Memory-efficient polynomial arithmetic

```
Pascal Giorgi^1 Bruno Grenet^1 Daniel S. Roche^2 LIP, ÉNS de Lyon — 11 apr. 2019
```

¹ LIRMM, Université de Montpellier

² CS Department, US Naval Academy

Polynomial arithmetic

- Multiplication: M(n)
 - Naïve: $2n^2 + 2n 1$
 - Karatsuba: $< 6.5 n^{\log_2 3}$
 - Toom-3: $< 18.75 n^{\log_3 5}$
 - FFT-based: $4.5n \log n + O(n)$ or $O(n \log n \log \log n)$

Polynomial arithmetic

- Multiplication: M(n)
 - Naïve: $2n^2 + 2n 1$
 - Karatsuba: $< 6.5 n^{\log_2 3}$
 - Toom-3: $< 18.75 n^{\log_3 5}$
 - FFT-based: $4.5n \log n + O(n)$ or $O(n \log n \log \log n)$
- Other tasks:
 - Euclidean division: 5M(n) + o(M(n))
 - GCD: $O(M(n) \log n)$
 - Evaluation & interpolation: $O(M(n) \log n)$
 - Power series computations: O(M(n)) or $O(M(n) \log n)$
 - . .

Polynomial arithmetic

- Multiplication: M(n)
 - Naïve: $2n^2 + 2n 1$
 - Karatsuba: $< 6.5 n^{\log_2 3}$
 - Toom-3: $< 18.75 n^{\log_3 5}$
 - FFT-based: $4.5n \log n + O(n)$ or $O(n \log n \log \log n)$
- Other tasks:
 - Euclidean division: 5M(n) + o(M(n))
 - GCD: $O(M(n) \log n)$
 - Evaluation & interpolation: $O(M(n) \log n)$
 - Power series computations: O(M(n)) or $O(M(n) \log n)$
 - ...

What about space complexity?

Space complexity of polynomial arithmetic

- Quadratic multiplication algorithm: $O(1)^1$
- Karatsuba, Toom-3, FFT: O(n)
- Other tasks: often O(n)

1. Models to be defined later.

Space complexity of polynomial arithmetic

- Quadratic multiplication algorithm: $O(1)^1$
- Karatsuba, Toom-3, FFT: *O*(*n*)
- Other tasks: often O(n)
- Improvements on Karatsuba's algorithm:
 - Thomé (2002): $n + O(\log n)$
 - Roche (2009): *O*(log *n*)
 - \rightarrow time complexity multiplied by a constant

^{1.} Models to be defined later.

Space complexity of polynomial arithmetic

- Quadratic multiplication algorithm: $O(1)^{1}$
- Karatsuba, Toom-3, FFT: O(n)
- Other tasks: often O(n)
- Improvements on Karatsuba's algorithm:
 - Thomé (2002): $n + O(\log n)$
 - Roche (2009): *O*(log *n*)
 - ightarrow time complexity multiplied by a constant
- Improvements on FFT-based algorithms:
 - Roche (2009): O(1) if $n = 2^k$
 - Harvey & Roche (2010): *O*(1)
 - ightarrow time complexity multiplied by a constant

^{1.} Models to be defined later.

- \rightarrow *Standard* registers of size $O(\log n)$
- $\rightarrow \textit{Algebraic} \text{ registers containing one coefficient}$

- \rightarrow *Standard* registers of size $O(\log n)$
- ightarrow Algebraic registers containing one coefficient
- Read-only input / write-only output
 - (Close to) classical complexity theory
 - Lower bound $\Omega(n^2)$ on time \times space for multiplication

- \rightarrow Standard registers of size $O(\log n)$
- → Algebraic registers containing one coefficient
- Read-only input / write-only output
 - (Close to) classical complexity theory
 - Lower bound $\Omega(n^2)$ on time \times space for multiplication
- Read-only input / read-write output
 - Thomé (2002), Roche (2009) and Harvey & Roche (2010)
 - Reasonable from a programmer's viewpoint

