftarrow h(\partial A) \star_m h(B) + h(A) \star_m h(\partial B)$ 
6          Compute  $Z \leftarrow h(\partial^2 A) \star_m h(B) + 2h(\partial A) \star_m h(\partial B) + h(A) \star_m h(\partial^2 B)$ 
7          Initialize  $R \leftarrow (0, \dots, 0)$ 
8          for each  $k \in [m]$  do
9              if  $X_k \neq 0$  and  $Y_k^2 = X_k \cdot Z_k$  then
10                  $z \leftarrow Y_k / X_k$ 
11                  $R_z \leftarrow R_z + X_k$ 
12  Let  $C$  be the coordinate-wise maximum of all vectors  $R$ 
13  if  $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$  then return  $C$ 

```

polynomial root finding, computing discrete logarithms and linear algebra with (transposed) Vandermonde systems. These computations can be carried out in near-linear time, but only using randomization [29, 38]. Our algorithm is similar to Prony’s method with two essential modifications: We replace the computationally expensive parts using the combinatorial trick (Part 3, here we critically use that the vectors are nonnegative) and derandomize the remaining steps (Parts 1 and 2) using classical methods.

1.2.2 Las Vegas Algorithms

Next, we outline the idea behind proving Theorem 2. Algorithm 1 is a simple Las Vegas algorithm with expected running time $O(t \log^2 t)$ as claimed in Theorem 2; however, to obtain the tail bound on the running time one has to slightly refine Algorithm 1. We provide this refinement along with a detailed analysis in Section 4; for the rest of the overview we will analyze the simple version in Algorithm 1.

To understand the pseudocode, we first clarify some notation: For a vector A , we denote by ∂A its *derivative* defined coordinate-wise as $(\partial A)_i = i \cdot A_i$. More generally, we denote by $\partial^d A$ its d -th *derivative* with $(\partial^d A)_i = i^d \cdot A_i$. This definition is in slight dissonance with the analogous definition for polynomials (which would require the derivative vector to be scaled and *shifted*), but we prefer this version as it leads to a slightly simpler algorithm.

Moreover, we define hashing for vectors: For a hash function $h : [n] \rightarrow [m]$ and a length- n vector A , define the length- m vector $h(A)$ via $h(A)_j = \sum_{i:h(i)=j} A_i$. The operator \star_m denotes convolution with wrap-around (see Section 2 for details).

Let us outline the high-level idea of Algorithm 1. The outer loop (Line 1) guesses the correct sparsity, i.e., as soon as the outer loop reaches a value $m \geq \Omega(t)$ we expect the algorithm to terminate. Each iteration of the repeat-loop (Line 2) is supposed to produce a vector R which closely approximates $A \star B$. More specifically, we prove that R satisfies the following two properties:

- 1** It always holds that $R \leq A \star B$ (coordinate-wise).
- 2** Equality is achieved at any coordinate with constant probability (provided that the outer loop has reached a sufficiently large value $m \geq \Omega(t)$).

It follows that C , the coordinate-wise maximum of several vectors R , also always satisfies $C \leq A \star B$. Hence, the algorithm never outputs an incorrect solution. Indeed, since C and $A \star B$ are nonnegative vectors, the vector $C = A \star B$ is the only one simultaneously satisfying $C \leq A \star B$ and $\|C\|_1 = \|A \star B\|_1 = \|A\|_1 \cdot \|B\|_1$. To see that Algorithm 1 terminates fast, note that the repeat-loop runs for $\Omega(\log m)$ iterations and thus, using the second claim we

correctly assign *all* coordinates with high probability.

The crucial part is to prove that R satisfies the claims 1 and 2. Intuitively, R consists of all nonzero entries from $A \star B$ which did not suffer from a collision with another nonzero entry. For a more formal argument, we analyze the inner-most loop (Line 8). For starters, focus on an iteration $k \in [m]$ and suppose that there is only a single nonzero entry in $A \star B$, say at z , which is hashed to the bucket k .² In this case we have $X_k = (A \star B)_z$, $Y_k = z \cdot (A \star B)_z$ and $Z_k = z^2 \cdot (A \star B)_z$. As a consequence, the conditions “ $X_k \neq 0$ ” and “ $Y_k^2 = X_k \cdot Z_k$ ” in Line 9 are satisfied. The algorithm then correctly identifies z in Line 10 and updates “ $R_z \leftarrow R_z + (A \star B)_z$ ” as intended.

However, to prove claim 1 (which is ultimately responsible for the Las Vegas guarantee), we have to be certain that Lines 10 and 11 are only executed if there is a single entry hashed to the k -th bucket (otherwise, the index z computed in Line 10 is likely to be nonsense). The key insight is that the simple test “ $Y_k^2 = X_k \cdot Z_k$ ” in Line 9 suffices, as can be proven by the following lemma (see Section 4 for a proof).

Lemma 4 (Testing 1-Sparsity). *If V be a nonnegative vector, then $\|\partial V\|_1^2 \leq \|V\|_1 \cdot \|\partial^2 V\|_1$. This inequality is tight if and only if $\|V\|_0 \leq 1$.*

This new tester is one of the reasons why we can achieve the claimed Las Vegas running time simplifying (and slightly improving) upon Cole and Hariharan’s algorithm. This concludes the overview of our simple Las Vegas algorithm (Theorem 2).

The insight behind our accelerated Las Vegas algorithm (Theorem 3) is that Algorithm 1 already reaches a very good approximation after much less than $O(\log m)$ iterations of the inner loop. Indeed, after only $O(\log \log n)$ iterations we expect that algorithm has already recovered $A \star B$ correctly up to a $(\log n)^{-\Omega(1)}$ fraction of the entries. At this point it becomes more efficient to switch to another recovery approach which exploits that $A \star B - C$ is already quite sparse, as in [11]. In particular, since $A \star B - C$ is a *nonnegative* vector and its sparsity is at most $t/\log n$, say, we can use the hash function $h(x) = x \bmod p$ for p being a random prime in $[t, 2t]$. This family of hash functions (1) satisfies that $h(A) \star_m h(B) - h(C) = h(A \star B - C)$ (and thus preserves all cancellations) and (2) isolates a constant fraction of elements in $A \star B - C$ with constant probability to clear up the rest of the elements. Note that it is important that $A \star B - C$ is $t/\log n$ sparse instead of t sparse for (2) to hold, because h is only $O(\log n)$ -universal. Choosing $O(\log t)$ different random primes and using the 1-sparsity testing we arrive at our desired algorithm. For the sparsity test we require that the vector $A \star B - C$ is nonnegative.

One catch is that this approach only gives a $O(t \log t \cdot \log \log n)$ -time algorithm (instead of the desired time with $\log \log t$ in place of $\log \log n$) due to the fact that $h(x)$ is $O(\log n)$ -universal and hence the random prime must be chosen in an interval that is also dependent on n rather than solely on t . To address this issue we apply the following precomputation: We hash to a $\text{poly}(t)$ -size universe and verify that this hashing was successful in Las Vegas randomized time, again using our 1-sparsity tester. The details of this step appear in Section 4.5.

1.3 Discussion and Open Problems

Our work raises several questions.

Better Deterministic Algorithms? By a closer inspection of the time analysis, our deterministic algorithm computes the convolution of sparse nonnegative vectors in time $O(t \log^5(n\Delta) \text{polyloglog}(n\Delta))$.

- 1** Can the running time be improved? In particular, is it possible to reduce the number of log factors or can we omit the dependence on n or Δ ?

² Strictly speaking, that condition is not sufficient because linear hashing is only “almost” additive. We ignore this technical issue in the overview and give the full analysis in Section 4.

2 Can the restriction to nonnegative vectors be removed, or equivalently, is it possible to achieve *sparse polynomial multiplication in deterministic near-linear time*? In our algorithm the only step which exploits nonnegativity is the recursive support computation (Part 3).

Better Las Vegas Algorithms? We proved that sparse nonnegative convolution is in Las Vegas time $O(t \log t \log \log t)$.

3 Is the restriction to nonnegative vectors necessary? This seems like a difficult question because we are not aware of algorithms that run in even slightly subquadratic time in t (without using heavy pre-computation).

4 Can one achieve $O(t \log t)$ Las Vegas running time matching the running time of the dense case? More specifically, can the sparsity-testing technique which lead to our Las Vegas algorithms be extended and incorporated to obtain the optimal running time? Recall that the key step in the analysis is the application of Lemma 4. This lemma can be generalized as follows: A nonnegative vector V is at most s -sparse if and only if the following positive-semidefinite matrix is nonsingular:

$$\begin{bmatrix} \|\partial^0 V\|_1 & \|\partial^1 V\|_1 & \cdots & \|\partial^s V\|_1 \\ \|\partial^1 V\|_1 & \|\partial^2 V\|_1 & \cdots & \|\partial^{s+1} V\|_1 \\ \vdots & \vdots & \ddots & \vdots \\ \|\partial^s V\|_1 & \|\partial^{s+1} V\|_1 & \cdots & \|\partial^{2s} V\|_1 \end{bmatrix};$$

see for instance [32, Theorem 3A] for a proof. One approach for an improved Las Vegas algorithm would be to hash to $t/\log t$ buckets using a linear hash function, recover each bucket as in [11] in $O(t \log t)$ time and, using the generalized sparsity-testing technique, verify that most buckets indeed have sparsity $O(\log t)$, which in turn means that all but a $1/\log t$ -fraction of $A \star B$ has been successfully recovered; then one can continue and recover the rest with $h(x) = x \bmod p$. Although promising, this approach suffers from precision issues (when implementing the $O(\log t)$ -tester the numbers get too large) and hence does not lead to the desired $O(t \log t)$ time. It would be very interesting to find a way to circumvent this obstacle and obtain the ideal $O(t \log t)$ Las Vegas running time.

