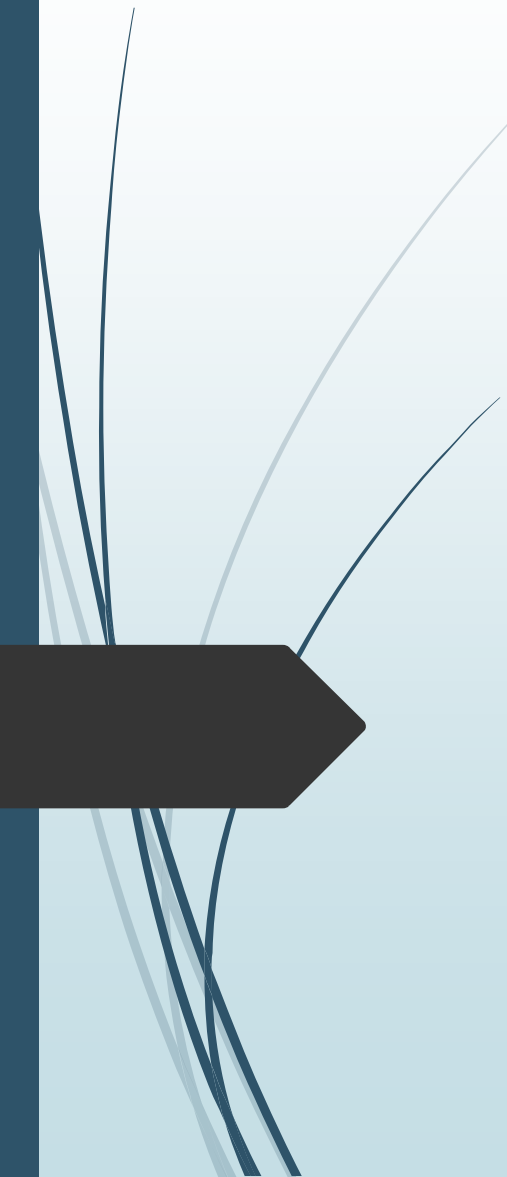


Fast Matrix Multiplication Algorithms

Why should we care?

Complexity of matrix multiplication = Complexity of "almost all" matrix problems

- Solving linear systems
- Evaluating determinants
- LU factorization
- Many more

P. Bürgisser, M. Clausen, M. A. Shokrollahi

Algebraic complexity theory.

A brief history…

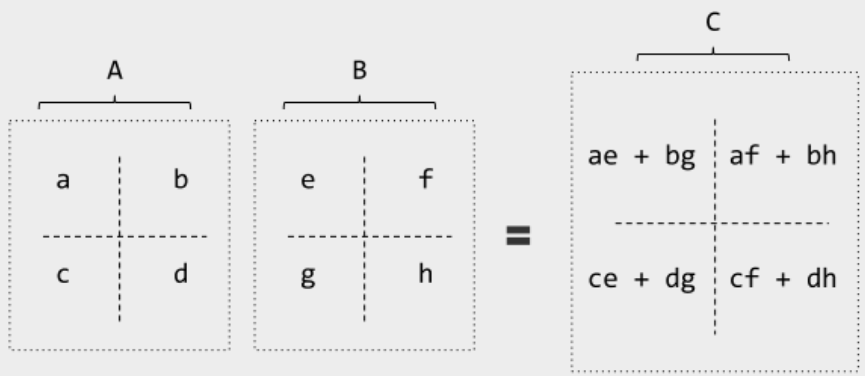
- Until the late 1960's: naïve algorithm, $O(n^3)$

A brief history…

- Until the late 1960's: naïve algorithm, $O(n^3)$
- 1969: Strassen's algorithm, $O(n^{2,808})$ (V. Strassen. Gaussian elimination is not optimal)