- \rightarrow *Standard* registers of size $O(\log n)$
- ightarrow Algebraic registers containing one coefficient
- Read-only input / write-only output
 - (Close to) classical complexity theory
 - Lower bound $\Omega(n^2)$ on time \times space for multiplication
- Read-only input / read-write output
 - Thomé (2002), Roche (2009) and Harvey & Roche (2010)
 - Reasonable from a programmer's viewpoint
- Read-write input and output
 - Too permissive in general
 - Variant: inputs must be restored at the end

- \rightarrow *Standard* registers of size $O(\log n)$
- → Algebraic registers containing one coefficient
- Read-only input / write-only output
 - (Close to) classical complexity theory
 - Lower bound $\Omega(n^2)$ on time \times space for multiplication
- ✓ Read-only input / read-write output
 - Thomé (2002), Roche (2009) and Harvey & Roche (2010)
 - Reasonable from a programmer's viewpoint
 - Read-write input and output
 - Too permissive in general
 - Variant: inputs must be restored at the end

- Karatsuba's algorithm:
 - Divide-and-Conquer: $(f_0 + X^{\frac{n}{2}}f_1) \cdot (g_0 + X^{\frac{n}{2}}g_1)$ = $f_0g_0 + ((f_0 + f_1)(g_0 + g_1) - f_0g_0 - f_1g_1)X^{\frac{n}{2}} + f_1g_1X^n$

- Karatsuba's algorithm:
 - Divide-and-Conquer: $(f_0 + X^{\frac{n}{2}}f_1) \cdot (g_0 + X^{\frac{n}{2}}g_1)$ = $f_0g_0 + ((f_0 + f_1)(g_0 + g_1) - f_0g_0 - f_1g_1)X^{\frac{n}{2}} + f_1g_1X^n$
 - Thomé'02: Careful use of n temp. registers + $O(\log n)$ stack
 - Roche'09: half-additive version \leadsto only $O(\log n)$ stack $(h_{\ell} \leftarrow h_{\ell} + fg \text{ where } \deg(h_{\ell}) < \deg(f), \deg(g))$

- Karatsuba's algorithm:
 - Divide-and-Conquer: $(f_0 + X^{\frac{n}{2}}f_1) \cdot (g_0 + X^{\frac{n}{2}}g_1)$ = $f_0g_0 + ((f_0 + f_1)(g_0 + g_1) - f_0g_0 - f_1g_1)X^{\frac{n}{2}} + f_1g_1X^n$
 - Thomé'02: Careful use of n temp. registers + $O(\log n)$ stack
 - Roche'09: half-additive version \leadsto only $O(\log n)$ stack $(h_{\ell} \leftarrow h_{\ell} + fg \text{ where } \deg(h_{\ell}) < \deg(f), \deg(g))$
- FFT-based algorithms:
 - $(F,G) \rightarrow (F(\omega^i), G(\omega^i))_i \rightarrow FG(\omega^i)_i \rightarrow FG$

- Karatsuba's algorithm:
 - Divide-and-Conquer: $(f_0 + X^{\frac{n}{2}}f_1) \cdot (g_0 + X^{\frac{n}{2}}g_1)$ = $f_0g_0 + ((f_0 + f_1)(g_0 + g_1) - f_0g_0 - f_1g_1)X^{\frac{n}{2}} + f_1g_1X^n$
 - Thomé'02: Careful use of n temp. registers + $O(\log n)$ stack
 - Roche'09: half-additive version \leadsto only $O(\log n)$ stack $(h_{\ell} \leftarrow h_{\ell} + fg \text{ where } \deg(h_{\ell}) < \deg(f), \deg(g))$
- FFT-based algorithms:
 - $(F,G) \rightarrow (F(\omega^i), G(\omega^i))_i \rightarrow FG(\omega^i)_i \rightarrow FG$
 - Every \rightarrow is in-place (overwriting) but # points is $1 + \deg(FG)$
 - \rightsquigarrow size $((F(\omega^i), G(\omega^i))_i) = 2$ size(FG)