2 Preliminaries

Machine Model. Throughout this paper we work over the Word RAM model. In particular, logical and arithmetic operations on machine words take constant time. Concerning the sparse convolution problem, we assume that both the indices and entries of the given vectors fit into a constant number of machine words.

Notation. Let \mathbf{Z} and \mathbf{N} denote the integers and nonnegative integers, respectively. For a prime power q , let \mathbf{F}_q denote the finite field with q elements. We set $[n] = \{0, \dots, n-1\}$. We write $\text{poly}(n) = n^{O(1)}$, $\text{polylog}(n) = (\log n)^{O(1)}$ and $\text{polyloglog}(n) = (\log \log n)^{O(1)}$.

We mostly denote vectors by A, B, C with A_i referring to the i -th coordinate in A . We define the *convolution* of two length- n vectors A and B as the vector $A \star B$ of length $2n-1$ with

$$(A \star B)_k = \sum_{\substack{i, j \in [n] \\ i+j=k}} A_i \cdot B_j.$$

The *cyclic convolution* $A \star_m B$ is the length- m vector with

$$(A \star_m B)_k = \sum_{\substack{i, j \in [n] \\ i+j \equiv k \pmod{m}}} A_i \cdot B_j.$$

We refer to $\text{supp}(A) = \{i \in [n] : A_i \neq 0\}$ as the *support* of A , we set $\|A\|_0 = |\text{supp}(A)|$ and say that A is s -sparse if $\|A\|_0 \leq s$. If A is a vector with real entries, then we also define $\|A\|_1 = \sum_i |A_i|$ and $\|A\|_\infty = \max_i |A_i|$ in the usual way, and we say that A is *nonnegative* if all of its entries are nonnegative. We often hash length- n vectors A using an arbitrary hash function $h : [n] \rightarrow [m]$ to shorter length- m vectors $h(A)$ defined by

$$h(A)_j = \sum_{\substack{i \in [n] \\ h(i)=j}} A_i.$$

Finite Field Arithmetic. Let $q = p^m$ be a prime power. Recall that the prime field \mathbf{F}_p can be represented as $\mathbf{Z}/p\mathbf{Z}$, the integers modulo p . The field \mathbf{F}_q can be represented as $\mathbf{F}_p[X]/\langle f \rangle$ where $f \in \mathbf{F}_p[X]$ is an arbitrary irreducible degree- m polynomial. There is a deterministic algorithm to precompute such an irreducible polynomial $f \in \mathbf{F}_p$ in time $\text{poly}(p, m)$ [39]; we will point out this step in our algorithms. Having precomputed f , we can perform the basic field operations in \mathbf{F}_q using polynomial arithmetic in time $\tilde{O}(\log q)$ [43].

Let us quickly recall some definitions from field theory. The *multiplicative order* of an element x is the smallest positive integer i such that $x^i = 1$; we also call x an i -th root of unity. The *minimal polynomial* of a field element $x \in \mathbf{F}$ is defined as the smallest-degree monic polynomial (i.e., with leading coefficient 1) over \mathbf{F} which vanishes at x . We say that two field elements x, y are *conjugate* if their minimal polynomials coincide.

3 Deterministic Algorithm

In this section we prove Theorem 1. We proceed in three steps, as outlined before.

3.1 The Key Step: Evaluation & Interpolation

The main algebraic ingredient to the algorithm is the following result about efficient computations with transposed Vandermonde matrices. For a proof, see e.g. [28, 31, 36].

Theorem 5 (Transposed Vandermonde Systems). *Let \mathbf{F} be a field. Given pairwise distinct elements $a_0, \dots, a_{n-1} \in \mathbf{F}$ and a vector $x \in \mathbf{F}^n$, let*

$$M = \begin{bmatrix} 1 & 1 & \dots & 1 \\ a_0 & a_1 & \dots & a_{n-1} \\ a_0^2 & a_1^2 & \dots & a_{n-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_0^{n-1} & a_1^{n-1} & \dots & a_{n-1}^{n-1} \end{bmatrix}.$$

Both Mx and $M^{-1}x$ can be computed in deterministic time $O(n \log^2 n)$ using $O(n \log^2 n)$ field operations.

We remark that the transposed Vandermonde matrix M is nonsingular if and only if the elements a_0, \dots, a_{n-1} are pairwise distinct. The next lemma reinterprets this result in terms of multi-point evaluation and interpolation of sparse polynomials. In analogy to the vector notation, we denote by $\text{supp}(A)$ the set of exponents i for which X^i has a nonzero coefficient in A , and we say that A is t -sparse if $|\text{supp}(A)| \leq t$.

Lemma 6 (Sparse Evaluation and Interpolation). *Let \mathbf{F} be a field and let $\omega \in \mathbf{F}$ have multiplicative order at least n . The following two computational problems can be solved in deterministic time $O(t \log^2 t + t \log n)$:*

- 1 Evaluation:** *Given a t -sparse degree- n polynomial A , evaluate $A(\omega^0), \dots, A(\omega^{t-1})$.*
- 2 Interpolation:** *Given $a_0, \dots, a_{t-1} \in \mathbf{F}$ and a size- t set $T \subseteq [n]$, interpolate the unique polynomial A with evaluations $A(\omega^i) = a_i$ for all $i \in [t]$ and $\text{supp}(A) \subseteq T$.*

► **Proof. 1 Evaluation:** Assume that A has the form $A(X) = \sum_{i=1}^t A_{x_i} X^{x_i}$. We precompute the powers $\omega^{x_1}, \dots, \omega^{x_t}$ by repeated squaring in time $O(t \log n)$. We can then compute the evaluations $A(\omega^0), \dots, A(\omega^{t-1})$ by computing the following transposed Vandermonde matrix-vector product:

$$\begin{bmatrix} A(\omega^0) \\ A(\omega^1) \\ \vdots \\ A(\omega^{t-1}) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \omega^{x_1} & \omega^{x_2} & \cdots & \omega^{x_t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{x_1(t-1)} & \omega^{x_2(t-1)} & \cdots & \omega^{x_t(t-1)} \end{bmatrix} \begin{bmatrix} A_{x_1} \\ A_{x_2} \\ \vdots \\ A_{x_t} \end{bmatrix}.$$

Since ω has order at least n , the elements $\omega^{x_1}, \dots, \omega^{x_t}$ are pairwise distinct. Therefore, this matrix is nonsingular and we may apply Theorem 5 to efficiently evaluate the product in time $O(t \log^2 t)$.

2 **Interpolation:** Let x_1, \dots, x_t denote the elements in T . We similarly prepare $\omega^{x_1}, \dots, \omega^{x_t}$ via repeated squaring. To interpolate A , we solve the following transposed Vandermonde equation system with indeterminates A_{x_1}, \dots, A_{x_t} :

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{t-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \omega^{x_1} & \omega^{x_2} & \cdots & \omega^{x_t} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{x_1(t-1)} & \omega^{x_2(t-1)} & \cdots & \omega^{x_t(t-1)} \end{bmatrix} \begin{bmatrix} A_{x_1} \\ A_{x_2} \\ \vdots \\ A_{x_t} \end{bmatrix}.$$

Again, this matrix is nonsingular and thus Theorem 5 applies to compute a solution in time $O(t \log^2 t)$. Setting $A = \sum_{i=1}^t A_{x_i} X^{x_i}$, we have clearly reconstructed a polynomial with the correct evaluations $A(\omega^i) = a_i$ and support set $\text{supp}(A) \subseteq T$. Moreover, A is the only polynomial satisfying these conditions since the equation system is nonsingular and therefore A is uniquely determined. ◀

Using Lemma 6 we obtain our main result assuming we know a superset of the support and an appropriate element ω .

Lemma 7 (Sparse Convolution over a Large Field). *Let \mathbf{F} be a field. Given $A, B \in \mathbf{F}^n$, a set $T \supseteq \text{supp}(A \star B)$ and an element $\omega \in \mathbf{F}$ with multiplicative order at least n , we can compute $A \star B$ in deterministic time $O(t \log^2 t + t \log n)$ using $O(t \log^2 t + t \log n)$ field operations. Here, $t = \|A\|_0 + \|B\|_0 + |T|$.*

► **Proof.** We follow an evaluation–interpolation approach. Let us identify vectors with polynomials via $A(X) = \sum_{i=0}^{n-1} A_i X^i$. In this correspondence, taking convolutions $A \star B$ corresponds to multiplying polynomials $A(X) \cdot B(X)$.

We first evaluate $A(\omega^0), \dots, A(\omega^{t-1})$ and $B(\omega^0), \dots, B(\omega^{t-1})$ using Lemma 6. We then apply Lemma 6 again to interpolate a polynomial $C(X)$ with $\text{supp}(C) \subseteq T$ and $C(\omega^i) = A(\omega^i) \cdot B(\omega^i)$ for all $i \in [t]$. One solution is the correct polynomial $C(X) = A(X) \cdot B(X)$, and Lemma 6 guarantees that this is the unique solution. The running time is $O(t \log^2 t + t \log n)$ as claimed. ◀

It remains to construct ω (see Section 3.2) and to find a superset of the support (see Section 3.3).

3.2 Finding Large-Order Elements

We next solve the sparse convolution problem for integer vectors A, B assuming that we know the support of $A \star B$, using what we have established in the last section. We start with the following two lemmas due to Cheng [17]; for completeness we include short proofs.