Strassen's algorithm

DIVIDE AND CONQUER



$$s = af + bh = P_1 + P_2$$

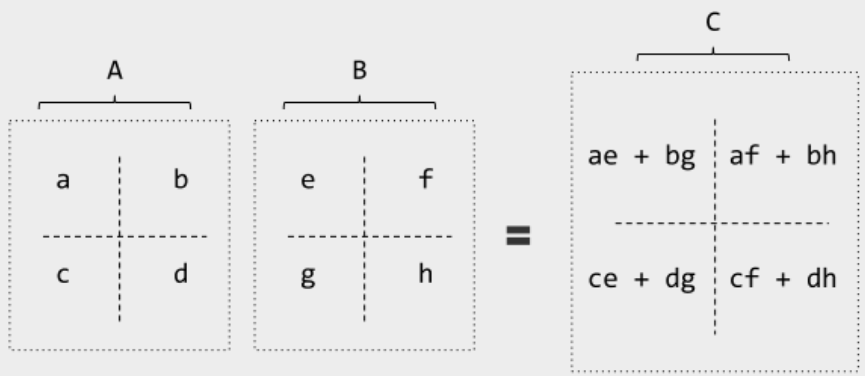
$$t = ce + dg = P_3 + P_4$$

$$r = ae + bg = P_5 + P_4 - P_2 + P_6$$

$$u = cf + dh = P_5 + P_1 - P_3 + P_7$$

Strassen's algorithm

DIVIDE AND CONQUER



$$s = af + bh = P_1 + P_2$$

$$t = ce + dg = P_3 + P_4$$

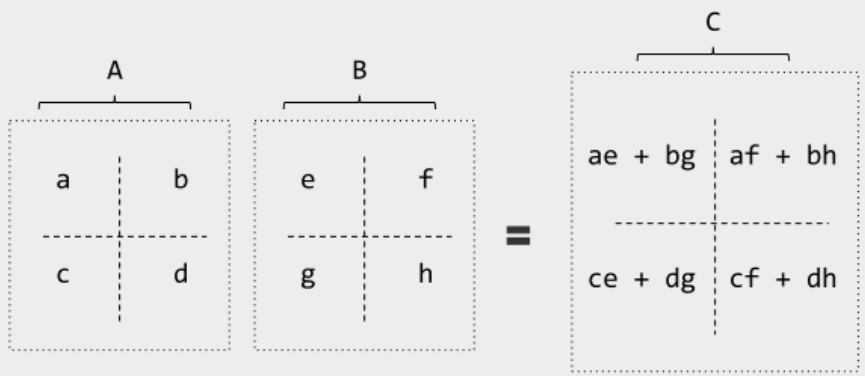
$$r = ae + bg = P_5 + P_4 - P_2 + P_6$$

$$u = cf + dh = P_5 + P_1 - P_3 + P_7$$

1. $P_1 = A_1B_1 = a(f - h) = af - ah$
2. $P_2 = A_2B_2 = (a + b)h = ah + bh$
3. $P_3 = A_3B_3 = (c + d)e = ce + de$
4. $P_4 = A_4B_4 = d(g - e) = dg - de$
5. $P_5 = A_5B_5 = (a + d)(e + h) = ae + ah + de + dh$
6. $P_6 = A_6B_6 = (b - d)(g + h) = bg + bh - dg - dh$
7. $P_7 = A_7B_7 = (a - c)(e + f) = ae + af - ce - cf$

Strassen's algorithm

DIVIDE AND CONQUER



$$s = af + bh = P_1 + P_2$$

$$t = ce + dg = P_3 + P_4$$

$$r = ae + bg = P_5 + P_4 - P_2 + P_6$$

$$u = cf + dh = P_5 + P_1 - P_3 + P_7$$

1. $P_1 = A_1B_1 = a(f - h) = af - ah$
2. $P_2 = A_2B_2 = (a + b)h = ah + bh$
3. $P_3 = A_3B_3 = (c + d)e = ce + de$
4. $P_4 = A_4B_4 = d(g - e) = dg - de$
5. $P_5 = A_5B_5 = (a + d)(e + h) = ae + ah + de + dh$
6. $P_6 = A_6B_6 = (b - d)(g + h) = bg + bh - dg - dh$
7. $P_7 = A_7B_7 = (a - c)(e + f) = ae + af - ce - cf$

$$T(n) = 7T\left(\frac{n}{2}\right) + \Theta(n^2)$$

Time complexity: $O(n^{2.808})$

A brief history…

- Until the late 1960's: naïve algorithm, $O(n^3)$
- 1969: Strassen's algorithm, $O(n^{2,808})$ (V. Strassen. Gaussian elimination is not optimal)
- 1978: Pan, $\omega < 2.796$ (V. Y. Pan. Strassen's algorithm is not optimal)

Bilinear Algorithms

Given two matrices A, B

$P_l = \left(\sum_{i,j} u_{ijl} A[i,j]\right)\left(\sum_{i,j} v_{ijl} B[i,j]\right)$, r linear combinations.

