Towards Optimal Toom-Cook Matrices for Integer and Polynomial Multiplication

Marco Bodrato, Alberto Zanoni

Centro "Vito Volterra" Università di Roma "Tor Vergata" Via Columbia 2 – 00133 Rome, Italy

{bodrato, zanoni}@volterra.uniroma2.it

ISSAC 2007 July 31st

1 Toom-Cook multiplication methods

- Multiplication algorithms and complexity
- Toom-Cook algorithm for polynomials, revisited
- Extension to unbalanced factors

2 Toom matrices

- The working model
- Operations and costs
- Graph searching

3 Results

- Toom-2.5 and Toom-3
- Higher Toom-Cook methods
- Some graphics

Long integer and polynomial multiplication

Some notation

Let **R** be \mathbb{Z} or $\mathbb{Z}[X]$,

$$a(x) = \sum_{i=0}^{d_a} a_i x^i \quad , \quad b(x) = \sum_{i=0}^{d_b} b_i x^i \in \mathbf{R}[x]$$

We want to compute their product

$$c(x) = \sum_{i=0}^{d_c} c_i x^i \in \mathbf{R}[x]$$

$$\deg(a)=d_a$$
 ; $\deg(b)=d_b$; $\deg(c)=d_c=d_a+d_b$

・ 同 ト ・ ヨ ト ・ ヨ ト

æ

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

・ロン ・回と ・ヨン ・ヨン

Long integer and polynomial multiplication The classical Toom-Cook approach

- Тоом-Cook methods concern univariate polynomials multiplication.
- Тоом-*n* method refers to factors having *n* parts each (degrees $d_a = d_b = n 1$).
- It is an absolutely standard procedure to apply these methods to general univariate polynomials and long integers multiplication by a simple base changing.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

イロン イヨン イヨン イヨン

Multiplication algorithms

Many algorithms are known for polynomial multiplication.



 $O(d^2)$

Each one has a different complexity, and its own range where it is the fastest one.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

Multiplication algorithms

Many algorithms are known for polynomial multiplication.

- Naïve
- Karatsuba (1962)



・ロト ・回ト ・ヨト ・ヨト

Each one has a different complexity, and its own range where it is the fastest one.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

イロト イヨト イヨト イヨト

Multiplication algorithms

Many algorithms are known for polynomial multiplication.

 Naïve 		$O(d^2)$
• Karatsuba (1962)	$O(d^{\log_2 3})$

• Schönhage-Strassen (1971) $O(d \log d \log \log d)$

Each one has a different complexity, and its own range where it is the fastest one.

In 2007, Martin Fürer announced a new algorithm that should have a better asymptotic complexity.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

Multiplication algorithms

Many algorithms are known for polynomial multiplication.

- Naïve
- Karatsuba (Тоом-2) (1962)
- Тоом-Cook-n (1963)
- Schönhage-Strassen (1971)

 $O(d^{2})$ $O(d^{\log_{2} 3})$ $O(d^{\log_{n}(2n-1)})$

 $O(d \log d \log \log d)$

・ロン ・回と ・ヨン ・ヨン

Each one has a different complexity, and its own range where it is the fastest one.

We aim to analyse the optimality of Toom-Cook methods within their respective ranges of applicability.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

イロン イヨン イヨン イヨン

Recall on Тоом-*n* algorithm

3 (core) phases

Splitting: choose a base and splitEvaluation

1 Recomposition: shift and add.

Phase 2, some linear algebra

Evaluate polynomials a(x), b(x) in 2n-1 different points $\{v_i\} \in \mathbb{Z}$.

This can be obtained by multiplying a (non square) Vandermonde matrix by the vector of coefficients.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

Recall on Тоом-*n* algorithm

3 (core) phases

- Splitting: choose a base and split
- **2** Evaluation: 2 × matrix-vector multiplication
- Multiplication
- **1** Recomposition: shift and add.