Lemma 8 ([17]). *Let $\beta \in \mathbf{F}_p$ be primitive. Then $X^{p-1} - \beta \in \mathbf{F}_p[X]$ is irreducible.*

► **Proof.** Let $f = X^{p-1} - \beta$ and let $f = f_1 \dots f_m$ denote its factorization into monic irreducibles. We first prove that all factors have the same degree. Let α be a root of f

(in a field extension). Then $\{x\alpha : x \in \mathbf{F}_p^\times\}$ must be the full set of roots of f . Indeed, $(x\alpha)^{p-1} - \beta = x^{p-1}\beta - \beta = 0$ by Fermat's Little Theorem, and there cannot be other roots since f has degree $p-1$. Pick arbitrary distinct indices $1 \leq i, j \leq m$; we prove that $\deg(f_i) \leq \deg(f_j)$. Let $x, y \in \mathbf{F}_p^\times$ be such that $x\alpha$ is a root of f_i and $y\alpha$ is a root of f_j . We can construct a polynomial $f'_j(X) = f_j(yx^{-1}X)$, which by construction has degree $\deg(f_j)$ and has $x\alpha$ as a root. But recall that f_i is irreducible (and monic) and therefore the minimal polynomial of $x\alpha$. It follows that $\deg(f_i) \leq \deg(f'_j) = \deg(f_j)$. Since i, j were arbitrary we conclude that all polynomials f_1, \dots, f_m must have common degree $d = \frac{p-1}{m}$.

Next, we prove that $m = 1$. Let $\alpha_1, \dots, \alpha_d$ denote the roots of f_1 (in a field extension). As observed before, we have that $\alpha_i \alpha_j^{-1} \in \mathbf{F}_p$ for all i, j . Moreover, $\prod_i \alpha_i$ is the constant coefficient of f_1 and thus $\prod_i \alpha_i \in \mathbf{F}_p$. It follows that $\alpha_1^d = \prod_{i=1}^d \alpha_1 \alpha_i^{-1} \alpha_i$ is an element of \mathbf{F}_p . Recall that α_1 is a root of f and hence $\alpha_1^{p-1} = (\alpha_1^d)^m = \beta$. Finally, any value $m > 1$ would contradict the primitivity of β . \blacktriangleleft

Lemma 9 ([17]). *Let $f = X^{p-1} - \beta \in \mathbf{F}_p[X]$ be an irreducible polynomial. Then $X + 1$ has multiplicative order at least 2^p in $\mathbf{F}_p[X]/\langle f \rangle$ provided that $p \geq 7$.*

► **Proof.** Let $\mathbf{F}_{p^{p-1}}$ denote the field $\mathbf{F}_p[X]/\langle f \rangle$. Let s denote the order of $X + 1 \in \mathbf{F}_{p^{p-1}}$ and let S denote the set of s -th roots of unity in $\mathbf{F}_{p^{p-1}}$ (that is, S is the set of all polynomials $g \in \mathbf{F}_p[X]/\langle f \rangle$ such that $g^s = 1 \pmod{f}$). We show that S must be large. We clearly have $X + 1 \in S$. More generally, for any $i \in \mathbf{F}_p^\times$ we also have $iX + 1 \in S$ since $X + 1$ and $iX + 1$ are conjugate over \mathbf{F}_p . Furthermore, S is closed under multiplication.

Let $E \subseteq \mathbf{N}^{p-1}$ be the set of all sequences $e = (e_1, \dots, e_{p-1})$ with entry sum $\sum_i e_i = p-2$. For any such sequence $e \in E$, we define $\phi(e) = \prod_{i=1}^{p-1} (iX + 1)^{e_i} \in \mathbf{F}_{p^{p-1}}$. By the previous paragraph, ϕ is a map $\phi : E \rightarrow S$. We claim that ϕ is injective. If $\phi(e) = \phi(e')$ for distinct $e, e' \in E$, then by definition

$$\prod_{i=1}^{p-1} (iX + 1)^{e_i} = \prod_{i=1}^{p-1} (iX + 1)^{e'_i} \pmod{f}.$$

Recall that f has degree $p-1$, but $\sum_i e_i = \sum_i e'_i < p-1$. It follows that the equation remains true even without computing modulo f :

$$\prod_{i=1}^{p-1} (iX + 1)^{e_i} = \prod_{i=1}^{p-1} (iX + 1)^{e'_i}.$$

However, this identity contradicts unique factorization in $\mathbf{F}_p[X]$. It follows that ϕ is injective and therefore $s \geq |S| \geq |E|$. Finally, by a simple counting argument one can show that $|E| = \binom{p-4}{p-2} \geq 2^p$, for all $p \geq 7$. \blacktriangleleft

For the rest of this section, we will analyze Algorithm 2.

Lemma 10 (Correctness of Algorithm 2). *Given integer vectors A, B and an arbitrary set $T \supseteq \text{supp}(A \star B)$, Algorithm 2 correctly returns $C = A \star B$.*

► **Proof.** First, focus on an arbitrary iteration i of the loop in Lines 3 to 8. We prove that the algorithm computes the vector $C^i \in \mathbf{F}_{p_i}$ which is obtained from $C = A \star B$ by reducing all coefficients modulo p_i . The polynomial $X^{p_i-1} - \beta$ computed in Lines 4 and 5 is indeed irreducible by Lemma 8, so we can represent \mathbf{F}_{q_i} as $\mathbf{F}_{p_i}/\langle X^{p_i-1} - \beta \rangle$ as claimed. Moreover, the element $\omega \in \mathbf{F}_{q_i}$ constructed in Line 6 has multiplicative order at least $2^{p_i} \geq n$ by Lemma 9. The preconditions of Lemma 7 are satisfied ($T \supseteq \text{supp}(A \star B) \supseteq \text{supp}(A^i \star B^i)$ and ω has order at least n), hence we correctly compute $C^i = A^i \star B^i$ in Line 8. Note that although we carry out the computations over the extension field \mathbf{F}_{q_i} , the vector C^i is guaranteed to have coefficients in \mathbf{F}_{p_i} .

We finally use the Chinese Remainder Theorem to recover C from its images modulo p_1, \dots, p_k . As $\prod_{i=1}^k p_i \geq 2^k \geq n \|A\|_\infty \|B\|_\infty$ exceeds the maximum coefficient in C , this recovery step correctly identifies $C = A \star B$. \blacktriangleleft

Algorithm 2**Input:** Vectors $A, B \in \mathbf{Z}^n$ and a set $T \supseteq \text{supp}(A \star B)$ **Output:** $C = A \star B$

- 1 Let $k = \lceil \log(n \|A\|_\infty \|B\|_\infty) \rceil$
- 2 Compute the smallest k primes p_1, \dots, p_k larger than $\lceil \log n \rceil$
- 3 **for** $i \leftarrow 1, \dots, k$ **do**
- 4 Find a primitive element $\beta \in \mathbf{F}_{p_i}$ by brute-force
- 5 Let $q_i = p_i^{p_i-1}$ and represent \mathbf{F}_{q_i} as $\mathbf{F}_{p_i}[X]/\langle X^{p_i-1} - \beta \rangle$
- 6 Let $\omega = X + 1 \in \mathbf{F}_{q_i}$
- 7 Reduce the coefficients of A, B modulo p_i to obtain $A^i, B^i \in \mathbf{F}_{p_i}^n \subseteq \mathbf{F}_{q_i}^n$
- 8 Compute $C^i \leftarrow A^i \star B^i$ over \mathbf{F}_{q_i} using Lemma 7 with T and ω
- 9 **for each** $x \in T$ **do**
- 10 Use Chinese Remaindering to recover $C_x \in \mathbf{Z}$ from $C_x^1 \in \mathbf{F}_{p_1}, \dots, C_x^k \in \mathbf{F}_{p_k}$
- 11 **return** C with entries C_x for $x \in T$ and zeros elsewhere

Lemma 11 (Running Time of Algorithm 2). *The running time of Algorithm 2 is bounded by $O(t \log^4 n \text{polyloglog } n)$ where $t = \|A\|_0 + \|B\|_0 + |T|$, assuming that $\|A\|_\infty, \|B\|_\infty \leq \text{poly}(n)$.*

► **Proof.** Assuming that $\|A\|_\infty, \|B\|_\infty \leq \text{poly}(n)$, we have $k = \lceil \log(n \|A\|_\infty \|B\|_\infty) \rceil \leq O(\log n)$. Note that $p_1, \dots, p_k \leq \tilde{O}(\log n)$ by the Prime Number Theorem, and therefore computing these primes in Line 2 takes time $\tilde{O}(\log n)$, using for instance Eratosthenes' sieve. Finding a primitive element $\beta \in \mathbf{F}_{p_i}$ in Line 4 takes time $\tilde{O}(\log n)$ as well and Lines 5 to 7 have negligible costs. Per iteration, running the convolution algorithm in Line 8 takes time $O(t \log^2 t + t \log n) = O(t \log^2 n)$ and requires the computation of at most $O(t \log^2 n)$ field operations in \mathbf{F}_{q_i} , thus amounting for time $O(t \log^3 n \text{polyloglog } n)$. In total the loop in Line 3 takes time $O(t \log^4 n \text{polyloglog } n)$. Finally, each call to the algorithmic Chinese Remainder Theorem in Line 10 takes time $O(\log^2(\prod_{i=1}^k p_i)) = O(\log^2 n \text{polyloglog } n)$ [43]. ◀

We remark that our algorithm can be somewhat simplified by exploiting the following result: For any finite field \mathbf{F}_{p^m} , one can construct in time $\text{poly}(p, m)$ a (simple-structured) set which is guaranteed to contain a primitive element [40, 41]. The drawback is that the running time worsens by a couple of log factors.

3.3 Recursively Computing the Support

We finally remove the assumption that the support of $A \star B$ is given as part of the input.