$$AB[i,j] = \sum_l w_{ijl} P_l$$

- The minimum number of products r that a bilinear algorithm can use to compute the product of two $n \times n$ matrices is called the *rank* of $n \times n$ matrix multiplication $R(\langle n, n, n \rangle)$
- The product of two $\mathrm{k}n \times kn$ matrices can be viewed as the product of two $k \times \mathrm{k}$ matrices the entries of which are $n \times n$ matrices
- We can create a recursive algorithm ALG for multiplication of $k \times \mathrm{k}$.
- View the $k^i \times k^i$ as $k \times \mathrm{k}$ matrices with entries $k^{i-1} \times k^{i-1}$ matrices
- The recursive approach using an upper bound of r on $R(\langle k, k, k \rangle)$ gives a bound $\omega < \log_k r$, (the number of additions that one has to do in each step is no more than $3rk^2$
- As long as $r < k^3$ we get a non trivial bound for ω
- Strassen: $k = 2, \mathrm{r} = 7$
- Pan: $k = 70, r = 143640$

A brief history…

- Until the late 1960's: naïve algorithm, $O(n^3)$
- 1969: Strassen's algorithm, $O(n^{2,808})$ (V. Strassen. Gaussian elimination is not optimal)
- 1978: Pan, $\omega < 2.796$ (V. Y. Pan. Strassen's algorithm is not optimal)
- 1979: Bini (border rank), $\omega < 2.78$ (D. Bini, M. Capovani, F. Romani, and G. Lotti. $O(n^{2.7799})$ complexity for n n approximate matrix multiplication)

A brief history…

- Until the late 1960's: naïve algorithm, $O(n^3)$
- 1969: Strassen's algorithm, $O(n^{2,808})$ (V. Strassen. Gaussian elimination is not optimal)
- 1978: Pan, $\omega < 2.796$ (V. Y. Pan. Strassen's algorithm is not optimal)
- 1979: Bini (border rank), $\omega < 2.78$ (D. Bini, M. Capovani, F. Romani, and G. Lotti. $O(n^{2.7799})$ complexity for n n approximate matrix multiplication)
- 1981: Shonhage τ-theorem(asymptotic sum inequality), $\omega < 2.548$

  in the same paper, $\omega < 2.522$ (A. Schonhage. Partial and total matrix multiplication)

Approximate Bilinear Algorithms (ABA)

- In bilinear algorithms the coefficients $u_{ijl}, v_{ijl}, w_{ijl}$ were constants.
- In ABA these coefficients are linear combinations of the integer powers of a indeterminate λ.
- The entries of AB are then only approximately computed:

$$AB[i,j] = \sum_l w_{ijl} P_l + O(\lambda)$$

$O(\lambda)$: linear combination of positive powers of λ.

- When $\lambda \to 0$, then the product is almost exactly.
- The minimum number of products r for an ABA to compute the product of two $n \times n$ matrices, is called *border rank* of a matrix multiplication $\underline{R}(\langle n, n, n \rangle)$

- Bini showed that when dealing with the asymptotic complexity of matrix multiplication, approximate algorithms suffice obtaining bounds for ω
- If $\underline{R}(\langle n, n, n\rangle) \leq r$, then $\omega \leq \log_k r$
- Bini used 10 entry products to multiply a 2×3 matrix with a 3×3 matrix $k = 12$, $r = 1000$
- Shonhage τ-theorem: Suppose we have an upper bound of r on the border rank of computing p independent instances of matrix multiplication with dimensions $k_i \times m_i$ by $m_i \times m_i$ for $i = 1, \ldots, p$. Then $\omega < 3\tau$, where $\sum_i (k_i m_i n_i)^\tau = r$
- In particular he showed that one can approximately compute the product of a 3×1 by 1×3 vector and the product of a 1×4 by 4×1 vector together using only 10 products, whereas any exact bilinear algorithm needs at least 13 products.