- Karatsuba's algorithm:
 - Divide-and-Conquer: $(f_0 + X^{\frac{n}{2}}f_1) \cdot (g_0 + X^{\frac{n}{2}}g_1)$ = $f_0g_0 + ((f_0 + f_1)(g_0 + g_1) - f_0g_0 - f_1g_1)X^{\frac{n}{2}} + f_1g_1X^n$
 - Thomé'02: Careful use of n temp. registers + $O(\log n)$ stack
 - Roche'09: half-additive version \leadsto only $O(\log n)$ stack $(h_{\ell} \leftarrow h_{\ell} + fg \text{ where } \deg(h_{\ell}) < \deg(f), \deg(g))$
- FFT-based algorithms:
 - $(F,G) \rightarrow (F(\omega^i), G(\omega^i))_i \rightarrow FG(\omega^i)_i \rightarrow FG$
 - Every \rightarrow is in-place (overwriting) but # points is $1 + \deg(FG)$
 - \rightsquigarrow size $((F(\omega^i), G(\omega^i))_i) = 2$ size(FG)
 - Roche'09: Compute half of the result + recurse
 - Harvey-Roche'10: same with TFT (vdH'04)

Can *every* polynomial multiplication algorithm be performed without extra memory?

Can *every* polynomial multiplication algorithm be performed without extra memory?

- O(1)-space Karatsuba's algorithm?
- What about Toom-Cook algorithm?

Can *every* polynomial multiplication algorithm be performed without extra memory?

- O(1)-space Karatsuba's algorithm?
- What about Toom-Cook algorithm?
- What about other products (short and middle)?

Can *every* polynomial multiplication algorithm be performed without extra memory?

- O(1)-space Karatsuba's algorithm?
- What about Toom-Cook algorithm?
- What about other products (short and middle)?

Results:

- Yes!
- Almost (for other products)

Outline

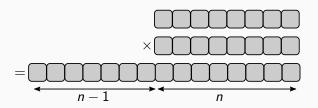
Polynomial products and linear maps

Space-preserving reductions

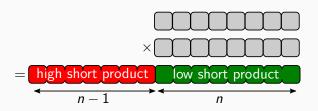
In-place algorithms from out-of-place algorithms

Polynomial products and linear maps

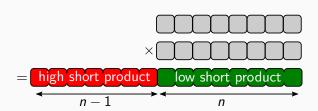
Short product



Short product

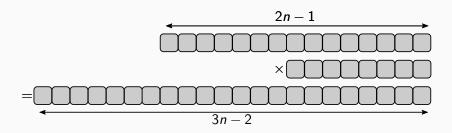


Short product

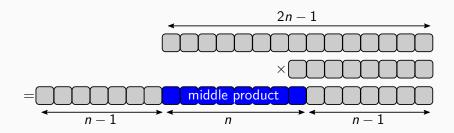


- Low short product: product of truncated power series
- Useful in other algorithms
- Time complexity: M(n)
- Space complexity: O(n)