Lemma 12 (Deterministic Sparse Nonnegative Convolution). *There is a deterministic algorithm to compute the convolution of two nonnegative vectors $A, B \in \mathbf{N}^n$ in time $O(t \log^5 n \text{polyloglog } n)$ where $t = \|A \star B\|_0$, assuming that $\|A\|_\infty, \|B\|_\infty \leq \text{poly}(n)$.*

► **Proof.** We construct a recursive algorithm for computing $C = A \star B$ using the scaling trick from [13]. In order to apply Algorithm 2, we first recursively compute a set $T \supseteq \text{supp}(C)$. To this end, let A', B' be vectors of length $\lceil \frac{n}{2} \rceil$ defined by $A'_i = A_i + A_{i+\lceil n/2 \rceil}$ and $B'_j = B_j + B_{j+\lceil n/2 \rceil}$ (that is, we fold A and B in half). We call the convolution algorithm recursively to compute $C' = A' \star B'$, and assign

$$T = \{k', k' + \lceil \frac{n}{2} \rceil, k' + 2 \cdot \lceil \frac{n}{2} \rceil : k' \in \text{supp}(C')\}.$$

We claim that indeed $T \supseteq \text{supp}(C)$. To see this, let $k \in \text{supp}(C)$ and write $k = i + j$ for some $i \in \text{supp}(A)$ and $j \in \text{supp}(B)$. By construction we have $i' = i \bmod \lceil \frac{n}{2} \rceil \in \text{supp}(A')$ and $j' = j \bmod \lceil \frac{n}{2} \rceil \in \text{supp}(B')$, and thus $k' = i' + j' \in \text{supp}(C')$. By definition it is immediate that $k' \in \{k, k - \lceil \frac{n}{2} \rceil, k - 2 \cdot \lceil \frac{n}{2} \rceil\}$ and therefore $k \in T$.

We finally analyze the running time. It takes time $O(|T| \log^4 n \text{polyloglog } n)$ to call Algorithm 2 once, by Lemma 10. Note that $|T| \leq 3|\text{supp}(C')| \leq 3t$, and thus each call takes time $O(t \log^4 n \text{polyloglog } n)$. Since the length of all vectors is halved in every step, after

$\log n$ recursion levels the problem has reached constant input size. The total running time is bounded by $O(t \log^5 n \text{polyloglog } n)$. ◀

The proof of Theorem 1 is now immediate from Lemma 12. To analyze the algorithm for vectors with entries of size Δ , simply view the vectors as having length $n' = \max\{n, \Delta\}$.

4 Las Vegas Algorithms

The goal of this section is to prove Theorems 2 and 3. We first gather some facts about hash functions (Section 4.1) and prove the sparsity testing lemma (Section 4.2). Then we prove Theorem 2 (Section 4.3) and Theorem 3 (Sections 4.4 and 4.5).

4.1 Hashing

Recall that a linear hash function $h : [n] \rightarrow [m]$ is defined by $h(x) = (ax \bmod N) \bmod m$, where $N \geq n$ is fixed and $a \in [N]$ is a random number. Typically N is a prime number, in which case it is easy to prove that the family is 2-universal. However, this choice is inefficient since precomputing a prime number N requires time $\text{polylog } N$. In many applications this overhead is negligible—in our case it would incur an additive $\text{polylog } n$ term to the running time which is otherwise independent of n . One can remove this overhead and perform linear hashing *without* prime numbers as proven e.g. in [20]. In Appendix A we provide a different self-contained proof.

Lemma 13 (Linear Hashing without Primes). *Let $n \geq m$ be arbitrary. There is a family of linear hash functions $h : [n] \rightarrow [m]$ with the following three properties.*

- 1 **Efficiency:** *Sampling and evaluating h takes constant time.*
- 2 **Uniform Differences:** *For any distinct keys $x, y \in [n]$ and for any $q \in [m]$, the probability that $h(x) - h(y) \equiv q \pmod{m}$ is at most $O(\frac{1}{m})$.*
- 3 **Almost-Additiveness:** *There exists a constant-size set $\Phi \subseteq [m]$ such that for all keys $x, y \in [n]$ it holds that $h(x) + h(y) \equiv h(x + y) + \phi \pmod{m}$ for some $\phi \in \Phi$.*

To prove Theorem 3 we additionally make use of another family of hash functions: For a random prime number p , the hash function $h(x) = x \bmod p$ satisfies similar properties.

Lemma 14 (Random Prime Hashing). *Let $n \geq m$ be arbitrary. The family of hash functions $h(x) = x \bmod p$ where $p \in [m, 2m]$ is a random prime satisfies the following three properties.*

- 1 **Efficiency:** *Sampling h takes time $\text{polylog } m$ and evaluating takes constant time.*
- 2 **Almost-Universality:** *For any distinct keys $x, y \in [n]$, the probability that $h(x) = h(y)$ is at most $O(\frac{\log n}{m})$.*
- 3 **Additiveness:** *For all keys $x, y \in [n]$ it holds that $h(x) + h(y) \equiv h(x + y) \pmod{p}$.*

4.2 Derivatives and Sparsity Testing

Recall that we define the *derivative* ∂A of a vector A coordinate-wise by $(\partial A)_i = i \cdot A_i$, and we define the d -th *derivative* $\partial^d A$ by $(\partial^d A)_i = i^d \cdot A_i$. The crucial ingredient for the Las Vegas guarantee is the following lemma about testing 1-sparsity of a vector, having access to its first and second derivatives.

Lemma 4 (Testing 1-Sparsity). *If V be a nonnegative vector, then $\|\partial V\|_1^2 \leq \|V\|_1 \cdot \|\partial^2 V\|_1$. This inequality is tight if and only if $\|V\|_0 \leq 1$.*

► **Proof.** Note that since V is nonnegative, we can rewrite $V_i = \sqrt{V_i} \cdot \sqrt{V_i}$. The proof is a straightforward application of the Cauchy-Schwartz inequality:

$$\|\partial V\|_1^2 = \left(\sum_i i V_i \right)^2 = \left(\sum_i \sqrt{V_i} \cdot i \sqrt{V_i} \right)^2 \leq \left(\sum_i V_i \right) \left(\sum_i i^2 V_i \right) = \|V\|_1 \cdot \|\partial^2 V\|_1.$$

Algorithm 3**Input:** Nonnegative vectors $A, B \in \mathbf{N}^n$ and a parameter m **Output:** A nonnegative vector $R \leq A \star B$, for details see Lemma 16

```

1  Sample a linear hash function  $h : [n] \rightarrow [m]$ 
2  Compute  $X \leftarrow h(A) \star_m h(B)$ 
3  Compute  $Y \leftarrow h(\partial A) \star_m h(B) + h(A) \star_m h(\partial B)$ 
4  Compute  $Z \leftarrow h(\partial^2 A) \star_m h(B) + 2h(\partial A) \star_m h(\partial B) + h(A) \star_m h(\partial^2 B)$ 
5  Initialize  $R \leftarrow (0, \dots, 0)$ 
6  for each  $k \in [m]$  do
7      if  $X_k \neq 0$  and  $Y_k^2 = X_k \cdot Z_k$  then
8           $z \leftarrow Y_k / X_k$ 
9           $R_z \leftarrow R_z + X_k$ 
10 return  $R$ 

```

Recall that the Cauchy-Schwartz inequality is tight if and only if the involved vectors V and $\partial^2 V$ are scalar multiples of each other. This is possible if and only if $\|V\|_0 \leq 1$. ◀

4.3 Simple Algorithm

We are finally ready to analyze Algorithm 1. For the ease of presentation, we have extracted the core part of Algorithm 1 (Lines 3 to 11) as Algorithm 3, and our first goal is a detailed analysis of that core part.

To increase clarity we shall adopt the following naming convention for the rest of this section: The indices $x, y, z \in [n]$ exclusively denote coordinates of large vectors, whereas $i, j, k \in [m]$ denote coordinates of the hashed vectors, or equivalently, buckets of a hash function h . The first lemma analyzes the vectors X, Y, Z computed by the algorithm.

Lemma 15. *Let h, X, Y, Z be as in Algorithm 3. Moreover, for a bucket $k \in [m]$ define the nonnegative vector $V^k \in \mathbf{N}^n$ by*

$$V_z^k = \sum_{\substack{x+y=z \\ h(x)+h(y) \equiv k \pmod{m}}} A_x \cdot B_y.$$

Then $X_k = \|V^k\|_1$, $Y_k = \|\partial V^k\|_1$ and $Z_k = \|\partial^2 V^k\|_1$.

Note that $A \star B = \sum_k V^k$. Intuitively, the vector V^k is that part of $A \star B$ which is hashed into the k -th bucket.

► **Proof.** We merely showcase that $Y_k = \|\partial V^k\|_1$; the other proofs are very similar. For convenience, let us denote equality modulo m by \equiv . It holds that:

$$\begin{aligned}
Y_k &= \left(h(\partial A) \star_m h(B) + h(A) \star_m h(\partial B) \right)_k \\
&= \sum_{i+j \equiv k} h(\partial A)_i \cdot h(B)_j + h(A)_i \cdot h(\partial B)_j \\
&= \sum_{\substack{x,y \\ h(x)+h(y) \equiv k}} (\partial A)_x \cdot B_y + A_x \cdot (\partial B)_y \\
&= \sum_{\substack{x,y \\ h(x)+h(y) \equiv k}} (x+y) \cdot A_x \cdot B_y \\
&= \sum_z z \cdot \sum_{\substack{x+y=z \\ h(x)+h(y) \equiv k}} A_x \cdot B_y \\
&= \sum_z z \cdot V_z^k \\
&= \|\partial V^k\|_1.
\end{aligned}$$

Next, we will prove that in every iteration the algorithm computes a feasible approximation R to the target vector $A \star B$.