A brief history…

- Until the late 1960's: naïve algorithm, $O(n^3)$
- 1969: Strassen's algorithm, $O(n^{2,808})$ (V. Strassen. Gaussian elimination is not optimal)
- 1978: Pan, $\omega < 2.796$ (V. Y. Pan. Strassen's algorithm is not optimal)
- 1979: Bini (border rank), $\omega < 2.78$ (D. Bini, M. Capovani, F. Romani, and G. Lotti. $O(n^{2.7799})$ complexity for n n approximate matrix multiplication)
- 1981: Shonhage τ-theorem(asymptotic sum inequality), $\omega < 2.548$

  in the same paper, $\omega < 2.522$ (A. Sch¨onhage. Partial and total matrix multiplication)
- 1982: Romani, $\omega < 2.517$ (F. Romani. Some properties of disjoint sums of tensors related to matrix multiplication.)

A brief history…

- Until the late 1960's: naïve algorithm, $O(n^3)$
- 1969: Strassen's algorithm, $O(n^{2,808})$ (V. Strassen. Gaussian elimination is not optimal)
- 1978: Pan, $\omega < 2.796$ (V. Y. Pan. Strassen's algorithm is not optimal)
- 1979: Bini (border rank), $\omega < 2.78$ (D. Bini, M. Capovani, F. Romani, and G. Lotti. $O(n^{2.7799})$ complexity for n n approximate matrix multiplication)
- 1981: Shonhage τ-theorem(asymptotic sum inequality), $\omega < 2.548$

  in the same paper, $\omega < 2.522$ (A. Sch¨onhage. Partial and total matrix multiplication)
- 1982: Romani, $\omega < 2.517$ (F. Romani. Some properties of disjoint sums of tensors related to matrix multiplication.)
- 1982: Coppersmith & Winograd, $\omega < 2.496$ (D. Coppersmith and S. Winograd. On the asymptotic complexity of matrix multiplication)

Continue…

- 1986: Strassen, laser method, $\omega < 2.479$ (entirely new approach on the matrix multiplication problem) (V. Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication)

Continue…

- 1986: Strassen, laser method, $\omega < 2.479$ (entirely new approach on the matrix multiplication problem) (V. Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication)
- 1989: Coppersmith & Winograd, combine Strassen's laser method with a novel from analysis based on large sets avoiding arithmetic progression, $\omega < 2.376$ (D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions.)

Continue…

- 1986: Strassen, laser method, $\omega < 2.479$ (entirely new approach on the matrix multiplication problem) (V. Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication)
- 1989: Coppersmith & Winograd, combine Strassen's laser method with a novel from analysis based on large sets avoiding arithmetic progression, $\omega < 2.376$ (D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions.)
- 2003: Cohn & Umans: group theoretic framework for designing and analyzing matrix multiplication algorithms

Continue…

- 1986: Strassen, laser method, $\omega < 2.479$ (entirely new approach on the matrix multiplication problem) (V. Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication)
- 1989: Coppersmith & Winograd, combine Strassen's laser method with a novel from analysis based on large sets avoiding arithmetic progression, $\omega < 2.376$ (D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions.)
- 2003: Cohn & Umans: group theoretic framework for designing and analyzing matrix multiplication algorithms
- 2005: Cohn, Umans, Kleinberg, Szegedy, $O(n^{2.41})$ (H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans. Group-theoretic algorithms for matrix multiplication)