Phase 3, recursive application

▶ see unbalanced

Evaluate the product by multiplying factors evaluations. $c(v_i) = a(v_i) \cdot b(v_i)$ (degree n - 1) × (degree n - 1) \rightsquigarrow degree 2n - 2. (n parts) × (n parts) $\rightsquigarrow 2n - 1$ parts. $\Rightarrow 2n - 1$ multiplications.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

・ロン ・回 と ・ ヨ と ・ ヨ と

Recall on Тоом-*n* algorithm

3 (core) phases

- Splitting: choose a base and split
- **2** Evaluation: 2 × matrix-vector multiplication
- **③** Multiplication: $(2n 1) \times$ smaller multiplication
- Interpolation
- **1** Recomposition: shift and add.

Phase 4, some more linear algebra

Interpolate to obtain coefficient of the product polynomial.

Obtain this by multiplying the inverse of a (square) Vandermonde matrix by the vector of evaluations.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

・ロン ・回 と ・ ヨ と ・ ヨ と

Recall on Тоом-*n* algorithm

3 (core) phases

- Splitting: choose a base and split
- **2** Evaluation: 2 × matrix-vector multiplication
- **③** Multiplication: $(2n 1) \times$ smaller multiplication
- Interpolation: inverse matrix-vector multiplication
- Secomposition: shift and add.

Phases 2 and 4 are critical

Splitting order *n* results in (2n-1) multiplications in phase 3, and asymptotic behaviour $\Theta(d^{\log_n(2n-1)})$. Rigidly. The hidden constant is determined by the evaluation/interpolation points and operation sequences for phases 2 and 4.

Multiplication algorithms and complexity Toom-Cook algorithm for polynomials, revisited Extension to unbalanced factors

Unbalanced operands

Factors with different degrees

Тоом-(n+m)/2

▲ back to balanced

 $(\text{degree } n-1) \times (\text{degree } m-1) \rightsquigarrow \text{degree } n+m-2$ $(n \text{ parts}) \times (m \text{ parts}) \rightsquigarrow n+m-1 \text{ parts}$

Toom methods can thus be applied also to polynomials with different degrees. The evaluation phase depends on m and n separately, while the interpolation phase only on n + m.

Тоом-2.5

Unbalanced Tooм-3

・ロト ・回ト ・ヨト ・ヨト

$$\begin{array}{c|c} (\deg 2) \times (\deg 1) \rightsquigarrow \deg 3 \\ (3 \text{ parts}) \times (2 \text{ parts}) \rightsquigarrow 4 \text{ parts} \end{array} \middle| \begin{array}{c} (\deg 3) \times (\deg 1) \rightsquigarrow \deg 4 \\ (4 \text{ parts}) \times (2 \text{ parts}) \rightsquigarrow 5 \text{ parts} \end{array} \right.$$

The working model Operations and costs Graph searching

Some examples for basic cases

The matrices of Toom-2.5 and Toom-3 interpolation phase are

$$A_{2.5} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad ; \qquad A_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 16 & 8 & 4 & 2 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Theorem

For $n \ge 3$, det (A_n) is not a power of 2 (one division is needed).

Theorem

Let A_n be generated by $\{\infty, 1, -1, v_4, \dots, v_{2n-2}, 0\}$. At most 2n-5 divisions are needed in the interpolation phase.

The working model Operations and costs Graph searching

The setting

GOAL: we want the best (most efficient) sequence of elementary operations on rows to transform matrix A_n into identity.

- There are ∞ possible inversion sequences (IS).
- We restrict the admissible operations by defining two criteria.
- They define a finite "model" such that an exhaustive search is possible.
- We describe this model as a weighted graph.
- The goal is reached by solving a shortest path problem on the graph.

イロン イヨン イヨン イヨン

The working model Operations and costs Graph searching

Some useful definitions

For a square matrix M:

 $\begin{array}{rcl} M[i,j] & : & \text{the entry in position } (i,j) \\ M^{(i)} & : & \text{the } i^{th} \text{ line} \\ M^{[j]} & : & \text{the } j^{th} \text{ column} \end{array}$

Definition

The **support** of $M^{(i)}$ is the set $s(M^{(i)})$ of column indexes $j \in \mathbb{N}$ such that $M[i,j] \neq 0$. Similarly for $M^{[i]}$. The **support** of M is the set s(M) of pairs $(i,j) \in \mathbb{N} \times \mathbb{N}$ such that $M[i,j] \neq 0$.