Lemma 16 (Correctness and Running Time of Algorithm 3). *Given nonnegative vectors A, B and any parameter m , Algorithm 3 runs in time $O(m \log m)$ and computes a vector R such that for every $z \in [n]$:*

- $R_z \leq (A \star B)_z$ (always), and
- $R_z < (A \star B)_z$ with probability at most $c \cdot \|A \star B\|_0/m$ for some constant c .

► **Proof.** Fix an iteration $k \in [m]$ of the loop (Line 6) and suppose that the condition in Line 7 is satisfied. Defining V^k as in the previous lemma, we claim that V^k is exactly 1-sparse. Indeed, on the one hand, V^k is not the all-zeros vector as $\|V^k\|_1 = X_k > 0$. On the other hand, since $\|V^k\|_1 \cdot \|\partial^2 V^k\|_1 = X_k \cdot Z_k = Y_k^2 = \|\partial V^k\|_1^2$ we have that $\|V^k\|_0 \leq 1$ by Lemma 4. Given that V^k is 1-sparse, it is easy to check that the value $z := Y_k/X_k$ as computed in Line 8 is the unique nonzero coordinate in V^k , i.e., $\text{supp}(V^k) = \{z\}$. It follows that the update in Line 9 is in fact an update of the form “ $R \leftarrow R + V^k$ ”. Recall that $\sum_k V^k = A \star B$, and thus the first item follows directly.

Next, we focus on the second item. We can assume that $z \in \text{supp}(A \star B)$ as otherwise the statement is trivial given the previous paragraph. Let $\Phi \subseteq [m]$ be the set from Lemma 13. We say that z *collides* with another index z' if there are $\phi, \phi' \in \Phi$ such that $h(z) + \phi \equiv h(z') + \phi' \pmod{m}$. If z does not collide with any other $z' \in \text{supp}(A \star B)$ then we say that z is *isolated*. The remaining proof splits into the following two statements:

- *Each index $z \in \text{supp}(A \star B)$ is isolated with probability $1 - O(\|A \star B\|_0/m)$.* If z collides with another index z' then we have $h(z) - h(z') \equiv q \pmod{m}$ for some $q = \phi - \phi'$, $\phi, \phi' \in \Phi$. For any fixed q this event occurs with probability at most $O(\frac{1}{m})$ by the uniform difference property of linear hashing (Lemma 13). Taking a union bound over the constant number of elements q , we conclude that z collides with z' with probability at most $O(\frac{1}{m})$. Hence the expected number of collisions is $O(\|A \star B\|_0/m)$. Using Markov’s inequality we finally obtain that a collision occurs with probability at most $O(\|A \star B\|_0/m)$ and only in that case z fails to be isolated.
- *Whenever z is isolated we have $R_z = (A \star B)_z$.* To see this, it suffices to argue that for all k of the form $k \equiv h(z) + \phi \pmod{m}$, for some $\phi \in \Phi$, the vectors V^k are at most 1-sparse. In that case the corresponding iterations k each perform the update “ $R \leftarrow R + V^k$ ” in Line 9 and the claim follows since $R_z = \sum_k V_z^k = (A \star B)_z$. So suppose that some vector V^k is at least 2-sparse, i.e., there exist $x, x' \in \text{supp}(A)$ and $y, y' \in \text{supp}(B)$ such that $h(x) + h(y) \equiv h(x') + h(y') \equiv k \pmod{m}$ and $x + y \neq x' + y'$. Then either $z' := x + y$ or $z' := x' + y'$ differs from z , and we have witnessed a collision between z and z' . This contradicts the assumption that z is isolated.

Finally, note that the running time is dominated by the six calls to FFT in Lines 2 to 4 taking time $O(m \log m)$. The loop (Line 6) only takes linear time. ◀

Recall that Algorithm 1 simply calls Algorithm 3 several times and returns the coordinate-wise maximum C of all computed vectors R as soon as $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$. The bucket size m increases from iteration to iteration. Given the analysis of Algorithm 3 it remains to prove that Algorithm 1 is correct and fast, thereby proving Theorem 2.

Lemma 17 (Correctness of Algorithm 1). *Whenever Algorithm 1 outputs a vector C , then $C = A \star B$ (with error probability 0).*

► **Proof.** In Line 11, C is computed as the coordinate-wise maximum of several vectors R computed by Algorithm 3. The previous lemma asserts that $R \leq A \star B$ (coordinate-wise) and therefore also $C \leq A \star B$. Moreover, since C was returned by the algorithm we must have $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$ (Line 13). In conjunction, these facts imply that $C = A \star B$, since both C and $A \star B$ are nonnegative vectors. ◀

Lemma 18 (Running Time of Algorithm 1). *The expected running time of Algorithm 1 is $O(t \log^2 t)$, where $t = \|A \star B\|_0$.*

Algorithm 4**Input:** Nonnegative vectors $A, B \in \mathbf{N}^n$ and a real parameter $\epsilon > 0$ **Output:** $C = A \star B$

```

1   $C \leftarrow (0, \dots, 0)$ 
2  for  $\mu \leftarrow 0, 1, 2, \dots, \infty$  do
3      for  $\nu \leftarrow 0, 1, 2, \dots, \mu$  do
4          repeat  $\mu \cdot 2^{\nu/(1+\epsilon)}$  times
5              Compute  $R$  by Algorithm 3 with parameter  $m = 2^{\mu-\nu}$ 
6              Update  $C \leftarrow \max\{C, R\}$  (coordinate-wise)
7              if  $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$  then return  $C$ 

```

► **Proof.** We first prove that the algorithm terminates with high probability as soon as the outer loop (Line 1) reaches a sufficiently large value. More precisely, let c be the constant from Lemma 13 and fix any iteration of the outer loop with value $m \geq 2ct$. We claim that the algorithm terminates within this iteration with probability at least $1 - t^{-1}$. To this end we analyze the probability of the event $C_z = (A \star B)_z$, for any fixed index z . Recall that C is the coordinate-wise maximum of $2 \log m \geq 2 \log t$ vectors R computed by Algorithm 3. For any such vector R , Lemma 16 guarantees that $R_z = (A \star B)_z$ with probability at least $1 - ct/m \geq \frac{1}{2}$. Hence, the probability that $C_z = (A \star B)_z$ is at least $1 - 2^{-2 \log t} = 1 - t^{-2}$. By a union bound over the t nonzero entries z , the probability that algorithm correctly computes $C = A \star B$ in this iteration is at least $1 - t^{-1}$.

The running time of a single iteration with value m is dominated by the $O(\log m)$ calls to Algorithm 3 taking time $O(m \log m)$. Sampling the hash functions h has negligible cost (by Lemma 13) and so does running the inner-most loop (Line 8). The previous paragraph in particular shows that the algorithm terminates before the η -th iteration after crossing the critical threshold $m \geq 2ct$, with probability at least $1 - t^{-\eta} \geq 1 - 4^{-\eta}$. Hence, we can bound the expected running time by the total time before this threshold ($m < 2ct$) plus the expected time after ($m = 2^\eta \cdot 2ct$) which can be bounded by a geometric series:

$$\sum_{\mu=0}^{\log(2ct)} O(2^\mu \log^2(2^\mu)) + \sum_{\eta=0}^{\infty} 4^{-\eta} \cdot O((2^\eta \cdot t) \cdot \log^2(2^\eta \cdot t)) = O(t \log^2 t). \quad \blacktriangleleft$$

This finished the analysis of Algorithm 1, but not yet the proof of Theorem 2 which additionally claims a tail bound on the running time. To get this additional guarantee, we can modify Algorithm 1 to increase m more carefully; see the pseudocode in Algorithm 4.

Lemma 19 (Correctness and Running Time of Algorithm 4). *Given nonnegative vectors $A, B \in \mathbf{N}^n$ and any parameter $\epsilon > 0$, Algorithm 4 correctly computes their convolution $A \star B$ in expected time $O(t \log^2 t)$, where $t = \|A \star B\|_0$. Moreover, with probability $1 - \delta$ it terminates in time*

$$O\left(t \log^2(t) \cdot \left(\frac{\log(t/\delta)}{\log t}\right)^{1+\epsilon+o(1)}\right).$$

► **Proof.** The correctness proof is exactly as in Lemma 17 and can therefore be omitted. We prove the improved running time bound. Let c be the constant from Lemma 16 and focus on the iterations of the outer loops (Lines 2 and 3) with values $\mu = M$ and $\nu = N$, where

$$N = \left\lceil (1 + \epsilon) \log \left(\frac{\log(t/\delta)}{\log t} \right) \right\rceil \quad \text{and} \quad M = \lceil \log(2ct) \rceil + N.$$

In that case we have $m = 2^{\mu-\nu} \geq 2ct$. We claim that the algorithm terminates in this iteration with probability at least $1 - \delta$. To prove this, we again analyze the probability of the event $C_z = (A \star B)_z$ for any fixed index z . The vector C is the coordinate-wise maximum of all vectors R computed in the inner loop (Line 2) and Lemma 16 proves that the event

$R_z = (A \star B)_z$ happens with probability at least $1 - ct/m \geq \frac{1}{2}$. Since the inner loop is repeated $\mu \cdot 2^{\nu/(1+\epsilon)}$ times, the event $C_z = (A \star B)_z$ happens with probability at least

$$1 - 2^{-\mu \cdot 2^{\nu/(1+\epsilon)}} \geq 1 - 2^{-\log(t) \cdot \log(t/\delta) / \log(t)} = 1 - \frac{\delta}{t}.$$

By a union bound over the t nonzero coordinates z , the algorithm computes $C = A \star B$ (and consequently terminates) with probability at least $1 - \delta$.