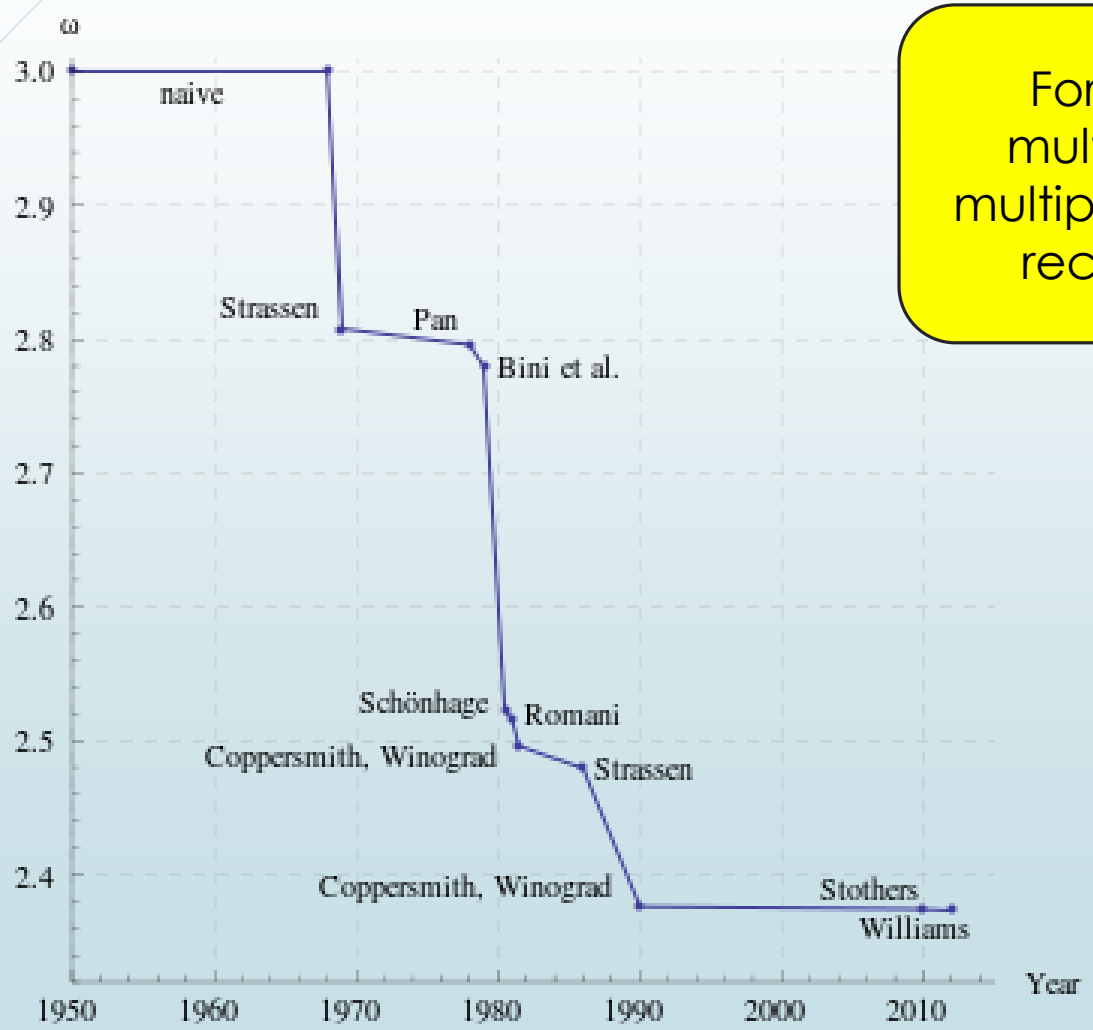
Conjectures that can lead to: $\omega = 2$.

Continue…

- 1986: Strassen, laser method, $\omega < 2.479$ (entirely new approach on the matrix multiplication problem) (V. Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication)
- 1989: Coppersmith & Winograd, combine Strassen's laser method with a novel from analysis based on large sets avoiding arithmetic progression, $\omega < 2.376$ (D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions.)
- 2003: Cohn & Umans: group theoretic framework for designing and analyzing matrix multiplication algorithms
- 2005: Cohn, Umans, Kleinberg, Szegedy, $O(n^{2.41})$ (H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans. Group-theoretic algorithms for matrix multiplication)

  Conjectures that can lead to: $\omega = 2$.
- 2014 Williams: $\omega < 2.373$ (Multiplying matrices in $O(n^{2.373})$ time)

Continue…



For some k they provide a way to multiply $k \times k$ matrices using $m \ll k^3$ multiplications and apply the technique recursively to show that $\omega < \log_k m$

Basic group theory definitions

Group

A group is a non empty set G with a binary operation $\cdot$ defined on G such that the following conditions hold:

1. For all $a, b, c \in G$, we have $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
2. There exists an element $1 \in G$ such that $1 \cdot a = a$ and $a \cdot 1 = a$ for all $a \in G$
3. For all $a \in G$ there exists an element $a^{-1} \in G$ such that $a \cdot a^{-1} = 1$ and $a^{-1} \cdot a = 1$

Order of a group

The order $|G|$ of a group is its cardinality, i.e. the number of elements in its set.

Cyclic Group

A group is said to be cyclic if it is generated by a single element.

(We say that X generates G if $G = \langle X \rangle$ if every element of G can be written as a finite product of elements from X and their inverses. Note that the order of an element a of a group is the order of the subgroup $\langle a \rangle$ it generates)

Abelian Group

The group G is said to be *abelian* if $a \cdot b = b \cdot a$ far all $a, b \in G$.