 ${}^{\#}M^{(i)} = \text{cardinality of } s(M^{(i)}).$ Similarly for ${}^{\#}M^{[i]}$ and ${}^{\#}M.$

・ロン ・回と ・ヨン ・ヨン

The working model Operations and costs Graph searching

The model criteria

$$\cdots \to M \xrightarrow{(i)} \widetilde{M} \to \cdots \to I$$

(A) Support reduction :

$$({}^{\#}\widetilde{M}^{(i)} < {}^{\#}M^{(i)}) \land (M[i,j] = 0 \Rightarrow \widetilde{M}[i,j] = 0)$$

At least one more zero entry. "Old" 0 entries are not modified.

(B) Regularisation : $\widetilde{M}[i,j]/M[i',j] = \widetilde{M}[i,j']/M[i',j']$. More entries differing from the corresponding ones in another line by a common multiplicative factor than before.

Example (A,B): in A_3 ,

・ 同 ト ・ ヨ ト ・ ヨ ト

The working model Operations and costs Graph searching

Operations we count on for linear algebra

Linear combinations

 $l_i \leftarrow (c_i \cdot l_i + c_j \cdot l_j)/d_i$, where c_i, c_j, d_i are "small" constants.

"small" actually means fixed: asymptotically small. Typically fits in 1 WORD.

Basic on long operands: linear operations

	C _i	$ c_j $	di	cost
Add/Sub	1	1	1	STEP
I.c of first type	1	2 ^k	1	STEP + (_1_2)
	2 ^{<i>k</i>}	1	1	STEP + (_1_2)
I.c of second type	1	$\neq 2^k$	1	STEP + (_1_X)
	$\neq 2^k$	1	1	STEP + (_1_X)
Division by 2^k (shift)	1	0	2 ^k	SHIFT
Exact division	1	0	$\neq 2^k$	DIV

・ 回 と ・ ヨ と ・ モ と …

э

The working model Operations and costs Graph searching

The Тоом graph

Let G = (V, E, w) the weighted graph such that

- V is the set of matrices obtained by A_n with →* subject to criteria (A) and (B).
- **2** *E* is the set of edges such that $(M, \widetilde{M}) \in E \Leftrightarrow \widetilde{M}$ can be obtained by *M* by means of an admissible linear combination.

The working model Operations and costs Graph searching

The Тоом graph

Let G = (V, E, w) the weighted graph such that

- V is the set of matrices obtained by A_n with →* subject to criteria (A) and (B).
- **2** *E* is the set of edges such that $(M, \widetilde{M}) \in E \Leftrightarrow \widetilde{M}$ can be obtained by *M* by means of an admissible linear combination.

Definition (weight function)

For $\varepsilon \in E$, $w(\varepsilon)$ is the cost of the corresponding linear combination. For a chain C, $w(C) = \sum_{\varepsilon \in C} w(\varepsilon)$. $w(M) = \min_{C(M,I)} \{w(C)\}$

・ロン ・回 と ・ ヨ と ・ ヨ と

Toom-Cook multiplication methods
Toom matrices
ResultsThe working model
Operations and costs
Graph searchingExample (Karatsuba graph):Let $(v_1 = \infty, v_2 = 1, v_3 = 0)$ $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ $\stackrel{\varepsilon_1}{\longrightarrow}$ $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ $\varepsilon_2 \downarrow$ $\downarrow \varepsilon_3$ $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ $\stackrel{\varepsilon_4}{\longrightarrow}$ I

Example (Knuth graph): Let $(v_1 = \infty, v_2 = -1, v_3 = 0)$

The working model Operations and costs Graph searching

Heuristics for search pruning

We use a recursive function f to visit G, keeping some vertexes for some time to benefit from already made evaluations.

The working model Operations and costs Graph searching

Heuristics for search pruning

We use a recursive function f to visit G, keeping some vertexes for some time to benefit from already made evaluations.