Now, to analyze the running time, we have to bound the running time until the algorithm reaches the required values $\mu = M$ and $\nu = N$. The running time of a single execution of the inner-most loop is dominated by the call to Algorithm 3 which takes time $O(m \log m)$ by Lemma 16. Thus, with probability $1 - \delta$ the total running time is bounded by

$$\sum_{\mu=0}^M \sum_{\nu=0}^{\mu} \mu \cdot 2^{\nu/(1+\epsilon)} \cdot O(2^{\mu-\nu} \log(2^{\mu-\nu})) \leq O\left(M^2 \sum_{\mu=0}^M 2^{\mu} \sum_{\nu=0}^{\mu} 2^{-\epsilon\nu/(1+\epsilon)}\right) \leq O(2^M \cdot M^2).$$

Plugging in the definition of M this becomes

$$O(2^M \cdot M^2) = O\left(t \cdot \left(\frac{\log(t/\delta)}{\log t}\right)^{1+\epsilon} \cdot M^2\right) = O\left(t \log^2(t) \cdot \left(\frac{\log(t/\delta)}{\log t}\right)^{1+\epsilon+o(1)}\right).$$

Finally, we derive from the previous paragraph that expected running time is bounded by $O(t \log^2 t)$. Indeed, the total running time exceeds $\ell \cdot t \log^2 t$ with probability $\ll O(\ell^{-3})$, and thus the expected running time is $t \log^2 t \cdot \sum_{\ell=1}^{\infty} (\ell + 1) \cdot O(\ell^{-3}) \leq O(t \log^2 t)$. \blacktriangleleft

This completes the proof of Theorem 2; it suffices to plug in some constant $0 < \epsilon < 1$ into Lemma 19 to obtain the claimed running time $O(t \log^2(t/\delta))$.

4.4 Accelerated Algorithm

We now speed up Algorithm 4 in expectation. The crucial subroutine in that algorithm is Algorithm 3 which computes a good approximation R of $A \star B$. For the improvement we design a similar subroutine which instead computes a good approximation of $A \star B - C$; see Algorithm 5.

Lemma 20 (Correctness and Running Time of Algorithm 5). *Given vectors $A, B, C \in \mathbf{Z}^n$ such that $A \star B - C$ is nonnegative, and any parameter m , Algorithm 5 runs in time $O(m \log m)$ and computes a vector R such that for every $z \in [n]$:*

- $R_z \leq (A \star B - C)_z$ (always), and
- $R_z < (A \star B - C)_z$ with probability at most $c \log n \cdot \|A \star B - C\|_0 / m$ for some constant c .

► **Proof.** Recall that by Lemma 14 the family of hash functions $h(x) = x \bmod p$ is truly additive, i.e., satisfies $h(x) + h(y) \equiv h(x + y) \pmod{p}$ for all keys x, y . As a consequence, it holds that $X = h(A) \star_p h(B) - h(C) = h(A \star B - C)$ and similarly $Y = h(\partial(A \star B - C))$ and $Z = h(\partial^2(A \star B - C))$; the proofs of these statements are straightforward calculations.

The rest of the proof is very similar to Lemma 16 and we merely sketch the differences. We analogously define vectors V^k by $V_z^k = (A \star B - C)_z$ if $z \equiv k \pmod{p}$ and $V_z^k = 0$ otherwise. Then, by the previous paragraph we have $X_k = \|V^k\|_1$, $Y_k = \|\partial V^k\|_1$ and $Z_k = \|\partial^2 V^k\|_1$. It follows by the same argument, using the sparsity tester (Lemma 4), that the recovered vector R is exactly $R = \sum_k V^k$, where the sum is over all vectors V^k which are at most 1-sparse. The first item is immediate since $\sum_{k \in [p]} V^k = A \star B - C$.

To prove the second item, it suffices to argue that with good probability each nonzero entry z does not collide with any other nonzero entry z' under h . In that case, the vector V^k for $k = h(z)$ is 1-sparse and the algorithm correctly computes $R_z = (A \star B - C)_z$. To see that each index z is likely isolated, we apply the $O(\log n)$ -universality of h (Lemma 14): The probability that z collides with some fixed index z' is at most $O(\log(n)/p) \leq O(\log(n)/m)$. Taking a union bound over the $\|A \star B - C\|_0$ nonzero entries z' yields the claimed bound.

Algorithm 5**Input:** Vectors $A, B, C \in \mathbf{Z}^n$ such that $A \star B - C$ is nonnegative, and a parameter m **Output:** A nonnegative vector $R \leq A \star B - C$, for details see Lemma 20

```

1  Sample a random prime  $p \in [m, 2m]$  and let  $h(x) = x \bmod p$ 
2  Compute  $X \leftarrow h(A) \star_p h(B) - h(C)$ 
3  Compute  $Y \leftarrow h(\partial A) \star_p h(B) + h(A) \star_p h(\partial B) - h(\partial C)$ 
4  Compute  $Z \leftarrow h(\partial^2 A) \star_p h(B) + 2h(\partial A) \star_p h(\partial B) + h(A) \star_p h(\partial^2 B) - h(\partial^2 C)$ 
5  Initialize  $R \leftarrow (0, \dots, 0)$ 
6  for each  $k \in [p]$  do
7      if  $X_k \neq 0$  and  $Y_k^2 = X_k \cdot Z_k$  then
8           $z \leftarrow Y_k / X_k$ 
9           $R_z \leftarrow R_z + X_k$ 
10 return  $R$ 

```

Algorithm 6**Input:** Nonnegative vectors $A, B \in \mathbf{N}^n$ **Output:** $C = A \star B$

```

1   $C \leftarrow (0, \dots, 0)$ 
2  for  $m \leftarrow 1, 2, 4, \dots, \infty$  do
3      repeat  $3 \log \log n$  times
4          Compute  $R$  by Algorithm 3 with inputs  $A, B$  and parameter  $m$ 
5          Update  $C \leftarrow \max\{C, R\}$  (coordinate-wise)
6      repeat  $2 \log m$  times
7          Compute  $R$  by Algorithm 5 with inputs  $A, B, C$  and parameter  $m' = \lceil \frac{m}{\log n} \rceil$ 
8          Update  $C \leftarrow C + R$ 
9      if  $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$  then return  $C$ 

```

Finally, observe that the running time is again dominated by the six calls to FFT in Lines 2 to 4, which take time $O(m \log m)$. Sampling h takes time $\text{polylog}(m)$ and the loop in Line 6 takes linear time. \blacktriangleleft

Next, to obtain the speed-up over Algorithm 4, we combine Algorithms 3 and 5. The rough idea is that Algorithm 4 reaches a good (but imperfect) approximation C of $A \star B$ after only $\log \log n$ iterations of the inner-most loop; after that point $A \star B - C$ is sufficiently sparse so that a few iterations with Algorithm 5 can correct the remaining errors. The resulting algorithm is summarized in Algorithm 6.

Lemma 21 (Correctness of Algorithm 6). *Whenever Algorithm 6 outputs a vector C , then $C = A \star B$ (with error probability 0).*

► Proof. We first prove that the algorithm maintains the invariant $0 \leq C \leq A \star B$. There are two types of updates. First, for a vector R computed by Algorithm 3, the algorithm updates “ $C \leftarrow \max\{C, R\}$ ”. Since R satisfies $0 \leq R \leq A \star B$ by Lemma 16, this update maintains the invariant. Second, for a vector R computed by Algorithm 5, the algorithm update “ $C \leftarrow C + R$ ”. Since R satisfies $0 \leq R \leq A \star B - C$ by Lemma 20, this update also upholds the invariant.

It is easy to conclude that the algorithm outputs the correct solution $C = A \star B$, as this is the only vector $0 \leq C \leq A \star B$ which also satisfies $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$. \blacktriangleleft

Lemma 22 (Running Time of Algorithm 6). *The expected running time of Algorithm 6 is $O(t \log t \log \log n)$, where $t = \|A \star B\|_0$.*

► Proof. For the analysis, we split the execution of the algorithm into two *phases*: The first and initial phase ends as soon as $\|A \star B - C\|_0 \leq t / \log^2 n$, and the second phase ends when the algorithm terminates. To analyze the expected running times of both phases, we assume that the outer loop (Line 2) has reached a value $m \geq 2ct$, where c is the maximum of the

constants in Lemmas 16 and 20. In this case we claim that a single execution of the loop body terminates both phases with probability at least $\frac{3}{4}$.

- 1** For the first phase we analyze the pseudocode in Lines 3 to 5. Fix an arbitrary index $z \in \text{supp}(A \star B)$. For a vector R computed by Algorithm 3 we have $R_z = (A \star B)_z$ with probability at least $1 - ct/m \geq \frac{1}{2}$, by Lemma 16. If any of the vectors R computed in Lines 3 to 5 satisfies $R_z = (A \star B)_z$, then we correctly assign “ $C_z \leftarrow \max\{C_z, R_z\}$ ” in Line 5 (and we never change that entry for the remaining execution of the algorithm). Since the loop runs for $3 \log \log n$ iterations, the probability that C_z remains incorrect is at most $2^{-3 \log \log n} = (\log n)^{-3}$. Therefore, the expected number of incorrectly assigned coordinates is at most $t/\log^3 n$ and by Markov’s inequality that number exceeds $t/\log^2 n$ with probability at most $1/\log n$. This is less than $\frac{1}{8}$ for sufficiently large n .
- 2** For the second phase we analyze the pseudocode in Lines 6 to 8. Assuming that the first phase is finished, we have $\|A \star B - C\|_0 \leq t/\log^2 n$. The argument is similar to the first phase: A vector R computed by Algorithm 5 satisfies $R_z = (A \star B - C)_z$ with probability at least $1 - c \log n \cdot \|A \star B - C\|_0/m' \geq 1 - ct/m \geq \frac{1}{2}$. If any vector of the vectors R computed in Lines 6 to 8 satisfies $R_z = (A \star B - C)_z$ then we correctly update “ $C_z \leftarrow C_z + R_z$ ” (and this entry is unchanged for the remaining execution). The probability that C_z is still incorrect after $2 \log m \geq 2 \log t$ iterations is $2^{-2 \log t} = t^{-2}$. By a union bound over the t nonzero entries z , we have correctly computed $C = A \star B$ after finishing the loop with probability at least $1 - t^{-1}$. For sufficiently large t , this is at least $\frac{7}{8}$.