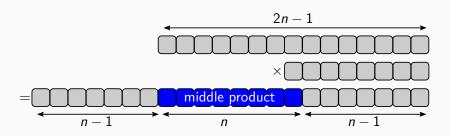
Middle product



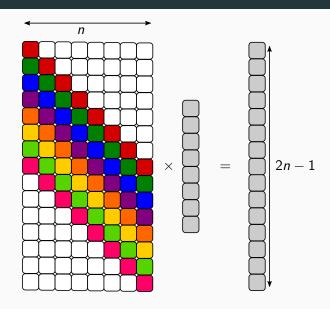
Middle product

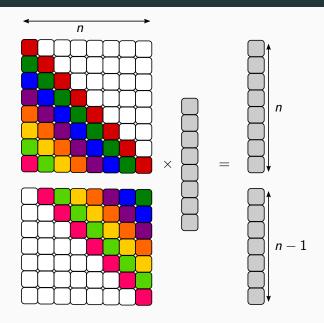


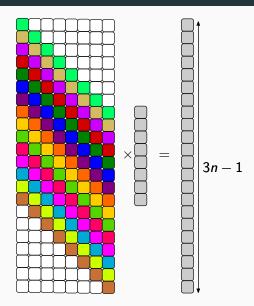
Middle product

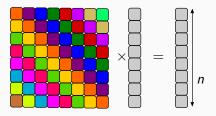


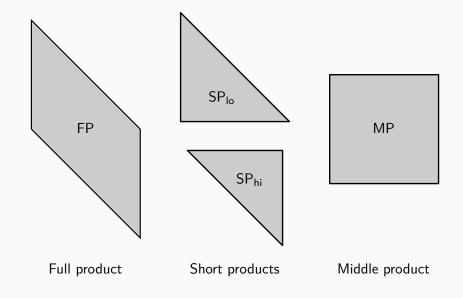
- Useful for Newton iteration
 - $G \leftarrow G(1 GF) \mod X^{2n}$ with $GF = 1 + X^nH$
 - division, square root, . . .
- Time complexity: $M(n) \rightarrow$ Tellegen's transposition
- Space complexity: O(n)
- ullet O(1) space in the most permissive model via transposition of Harvey-Roche algorithm (Bostan-Lecerf-Schost'03)











Space-preserving reductions

Relative difficulties of products

- Without space restrictions:
 - $SP \le FP$ and $FP \le SP_{lo} + SP_{hi}$
 - MP ≡ FP (transposition)
 - $MP \le SP_{lo} + SP_{hi} + (n-1)$ additions

Relative difficulties of products

- Without space restrictions:
 - $SP \le FP$ and $FP \le SP_{lo} + SP_{hi}$
 - MP ≡ FP (transposition)
 - MP \leq SP_{lo} + SP_{hi} + (n-1) additions
- Size of inputs and outputs:
 - FP: $n \times n \rightarrow 2n-1$
 - $SP_{lo}: n \times n \rightarrow n$
 - $SP_{hi}: n-1 \times n-1 \to n-1;$
 - MP: $2n 1 \times n \rightarrow n$

Relative difficulties of products

- Without space restrictions:
 - $SP \le FP$ and $FP \le SP_{lo} + SP_{hi}$
 - MP ≡ FP (transposition)
 - $MP \le SP_{lo} + SP_{hi} + (n-1)$ additions
- Size of inputs and outputs:
 - FP: $n \times n \rightarrow 2n-1$
 - $SP_{lo}: n \times n \rightarrow n$
 - $SP_{hi}: n-1 \times n-1 \to n-1;$
 - MP: $2n 1 \times n \rightarrow n$

Reductions unusable in space-restricted settings!

A relevant notion of reduction

Definitions

- TISP(t(n), s(n)): computable in time t(n) and space s(n)
- $A \leq_c B$: A computable with oracle B and
 - constant number c of calls to oracle
 - negligible extra time
 - without extra space (O(1))
- $A \equiv_c B$: $A \leq_c B$ and $B \leq_c A$

A relevant notion of reduction

Definitions

- TISP(t(n), s(n)): computable in time t(n) and space s(n)
- $A \leq_c B$: A computable with oracle B and
 - constant number c of calls to oracle
 - negligible extra time
 - without extra space (O(1))
- $A \equiv_c B$: $A \leq_c B$ and $B \leq_c A$