- We make **estimates** (from below) e(M) of w(M) by exploiting various heuristics (matrix support cardinality, determinant value, submatrices, etc).
- We introduce a **threshold** t (parameter for f) to avoid analysing not interesting subgraphs. If e(M) > t the subgraph under M is not analysed (no better IS can be drawn from it).
- t is updated while f visits G: if $M \xrightarrow{\varepsilon} \widetilde{M}$ and f(M, t) calls itself, then the recursive call is $f(\widetilde{M}, t w(\varepsilon))$.

The working model Operations and costs Graph searching

Heuristics for search pruning

We use a recursive function f to visit G, keeping some vertexes for some time to benefit from already made evaluations.

- We make estimates (from below) e(M) of w(M) by exploiting various heuristics (matrix support cardinality, determinant value, submatrices, etc).
- We introduce a **threshold** t (parameter for f) to avoid analysing not interesting subgraphs. If e(M) > t the subgraph under M is not analysed (no better IS can be drawn from it).
- t is updated while f visits G: if $M \xrightarrow{\varepsilon} \widetilde{M}$ and f(M, t) calls itself, then the recursive call is $f(\widetilde{M}, t w(\varepsilon))$.

The working model Operations and costs Graph searching

Heuristics for search pruning

We use a recursive function f to visit G, keeping some vertexes for some time to benefit from already made evaluations.

- We make estimates (from below) e(M) of w(M) by exploiting various heuristics (matrix support cardinality, determinant value, submatrices, etc).
- We introduce a **threshold** t (parameter for f) to avoid analysing not interesting subgraphs. If e(M) > t the subgraph under M is not analysed (no better IS can be drawn from it).
- t is updated while f visits G: if $M \xrightarrow{\varepsilon} \widetilde{M}$ and f(M, t) calls itself, then the recursive call is $f(\widetilde{M}, t w(\varepsilon))$.

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-2.5 optimal IS

 $A_{2.5}$ generated by $\{\infty, -1, 1, 0\}$, with det $(A_{2.5}) = 2$. A Toom-graph with 17 nodes was built. The weight is

 $4 \cdot \text{STEP} + \text{SHIFT}$

$$A_{2.5} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{2 = -3} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{2 \gg (1)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{3 = -2}_{2 = -4} I$$

There are 16 minimal equivalent IS.

・ロト ・回ト ・ヨト ・ヨト

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-3 optimal IS

 A_3 generated by $\{\infty, 2, -1, 1, 0\}$, with det $(A_3) = 12$.

The IS implemented in GMP 4.2.1 uses both criteria. Its weight is

 $w_{GMP} = 8 \cdot \text{STEP} + \text{DIV} + 2 \cdot \text{SHIFT} + 2 \cdot (-1-2)$

・ロト ・回ト ・ヨト ・ヨト

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-3 optimal IS

 A_3 generated by $\{\infty, 2, -1, 1, 0\}$, with det $(A_3) = 12$. The IS implemented in GMP 4.2.1 uses both criteria. Its weight is

 $w_{GMP} = 8 \cdot \text{STEP} + \text{DIV} + 2 \cdot \text{SHIFT} + 2 \cdot (_1_2)$

The solution we found uses only criterion (A) and its weight is

 $w_{BZ} = 8 \cdot \text{STEP} + \text{DIV} + \text{SHIFT} + \min(_1_X, \text{SHIFT}) + _1_2$

depending on which of _1_X, SHIFT is smaller.

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Toom-3 optimal IS, when SHIFT < _1_X

æ

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Toom-3.5 $(4 \times 3 \text{ or } 5 \times 2 \text{ unbalanced multiplications})$

 $A_{3.5}$ generated by $\{\infty, 2, -2, 1, -1, 0\}$. The weight is

$$12 \cdot \text{STEP} + 2 \cdot \text{DIV} + 2 \cdot \text{SHIFT} + 2 \cdot (_1_2)$$

$$A_{3.5} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 32 & 16 & 8 & 4 & 2 & 1 \\ -32 & 16 & -8 & 4 & -2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

One regularisation step (B) is needed.