Group Algebra

The group algebra $F[G]$ of G is defined to be the F-vector space with basis the elements of G endowed with the multiplication extending that on G. Thus:

1. An element of $F[G]$ is a sum $\sum_{g\in G} c_g g,\ c_g \in F$
2. Two elements $\sum_{g\in G} c_g g, \sum_{g\in G} c'_g g$ of $F[G]$ are equal if and only if $c_g = c'_g$ for all g
3. $\left(\sum_{g\in G} c_g g\right)\left(\sum_{g\in G} c'_g g\right) = \sum_{g\in G} c''_g g\ ,\ c''_g = \sum g_1 g_2 = g c_g c'_g$

Homomorphism

A homomorphism for a group G to a group G' is a map $a: G \to G'$ such that $a(ab) = a(a)a(b)$ for all $a, b \in G$. An isomorphism is a bijective homomorphism.

Multiplying polynomials via FFT

- Standard method requires time complexity of $O(n^2)$
- We think of the coefficient vectors of the polynomials as elements of the group algebra $\mathbb{C}[G]$ of a finite group G
- If the group is large (order at least $2n$), convolution of two vectors in the group algebra corresponds to the polynomial product.

Multiplying polynomials via FFT

- Standard method requires time complexity of $O(n^2)$
- We think of the coefficient vectors of the polynomials as elements of the group algebra $\mathbb{C}[G]$ of a finite group G
- If the group is large (order at least $2n$), convolution of two vectors in the group algebra corresponds to the polynomial product.

Discrete convolution

Suppose we have two complex vectors in E^N:

$$Z = (z_0\ z_1\ \cdots\ z_{N-1})^T \qquad Y = (y_0\ y\ \cdots\ y_{N-1})^T$$

The discrete convolution of these two vectors is another vector, which we denote $Z * Y$, defined componentwise by $(Z * Y)_k = \sum_{j=0}^{N-1} z_{k-j} y_j, k = 0,1,2, \ldots$

Multiplying polynomials via FFT

- Standard method requires time complexity of $O(n^2)$
- We think of the coefficient vectors of the polynomials as elements of the group algebra $\mathbb{C}[G]$ of a finite group G
- If the group is large (order at least $2n$), convolution of two vectors in the group algebra corresponds to the polynomial product.
- Convolution in the group algebra can be computed quickly using the FFT.
- Time complexity of FFT and inverse FFT: $O(nlogn)$

Discrete Fourier Transform for polynomials

Discrete Fourier Transform

Embed polynomials as elements of the group algebra $\mathbb{C}[G]$:

Let $G = \langle z \rangle$ be a cyclic group of order $m \geq 2n$. Define

$\bar{A} = \sum_{i=0}^{n-1} a_i z^i \quad \bar{B} = \sum_{i=0}^{n-1} b_i z^i$

Discrete Fourier Transform is an invertible linear transformation

D: $\mathbb{C}[G] \to \mathbb{C}^{|G|}$, such that

$D(\bar{A}) = (\sum_{i=0}^{n-1} a^i x_0^i, \sum_{i=0}^{n-1} a^i x_1^i, \ \ldots, \sum_{i=0}^{n-1} a^i x_{n-1}^i,),\ x_k = e^{\frac{2\pi i}{n}k}$

Then $\bar{A}\bar{B} = D^{-1}(D(\bar{A})D(\bar{B}))$

Embedding matrices A, B into elements $\bar{A}, \bar{B}$ of the group algebra $\mathbb{C}[G]$

Cohn & Umans

Matrix multiplication can be embedded into the group algebra of a finite group G (G must be non-abelian)

Let F be a field and S, T and U be subsets of G.

$A = (a_{s,t})_{s \in S, t \in T}$ and B $= (b_{t,u})_{t \in T, u \in U}$

are $|S| \times |T|$ and $|T| \times |U|$, indexed by elements of S, T and T, U respectively.

Then embed A, B as elements $\bar{A}, \bar{B} \in F[G]$:

$\bar{A} = \sum_{s \in S, t \in T} a_{s,t} s^{-1} t$ and $\bar{B} = \sum_{t \in T, u \in U} b_{t,u} t^{-1} u$

Using the FFT

As with the polynomial method, the Fourier transform provides an efficient way to compute the convolution product.

For a non-abelian group a fundamental theorem of Weddeburn says that the group algebra is isomorphic, via a Fourier transform, to an algebra of block diagonal matrices having block dimensions $d_1 \ldots d_k$, with $\sum d_i^2 = |G|$.