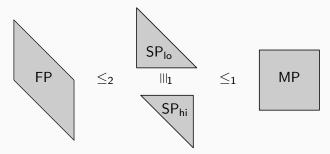
Proposition

If $B \in \mathsf{TISP}(t(n), s(n))$ and $A \leq_c B$, then

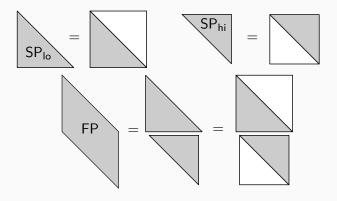
$$A \in \mathsf{TISP}(c\ t(n) + o(t(n)), s(n) + O(1))$$

Results

Theorem

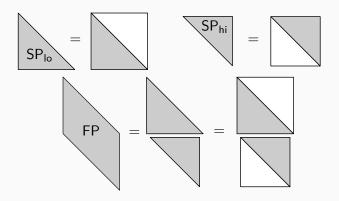


Visual proof



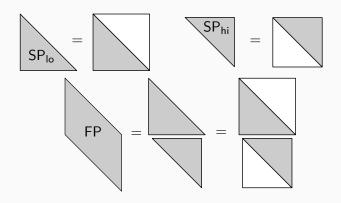
• Use of fake padding (in input, **not** in output!)

Visual proof



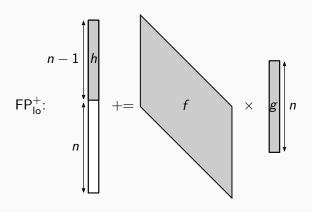
- Use of fake padding (in input, not in output!)
- $SP_{lo}(n) \le MP(n)$; $SP_{hi}(n) \le MP(n-1)$

Visual proof

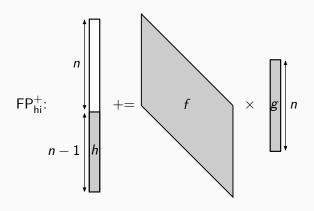


- Use of fake padding (in input, not in output!)
- $SP_{lo}(n) \leq MP(n)$; $SP_{hi}(n) \leq MP(n-1)$
- $\qquad \mathsf{FP}(n) \leq \mathsf{SP}_{\mathsf{hi}}(n) + \mathsf{SP}_{\mathsf{lo}}(n) \leq \mathsf{MP}(n) + \mathsf{MP}(n-1)$

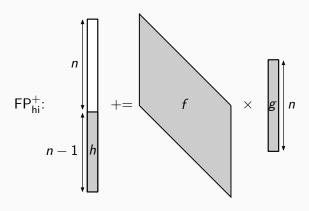
Half-additive full product: $h \leftarrow h + f \cdot g$



Half-additive full product: $h \leftarrow h + f \cdot g$

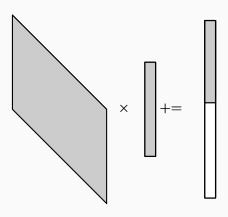


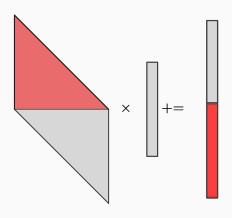
Half-additive full product: $h \leftarrow h + f \cdot g$

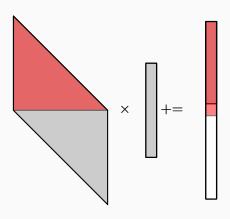


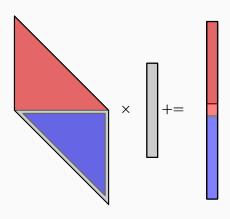
$$\textbf{Remark} \ \mathsf{FP}^+_{\mathsf{lo}} \equiv_1 \mathsf{FP}^+_{\mathsf{hi}}$$

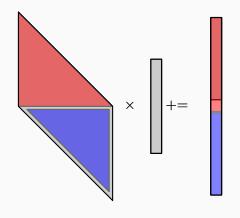
Theorem $FP^+ \leq_{3/2} SP$ and $SP \leq_2 FP^+$