In combination, with probability $\frac{3}{4}$ both phases finish and therefore the algorithm terminates within a single iteration of the outer loop. Each iteration takes time $O(m \log m \cdot \log \log n)$ (Lemma 16) plus $O(m' \log m' \cdot \log m) = O(m \log^2 m / \log n)$ (Lemma 20). To bound the total running time, we use that only with probability $4^{-\eta}$ the algorithm continues for another η iterations of the outer loop after crossing the critical threshold $m \geq 2ct$. Hence, the expected running time is bounded by $O(t \log t \log \log n)$ before that threshold and by

$$\sum_{\eta=1}^{\infty} 4^{-\eta} \cdot O\left((2^\eta \cdot t) \log(2^\eta \cdot t) \log \log n + (2^\eta \cdot t) \frac{\log^2(2^\eta \cdot t)}{\log n}\right) = O(t \log t \log \log n)$$

after. In total, the expected time is $O(t \log t \log \log n)$ as claimed. \blacktriangleleft

4.5 Las Vegas Length Reduction

As a final step, we can reduce the running time of Lemma 22 by replacing the $\log \log n$ factor with $\log \log t$. To this end, we implement a length reduction which reduces the convolution of arbitrary-length vectors to a small number of convolutions of length-poly(t) vectors. The pseudocode is given in Algorithm 7.

Lemma 23 (Correctness and Running Time of Algorithm 7). *Given nonnegative vectors A, B , Algorithm 7 correctly computes their convolution $A \star B$. The expected running time is $O(t \log t \log \log t)$, where $t = \|A \star B\|$.*

► Proof. We skip the correctness part since the proof is exactly like the correctness argument of Algorithm 1; the only difference here is that X, Y, Z are computed by Algorithm 6 instead of FFT, however, Algorithm 6 is a Las Vegas algorithm and therefore also always correct.

To analyze the running, we start by lower bounding the probability that any iteration terminates the algorithm. We say that a linear hash function h as sampled in Line 3 is *good* if for all distinct $z, z' \in \text{supp}(A \star B)$ and all $\phi, \phi' \in \Phi$ it holds that $h(z) + \phi \not\equiv h(z') + \phi' \pmod{m}$; here Φ is the set in Lemma 13. Following the same arguments as in Section 4.3 one can prove that Algorithm 7 terminates as soon as a good hash function is sampled. Therefore, we now lower bound the probability that a random linear hash function h is good. For fixed z, z', ϕ, ϕ' , the probability that $h(z) + \phi \equiv h(z') + \phi' \pmod{m}$ is at most $O(\frac{1}{m})$. We take a union bound over the $O(t^2)$ choices of z, z', ϕ, ϕ' and conclude that a random function h is

Algorithm 7**Input:** Nonnegative vectors $A, B \in \mathbf{N}^n$ **Output:** $C = A \star B$

```

1  repeat
2      Let  $m = \|A\|_0^3 \cdot \|B\|_0^3$ 
3      Sample a linear hash function  $h : [n] \rightarrow [m]$ 
4      Compute  $X \leftarrow h(A) \star_m h(B)$  by Algorithm 6
5      Compute  $Y \leftarrow h(\partial A) \star_m h(B) + h(A) \star_m h(\partial B)$  by Algorithm 6
6      Compute  $Z \leftarrow h(\partial^2 A) \star_m h(B) + 2h(\partial A) \star_m h(\partial B) + h(A) \star_m h(\partial^2 B)$  by Alg. 6
7      Initialize  $C \leftarrow (0, \dots, 0)$ 
8      for each  $k \in [m]$  do
9          if  $X_k \neq 0$  and  $Y_k^2 = X_k \cdot Z_k$  then
10              $z \leftarrow Y_k / X_k$ 
11              $C_z \leftarrow C_z + X_k$ 
12  if  $\|C\|_1 = \|A\|_1 \cdot \|B\|_1$  then return  $C$ 

```

good with probability at least $1 - O(t^2/m)$. Observe that $\|A\|_0 + \|B\|_0 - 1 \leq t \leq \|A\|_0 \cdot \|B\|_0$, and thus $t^3 \leq m \leq t^6$. Therefore, for sufficiently large t each iteration of the loop terminates the algorithm with probability at least $\frac{1}{2}$.

The running time of each iteration i is dominated by the six convolutions computed by Algorithm 6. Let $T_{i,1}, \dots, T_{i,6}$ denote the running times of these calls, respectively. Moreover, let S_i denote the random variable which indicates whether the i -th iteration takes place (or whether the algorithm has terminated before). By the previous paragraph we have that $\mathbf{P}(S_i = 1) \leq 2^{-i}$. The total running time is bounded by

$$\sum_{i=1}^{\infty} S_i \cdot \sum_{j=1}^6 T_{i,j}.$$

Hence, by linearity of expectation and since the random variables S_i and $T_{i,j}$ are independent, the expected running time is at most

$$\sum_{i=1}^{\infty} \mathbf{E}(S_i) \cdot \sum_{j=1}^6 \mathbf{E}(T_{i,j}) \leq \sum_{i=1}^{\infty} 2^{-i} \cdot O(t \log t \log \log m) = O(t \log t \log \log t).$$

Here, we used the expected time bound from Lemma 22 to bound $\mathbf{E}(T_{i,j})$. ◀

Lemma 23 completes the proof of Theorem 3.

References

- 1 Peyman Afshani, Casper B. Freksen, Lior Kamma, and Kasper G. Larsen. Lower bounds for multiplication via network coding. In *Proceedings of the 46th International Colloquium Automata, Languages, and Programming*, ICALP '19, pages 10:1–10:12. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2019.
- 2 Nir Ailon. A lower bound for Fourier transform computation in a linear model over 2×2 unitary gates using matrix entropy. *Chicago Journal of Theoretical Computer Science*, 19, 05 2013.
- 3 Amihod Amir, Ayelet Butman, and Ely Porat. On the relationship between histogram indexing and block-mass indexing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372, 2014.
- 4 Amihod Amir, Oren Kapah, and Ely Porat. Deterministic length reduction: Fast convolution in sparse data and applications. In *Proceedings of the 18th Symposium on Combinatorial Pattern Matching*, CPM '07, pages 183–194. Springer, 2007.
- 5 Andrew Arnold and Daniel S. Roche. Output-sensitive algorithms for sumset and sparse polynomial multiplication. In *Proceedings of the 40th International Symposium on Symbolic and Algebraic Computation*, ISSAC '15, pages 29–36. ACM, 2015.

- 6 Kyriakos Axiotis, Arturs Backurs, Karl Bringmann, Ce Jin, Vasileios Nakos, Christos Tzamos, and Hongxun Wu. Fast and simple modular subset sum. In *Proceedings of the 4th Symposium on Simplicity in Algorithms*, SOSA '21, pages 57–67. SIAM, 2021.
- 7 Michael Ben-Or and Prasoona Tiwari. A deterministic algorithm for sparse multivariate polynomial interpolation. In *Proceedings of the 20th ACM Symposium on Theory of Computing*, STOC '88, pages 301–309. ACM, 1988.
- 8 Leo I. Bluestein. A linear filtering approach to the computation of discrete Fourier transform. *IEEE Transactions on Audio and Electroacoustics*, 18(4):451–455, 1970.
- 9 Karl Bringmann. A near-linear pseudopolynomial time algorithm for subset sum. In *Proceedings of the 28th ACM-SIAM Symposium on Discrete Algorithms*, SODA '17, pages 1073–1084. SIAM, 2017.
- 10 Karl Bringmann, Nick Fischer, Danny Hermelin, Dvir Shabtay, and Philip Wellnitz. Faster minimization of tardy processing time on a single machine. In *Proceedings of the 47th International Colloquium Automata, Languages, and Programming*, ICALP '20, pages 19:1–19:12. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020.
- 11 Karl Bringmann, Nick Fischer, and Vasileios Nakos. Sparse nonnegative convolution is equivalent to dense nonnegative convolution. In *Proceedings of the 53rd ACM Symposium on Theory of Computing*, STOC '21, pages 1711–1724. ACM, 2021.
- 12 Karl Bringmann and Vasileios Nakos. Top- k -convolution and the quest for near-linear output-sensitive subset sum. In *Proceedings of the 52nd ACM Symposium on Theory of Computing*, STOC '20, pages 982–995. ACM, 2020.
- 13 Karl Bringmann and Vasileios Nakos. Fast n -fold Boolean convolution via additive combinatorics. In *Proceedings of the 48th International Colloquium Automata, Languages, and Programming*, ICALP '21, 2021. To appear.
- 14 Karl Bringmann and Vasileios Nakos. A fine-grained perspective on approximating subset sum and partition. In *Proceedings of the 32th ACM-SIAM Symposium on Discrete Algorithms*, SODA '21, pages 1797–1815. SIAM, 2021.
- 15 Timothy M. Chan and Moshe Lewenstein. Clustered integer 3SUM via additive combinatorics. In *Proceedings of the 47th ACM Symposium on Theory of Computing*, STOC '15, pages 31–40. ACM, 2015.
- 16 Qi Cheng. On the construction of finite field elements of large order. *Finite Fields and Their Applications*, 11(3):358–366, 2005.
- 17 Qi Cheng. Constructing finite field extensions with large order elements. *SIAM J. Discret. Math.*, 21(3):726–730, 2007.