Convolution in $\mathbb{C}[G]$ is thus transformed into block diagonal matrix multiplication.

Embed matrices A, B into elements $\bar{A}, \bar{B}$ of the group algebra $\mathbb{C}[G]$

Multiplication of $\bar{A}$ and $\bar{B}$ in the group algebra is carried out in the Fourier domain after performing the Discrete Fourier Transform (DFT) of $\bar{A}$ and $\bar{B}$

FFT

The product $\bar{A}\bar{B}$ is found by performing the inverse DFT

Inverse FFT

Entries of the matrix AB can be read off from the group algebra product $\bar{A}\bar{B}$

Triple Product Property

The approach works only if the group G admits an embedding of matrix multiplication into its group algebra.

The coefficients of the convolution product correspond to the entries of the product matrix.

Such an embedding is possible whenever the group G has three subgroups, $H_1,\ H_2,\ H_3$ with the property that whenever $h_1 \epsilon\ H_1, h_2 \epsilon\ H_2$ and $h_3 \epsilon\ H_3$ with $h_1 h_2 h_3 = 1$, then $h_1 = h_2 = h_3 = 1$

(The condition can be generalized to subsets of G rather than subgroups)

Beating the sum of cubes

In order for ω to be less than 3, the group must satisfy more conditions.

In particular, it must be the case that:

$|H_1||H_2||H_3| > \sum d_i^3$,

d_i: the block dimensions of the block matrices

Group Theory Definitions

Permutation Groups

Let S be a set and let $Sym(S)$ be the set of bijections $a: S \to S$

$Sym(S)$ is a group, called the group of symmetries of S. For example, the permutation group on n letters S_n is defined to be the group of symmetries of the set$\{1, \dots, n\}$— it has order $n!$.

Groups Acting on Sets

Let X be a set and let G be a group. A left action of G on X is a mapping $(g, x) \mapsto gx: G \times X \to X$ such that

a) $1x = x$, for all $x \in X$

b) $(g_1 g_2)x = g_1(g_2 x)$, all $g_1, g_2 \in G, x \in X$

A set together with a (left) action of G is called a (left) G-set. An action is trivial if $gx = x$ for all $g \in G$

Direct product

When G and H are groups, we can construct a new group $G \times H$, called the (direct) product of G and H. As a set, it is the Cartesian product of G and H, and multiplication is defined by: $(g, h)(g', h') = (gg', hh')$

Normal subgroups

A subgroup N of G is normal, denoted $N \triangleleft G$, if $gNg^{-1} = N$ for all $g \in G$

Semidirect product

A group G *is* a semidirect product of its subgroups N and Q if N is normal and the homomorphism $G \to {}^{G}/_{N}$ induces an isomorphism $Q \to {}^{G}/_{N}$.

We write $G = N \rtimes Q$.

Wreath Product

The wreath product of two groups A and B is constructed in the following way.

Let A^B be the set of all functions defined on B with values in A.

With respect to the componentwise multiplication, this set is a group which is the complete direct product of $|B|$ copies of A.