・ロン ・回 と ・ ヨ と ・ ヨ と

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-4

$$(4 \times 4 \text{ or } 5 \times 3 \text{ or } 6 \times 2)$$

$$A_4$$
 generated by $\left\{\infty, 2, 1, -1, \frac{1}{2}, -\frac{1}{2}, 0\right\}$. The weight is

 $18 \cdot \text{STEP} + 3 \cdot \text{DIV} + \text{SHIFT} + \min(-1 X, \text{SHIFT}) + 2 \cdot (-1 X) + 4 \cdot (-1 2)$

$$A_4=egin{pmatrix} 1&0&0&0&0&0&0\ 64&32&16&8&4&2&1\ 1&1&1&1&1&1&1\ 1&-1&1&-1&1&1\ 1&2&4&8&16&32&64\ 1&-2&4&-8&16&-32&64\ 0&0&0&0&0&1\ \end{pmatrix}$$

One regularisation step (B) is used.

・ロン ・回と ・ヨン・

æ

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-4.5

 $(5 \times 4 \text{ or } 6 \times 3 \text{ or } 7 \times 2)$

$$\mathcal{A}_{4.5}$$
 generated by $\left\{\infty,-1,-2,\frac{1}{2},1,2,-\frac{1}{2},0\right\}.$ The weight is

 $22 \cdot \text{STEP} + 4 \cdot \text{DIV} + \text{SHIFT} + 3 \cdot (-1 X) + 6 \cdot (-1 2)$

$$A_{4.5} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ -128 & 64 & -32 & 16 & -8 & 4 & -2 & 1 \\ 1 & 2 & 4 & 8 & 16 & 32 & 64 & 128 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 128 & 64 & 32 & 16 & 8 & 4 & 2 & 1 \\ 1 & -2 & 4 & -8 & 16 & -32 & 64 & -128 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-5

$(5 \times 5 \text{ or } 6 \times 4 \text{ or } 7 \times 3 \text{ or } 8 \times 2)$

$$A_5$$
 generated by $\left\{\infty, -2, \frac{1}{2}, 4, 2, -1, 1, -\frac{1}{2}, 0\right\}$. The weight is

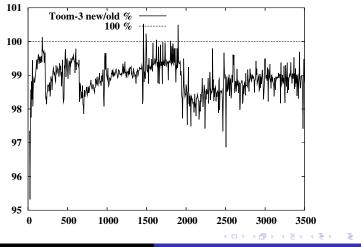
 $32 \cdot \text{STEP} + 5 \cdot \text{DIV} + 2 \cdot \text{SHIFT} + 6 \cdot (-1 X) + 8 \cdot (-1 2)$

・ロン ・回 と ・ ヨ と ・ ヨ と

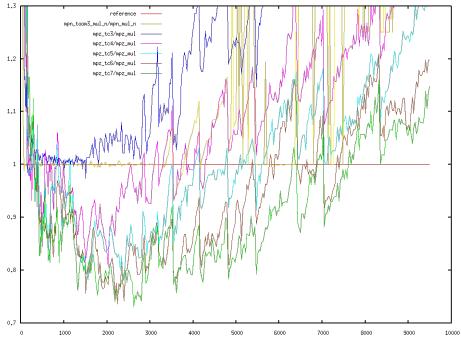
Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Тоом-3 gain

We implemented new GMP code for Toom-3 with the new IS.



Marco Bodrato, Alberto Zanoni Towards Optimal Toom-Cook Matrices



- -) 4 (-

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

Related results

Complete matrix inversion sequences for Tоом-3.5, Tоом-4, Tоом-4.5, Tоом-5 in What about Toom-Cook matrices optimality ? Technical Report 605, Centro "Vito Volterra", Università di Roma "Tor Vergata", October 2006. http://bodrato.it/papers/#CIVV2006.

Analysis for evaluation sequences,

result for Tooм-3 in

Towards Optimal Toom-Cook Multiplication for Univariate and Multivariate Polynomials in Characteristic 2 and 0 presented at WAIFI 2007, Madrid, España, June 21-22, 2007. http://bodrato.it/papers/#WAIFI2007.

Toom-2.5 and Toom-3 Higher Toom-Cook methods Some graphics

That's all folks !

Thank you very much for your very kind attention

Questions ?

Presentation will be available on the web: http://bodrato.it/papers/#ISSAC2007,

released under a CreativeCommons BY-NC-SA licence.