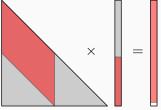




$$\mathsf{FP}^+_{\mathsf{lo}}(\mathit{n}) \leq \mathsf{SP}_{\mathsf{lo}}(\mathit{n}) + \mathsf{SP}_{\mathsf{hi}}(\mathit{n}) + \mathit{n} - 1$$

$$\left(f_0 + X^{\lceil n/2 \rceil} f_1\right) \cdot \left(g_0 + X^{\lceil n/2 \rceil} g_1\right) = f_0 g_0 + X^{\lceil n/2 \rceil} (f_0 g_1 + f_1 g_0) \mod X^n$$

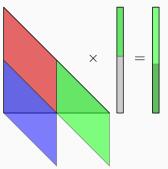
$$\left(f_0 + X^{\lceil n/2 \rceil} f_1\right) \cdot \left(g_0 + X^{\lceil n/2 \rceil} g_1\right) = f_0 g_0 + X^{\lceil n/2 \rceil} \left(f_0 g_1 + f_1 g_0\right) \mod X^n$$



$$\left(f_0 + X^{\lceil n/2 \rceil} f_1\right) \cdot \left(g_0 + X^{\lceil n/2 \rceil} g_1\right) = f_0 g_0 + X^{\lceil n/2 \rceil} \left(f_0 g_1 + f_1 g_0\right) \mod X^n$$

$$\times = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

$$\left(f_0 + X^{\lceil n/2 \rceil} f_1\right) \cdot \left(g_0 + X^{\lceil n/2 \rceil} g_1\right) = f_0 g_0 + X^{\lceil n/2 \rceil} \left(f_0 g_1 + f_1 g_0\right) \mod X^n$$



$$\mathsf{SP}_{\mathsf{lo}}(n) \leq \mathsf{FP}(\lfloor n/2 \rfloor) + \mathsf{FP}_{\mathsf{lo}}^+(\lfloor n/2 \rfloor) + \mathsf{FP}_{\mathsf{hi}}^+(\lceil n/2 \rceil)$$

Converse directions?

- From FP to SP:
 - problem with the output size
 - without space restriction: is $SP(n) \simeq FP(n/2)$?

Converse directions?

- From FP to SP:
 - problem with the output size
 - without space restriction: is $SP(n) \simeq FP(n/2)$?
- From SP to MP:
 - partial result:
 - up to log(n) increase in time complexity
 - techniques from next part
 - without space restriction or in a permissive model
 - FP to MP through Tellegen's transposition principle

In-place algorithms from out-of-place algorithms

Framework

- In-place algorithms parametrized by out-of-place algorithm
 - Out-of-place: Uses *cn* extra space
 - Constant c known to the algorithm

Framework

- In-place algorithms parametrized by out-of-place algorithm
 - Out-of-place: Uses cn extra space
 - Constant c known to the algorithm
- Goal:
 - Space complexity: O(1)
 - Time complexity: closest to the out-of-place algorithm

Framework

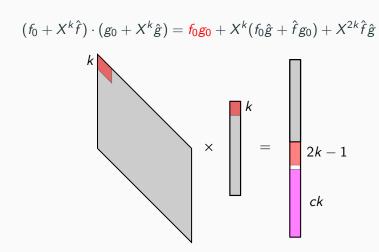
- In-place algorithms parametrized by out-of-place algorithm
 - Out-of-place: Uses cn extra space
 - Constant c known to the algorithm
- Goal:
 - Space complexity: *O*(1)
 - Time complexity: closest to the out-of-place algorithm
- Technique:
 - Oracle calls in smaller size
 - Tail recursive call
 - Fake padding