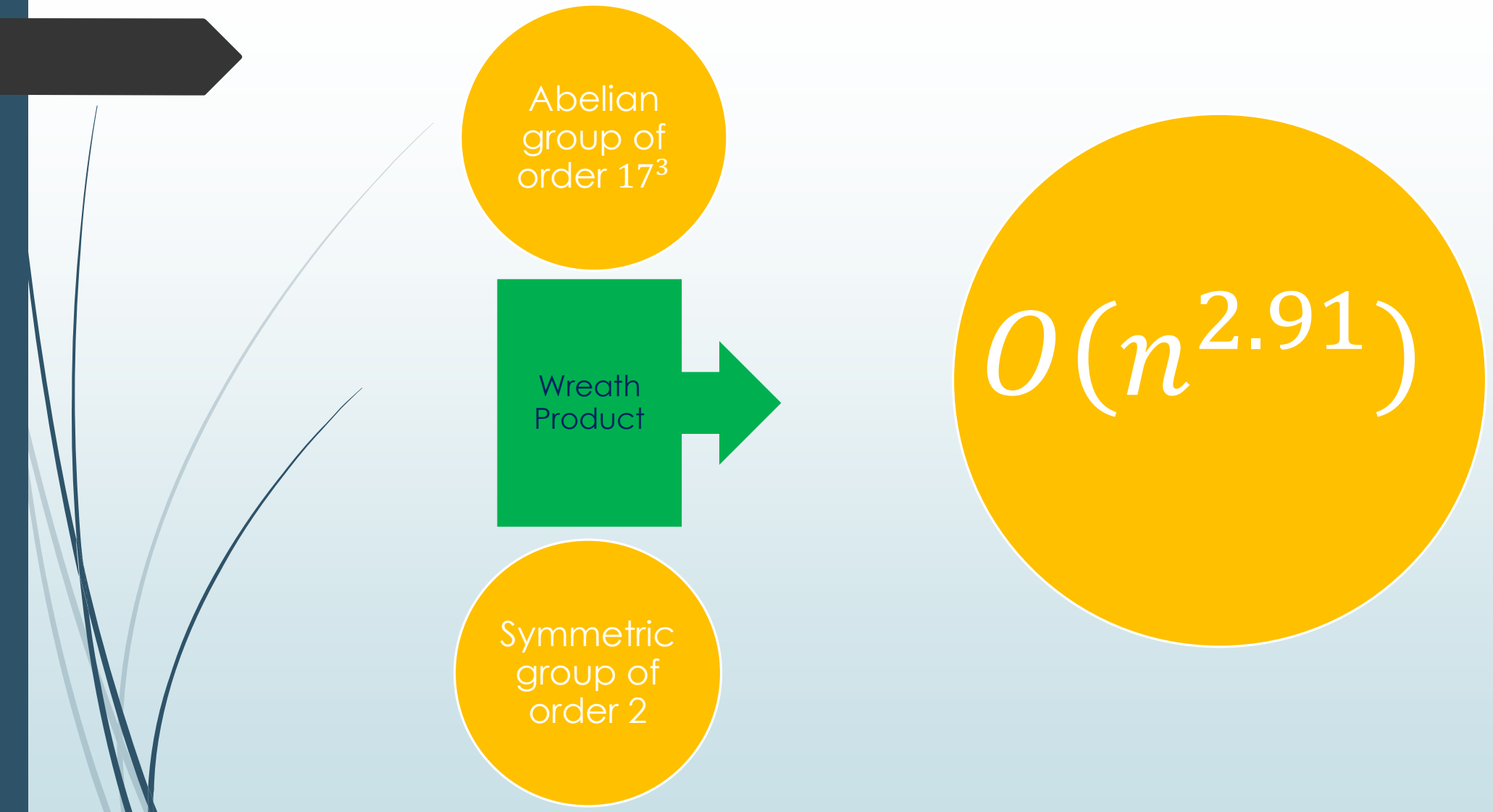
The semidirect product W of B and A^B is called the Cartesian wreath product of A and B, and is denoted by $A\ Wr\ B$.

If instead of A^B one takes the smaller group $A^{(B)}$ consisting of all functions with finite support, that is, functions taking only non-identity values on a finite set of points, then one obtains a subgroup of W called the wreath product of A and B and is denoted by $A\ wr\ B$.

Beating the sum of cubes, finally…

The elusive group G that managed to "beat the sum of cubes" turned out to be a wreath product of:

- Abelian group of order 17^3
- Symmetric group of order 2


Abelian group of order 17^3
Wreath Product
Symmetric group of order 2
$O(n^{2.91})$

Beating the sum of cubes, finally

The elusive group G that managed to "beat the sum of cubes" turned out to be a wreath product of:

- Abelian group of order 17^3
- Symmetric group of order 2

Elementary fact of group representation theory

The index of the largest abelian subgroup of a group is an upper bound on the size of the maximum block of the block diagonal matrix representation given by Wedderburn's theorem.

For a non-abelian group a fundamental theorem of Weddeburn says that the group algebra is isomorphic, via a Fourier transform, to an algebra of block diagonal matrices having block dimensions $d_1 \ldots d_k$, with $\sum d_i^2 = |G|$.

Improving the bounds for ω

- Szegedy realized that some of the combinatory structures of the 1987 Coppersmith - Winograd paper could be used to select the three subsets in the wreath product groups in amore sophisticated way.
- The researchers managed to achieve exponential bound: $\omega < 2.48$
- The researchers distilled their insights into two conjectures, one that has an algebraic flavor and one that has a combinatorial.

A 6×6 strong USP, along with 2 of its 18 pieces