- 18 Richard Cole and Ramesh Hariharan. Verifying candidate matches in sparse and wildcard matching. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, STOC '02, pages 592–601. ACM, 2002.
- 19 Gaspard R. de Prony. Essai expérimental et analytique: Sur les lois de la dilatibilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique Floréal et Plairial*, 1:24–76, 1795.
- 20 Martin Dietzfelbinger. *Universal Hashing via Integer Arithmetic Without Primes, Revisited*, pages 257–279. Springer, 2018.
- 21 Michael J. Fischer and Michael S. Paterson. String matching and other products. *Complexity of Computation*, 7:113–125, 1974.
- 22 Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- 23 Anna Gilbert and Piotr Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.
- 24 Pascal Giorgi, Bruno Grenet, and Armelle Perret du Cray. Essentially optimal sparse polynomial multiplication. In *Proceedings of the 45th International Symposium on Symbolic and Algebraic Computation*, ISSAC '20, pages 202–209. ACM, 2020.
- 25 Haitham Hassanieh, Fadel Adib, Dina Katabi, and Piotr Indyk. Faster GPS via the sparse Fourier transform. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, MobiCom '12, pages 353–364, 2012.
- 26 Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Nearly optimal sparse Fourier transform. In *Proceedings of the 44th ACM Symposium on Theory of Computing*, STOC '12, pages 563–578. ACM, 2012.
- 27 Piotr Indyk. Faster algorithms for string matching problems: matching the convolution bound. In *Proceedings of the 39th IEEE Annual Symposium on Foundations of Computer Science*, FOCS '98, pages 166–173. IEEE Computer Society, 1998.

- 28 Erich Kaltofen and Yagati N. Lakshman. Improved sparse multivariate polynomial interpolation algorithms. In *Proceedings of the 13th International Symposium on Symbolic and Algebraic Computation*, ISAAC '88, pages 467–474. Springer, 1988.
- 29 Erich L. Kaltofen. Fifteen years after dsc and wss2 what parallel computations i do today. In *Proceedings of the 4th International Workshop on Parallel and Symbolic Computation*, PASCO '10, pages 10–17. ACM, 2010.
- 30 Konstantinos Koiliaris and Chao Xu. Faster pseudopolynomial time algorithms for subset sum. *ACM Trans. Algorithms*, 15(3):40:1–40:20, 2019.
- 31 Lei Li. On the arithmetic operational complexity for solving Vandermonde linear equations. *Japan Journal of Industrial and Applied Mathematics*, 17, 2000.
- 32 Bruce G. Lindsay. On the determinants of moment matrices. *The Annals of Statistics*, 17(2):711–721, 1989.
- 33 Michael Monagan and Roman Pearce. Parallel sparse polynomial multiplication using heaps. In *Proceedings of the 34th International Symposium on Symbolic and Algebraic Computation*, ISSAC '09, pages 263–270. ACM, 2009.
- 34 Shanmugavelayutham Muthukrishnan. New results and open problems related to non-standard stringology. In *Proceedings of the 6th Symposium on Combinatorial Pattern Matching*, CPM '95, pages 298–317. Springer, 1995.
- 35 Vasileios Nakos. Nearly optimal sparse polynomial multiplication. *IEEE Trans. Inf. Theory*, 66(11):7231–7236, 2020.
- 36 Victor Pan. *Structured matrices and polynomials: Unified superfast algorithms*. 2001.
- 37 Daniel S. Roche. Adaptive polynomial multiplication. *Proceedings of Milestones in Computer Algebra*, pages 65–72, 2008.
- 38 Daniel S. Roche. What can (and can't) we do with sparse polynomials? In *Proceedings of the 43rd International Symposium on Symbolic and Algebraic Computation*, ISSAC '18, pages 25–30. ACM, 2018.
- 39 Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. In *Proceedings of the 29th IEEE Annual Symposium on Foundations of Computer Science*, FOCS '88, pages 283–290. IEEE Computer Society, 1988.
- 40 Victor Shoup. Searching for primitive roots in finite fields. In *Proceedings of the 22nd ACM Symposium on Theory of Computing*, STOC '90, pages 546–554. ACM, 1990.
- 41 Igor E. Shparlinskii. On primitive elements in finite fields and on elliptic curves. *Mathematics of the USSR-Sbornik*, 71(1):41–50, 1992.
- 42 Joris Van Der Hoeven and Grégoire Lecerc. On the complexity of multivariate blockwise polynomial multiplication. In *Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation*, ISSAC '12, pages 211–218. ACM, 2012.
- 43 Joachim von zur Gathen and Jürgen Gerhard. *Modern Computer Algebra*. Cambridge University Press, 3 edition, 2013.

A Linear Hashing without Primes

In this section we provide the proof for Lemma 13.

Lemma 13 (Linear Hashing without Primes). *Let $n \geq m$ be arbitrary. There is a family of linear hash functions $h : [n] \rightarrow [m]$ with the following three properties.*

- 1 **Efficiency:** *Sampling and evaluating h takes constant time.*
- 2 **Uniform Differences:** *For any distinct keys $x, y \in [n]$ and for any $q \in [m]$, the probability that $h(x) - h(y) \equiv q \pmod{m}$ is at most $O(\frac{1}{m})$.*
- 3 **Almost-Additiveness:** *There exists a constant-size set $\Phi \subseteq [m]$ such that for all keys $x, y \in [n]$ it holds that $h(x) + h(y) \equiv h(x + y) + \phi \pmod{m}$ for some $\phi \in \Phi$.*

For the proof we need another result. We say that a set $A = \{r + ia : i \in [|A|]\} \subseteq \mathbf{Z}$ is an *arithmetic progression* with *step-width* a . The following lemma proves that two arithmetic progressions with coprime step-widths are as uncorrelated as possible.

Lemma 24. *Let A and B be arithmetic progressions with coprime step-widths a and b , respectively. Then $|A \cap B| \leq \min\{\frac{|A|-1}{b}, \frac{|B|-1}{a}\} + 1$.*

► **Proof.** We may assume that $A = \{0, a, \dots, (|A| - 1)a\}$ and $B = \{0, b, \dots, (|B| - 1)b\}$ (remove all points before the first common element of A and B and shift such that the

first common element becomes zero). Since a and b are coprime, the intersection $A \cap B$ consists only of multiples of ab and thus $|A \cap B| \leq \lfloor \frac{(|A|-1)a}{ab} \rfloor + 1$. The same bound holds symmetrically for B . ◀

► **Proof of Lemma 13.** First, assume that m is odd. Let N be the smallest power of two larger than $n \cdot m$. We then define the family of hash functions as $h(x) = (ax \bmod N) \bmod m$, where $a \in [N]$ is a random *odd* number. We will now prove the three claimed properties for this family.

1 Efficiency: Sampling h only involves constructing N and sampling a random odd number. Both operations take constant time in the Word RAM model. Evaluating h is also in constant time.

3 Almost-Additiveness: Fix any keys $x, y \in [n]$. Then for one of the two choices $\phi \in \{0, N\}$ it holds that $(ax \bmod N) + (ay \bmod N) = (a(x+y) \bmod N) + \phi$. By reducing this equation modulo m , it follows that $\Phi = \{0, N \bmod m\}$ is a suitable choice.

2 Uniform Differences: To prove that h satisfies the uniform difference property, it suffices to prove that h is $O(1)$ -uniform, that is, $\mathbf{P}(h(z) = \psi) \leq O(\frac{1}{m})$ for all $z \in [n]$ and $\psi \in [m]$. Indeed, by the previous paragraph we have $h(x) - h(y) \equiv q \pmod{m}$ only if $h(x-y) \equiv q - \phi \pmod{m}$ for some $\phi \in \Phi$. Taking a union bound over the constant number of elements ϕ then yields the claim.

To check that h is $O(1)$ -uniform, we write $z = 2^k \cdot w$ where w is odd. Then $az \bmod N$ is uniformly distributed in $A = \{2^k \cdot i : i \in [2^{-k} \cdot N]\}$. Indeed, by identifying $[N]$ with the finite ring $\mathbf{Z}/N\mathbf{Z}$, A is the smallest additive subgroup of $\mathbf{Z}/N\mathbf{Z}$ which contains z , and thus multiplying with a random unit $a \in (\mathbf{Z}/N\mathbf{Z})^\times$ randomly permutes z within that subgroup. It follows that

$$\mathbf{P}(h(z) = \psi) = \frac{2^k \cdot |A \cap B|}{N},$$

where $B \subseteq [N]$ consists of all numbers equal to ψ modulo m . Observe that A and B are both arithmetic progressions with step-widths 2^k and m , respectively. Recall that m is odd, therefore 2^k and m are coprime and Lemma 24 applies and yields $|A \cap B| \leq \frac{N}{2^k m} + 1$. We finally obtain $\mathbf{P}(h(z) = \psi) \leq \frac{1}{m} + \frac{n}{N} \leq O(\frac{1}{m})$.

Finally, we remove the assumption that m is odd. If m is even, then we simply apply the previous construction for $m-1$ to obtain a linear hash function $h : [n] \rightarrow [m-1]$ and reinterpret this as a function $h : [n] \rightarrow [m]$. It is easy to see that this preserves efficiency and uniform differences, and we claim that it also preserves almost-additiveness. Indeed, for arbitrary keys x, y we know that $h(x) + h(y) \equiv h(x+y) + \phi \pmod{m-1}$ for some $\phi \in \Phi$. Both sides of the equation are integers less than $2m-2$ and hence their images modulo $m-1$ and m , respectively, differ by at most 1. Therefore, we have $h(x) + h(y) \equiv h(x+y) + \phi' \pmod{m}$ for some $\phi' \in \Phi' = \{\phi + \sigma : \phi \in \Phi, \sigma \in \{-1, 0, 1\}\}$. ◀