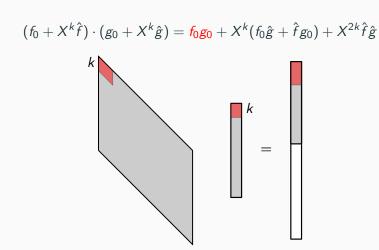
Tail recursion and fake padding

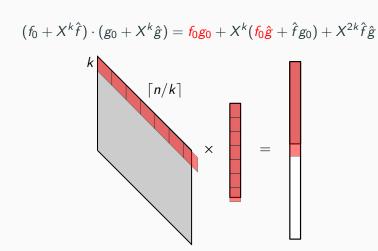
- Tail recursion:
 - Only one recursive call + last (or first) instruction
 - No need of recursive stack \rightsquigarrow avoid $O(\log n)$ extra space

Tail recursion and fake padding

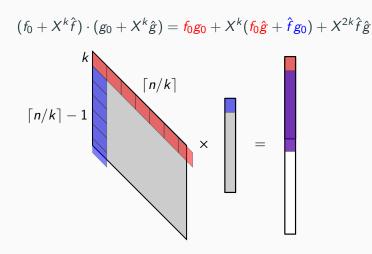
- Tail recursion:
 - Only one recursive call + last (or first) instruction
 - No need of recursive stack \rightsquigarrow avoid $O(\log n)$ extra space
- Fake padding:
 - Pretend to pad inputs with zeroes
 - Make the data structure responsible for it
 - O(1) increase in memory
 - Cf. strides in dense linear algebra
 - OK in inputs, not in outputs!



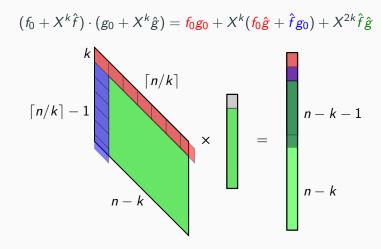




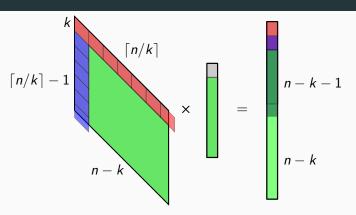
In-place FP⁺ from out-of-place FP



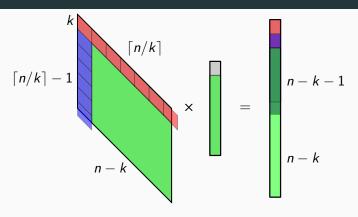
In-place FP⁺ from out-of-place FP



Analysis



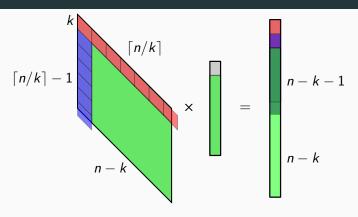
Analysis



•
$$ck + 2k - 1 \le n - k \implies k \le \frac{n+1}{c+3}$$

•
$$T(n) = (2\lceil n/k \rceil - 1)(M(k) + 2k - 1) + T(n - k)$$

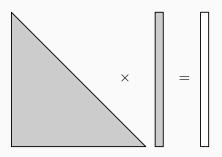
Analysis

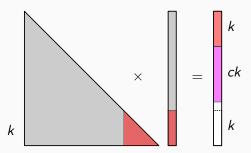


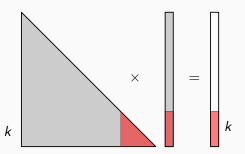
•
$$ck + 2k - 1 \le n - k \implies k \le \frac{n+1}{c+3}$$

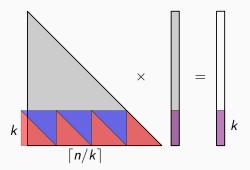
•
$$T(n) = (2\lceil n/k \rceil - 1)(M(k) + 2k - 1) + T(n - k)$$

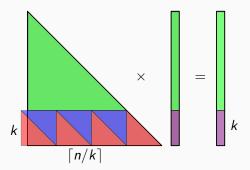
$$T(n) \le (2c+7)M(n) + o(M(n))$$

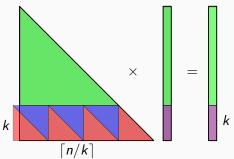




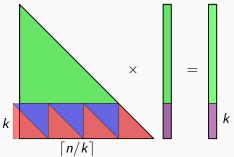






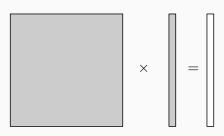


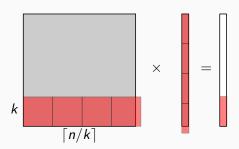
- $k \le n/(c+2)$
- $T(n) = \lceil n/k \rceil M(k) + (\lceil n/k \rceil 1) M(k-1) + 2k(\lceil n/k \rceil 1) + T(n-k)$

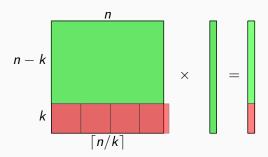


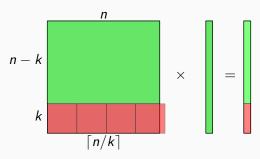
- $k \le n/(c+2)$
- $T(n) = \lceil n/k \rceil M(k) + (\lceil n/k \rceil 1) M(k-1) + 2k(\lceil n/k \rceil 1) + T(n-k)$

$$T(n) \le (2c+5)\mathsf{M}(n) + o(\mathsf{M}(n))$$

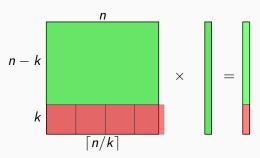








- Recursive call on part of f... but on full g!
- $T(n,m) = \lceil n/k \rceil M(k) + T(n,m-k)$



- Recursive call on part of f... but on full g!
- $T(n,m) = \lceil n/k \rceil M(k) + T(n,m-k)$

$$T(n,n) \le egin{cases} \mathsf{M}(n)\log_{\frac{c+2}{c+1}}(n) + o(\mathsf{M}(n)\log n) & \text{if } \mathsf{M}(n) \text{ is quasi-linear} \\ O(\mathsf{M}(n)) & \text{otherwise} \end{cases}$$

Other operations

Work in progress!

Other operations

Work in progress!

- Use our in-place algorithms as building blocks
 - Newton iteration: division, square root, . . .
 - Evaluation & interpolation
 - \rightarrow (at most) $\log(n)$ increase in complexity

Other operations

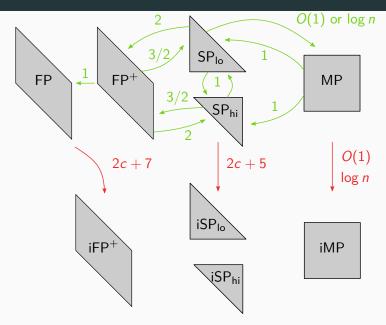
Work in progress!

- Use our in-place algorithms as building blocks
 - Newton iteration: division, square root, ...
 - Evaluation & interpolation
 - \rightarrow (at most) $\log(n)$ increase in complexity

Remark

- In place: division with remainder
- Only quotient or only remainder: not clear
- Main difficulty: size of the output

Summary



- TISP-reductions between polynomial products
- Self-reductions to obtain in-place algorithms

- TISP-reductions between polynomial products
- Self-reductions to obtain in-place algorithms

Comparisons

- Better use specialized in-place algorithms. . .
- ... when they exist!

- TISP-reductions between polynomial products
- Self-reductions to obtain in-place algorithms

Comparisons

- Better use specialized in-place algorithms. . .
- ... when they exist!

Main open problems

- Remove the log(n) for middle product or prove a lower bound
- General result on Tellegen's transposition principle
- What about integer multiplication?

- TISP-reductions between polynomial products
- Self-reductions to obtain in-place algorithms

Comparisons

- Better use specialized in-place algorithms. . .
- ... when they exist!

Main open problems

- Remove the log(n) for middle product or prove a lower bound
- General result on Tellegen's transposition principle
- What about integer multiplication?

Thank you!