François Le Gall

Graduate School of Informatics
Kyoto University

ADFOCS17 - Lecture 3
24 August 2017
Overview of the Lectures

- Fundamental techniques for fast matrix multiplication (1969~1987)
  - Basics of bilinear complexity theory: exponent of matrix multiplication, Strassen’s algorithm, bilinear algorithms
  - First technique: tensor rank and recursion
  - Second technique: border rank
  - Third technique: the asymptotic sum inequality
  - Fourth technique: the laser method

- Recent progress on matrix multiplication (1987~)
  - Laser method on powers of tensors
  - Other approaches
  - Lower bounds
  - Rectangular matrix multiplication

- Applications of matrix multiplications, open problems
Overview of the Lectures

- Fundamental techniques for fast matrix multiplication (1969~1987)
  - Basics of bilinear complexity theory: exponent of matrix multiplication, Strassen’s algorithm, bilinear algorithms
  - First technique: tensor rank and recursion
  - Second technique: border rank
  - Third technique: the asymptotic sum inequality
  - Fourth technique: the laser method

- Recent progress on matrix multiplication (1987~)
  - Laser method on powers of tensors
  - Other approaches
  - Lower bounds
  - Rectangular matrix multiplication

- Applications of matrix multiplications, open problems
Lower Bounds for Matrix Multiplication

\( R(\langle 2,2,2 \rangle) \geq 7 \) [Hopcroft and Kerr 1971] [Winograd 1971]
\( R(\langle 2,2,2 \rangle) \geq 7 \) [Landsberg 2005]

\[
R(\langle n, n, n \rangle) \geq \frac{3}{2} n^2 \quad [\text{Strassen 1983}]
\]

\[
R(\langle n, n, n \rangle) \geq \frac{3}{2} n^2 + \frac{1}{2} n - 1 \quad [\text{Lickteig 1984}]
\]

\[
R(\langle n, n, n \rangle) \geq \frac{3}{2} n^2 - 2 \quad [\text{Bürgisser, Ikenmeyer 2011}]
\]

\[
R(\langle n, n, n \rangle) \geq 2n^2 - n \quad [\text{Landsberg, Ottaviani 2011}]
\]

\[
R(\langle n, n, n \rangle) \geq \frac{5}{2} n^2 - 3n \quad [\text{Bläser 1999}]
\]

\[
R(\langle n, n, n \rangle) \geq 3n^2 - 4n^{2/3} - n \quad [\text{Landsberg 2012}]
\]

\[
R(\langle n, n, n \rangle) \geq 3n^2 - 2\sqrt{2}n^{2/3} - 3n \quad [\text{Massaranti, Raviolo 2012}]
\]
Theorem ([Raz 2002])

Any arithmetic circuit that computes the product of two $n \times n$ real matrices has size $\Omega(n^2 \log n)$, as long as the circuit does not use products with field elements of absolute value larger than 1.

Space-Time tradeoff (see, e.g., [Abrahamson 1991])

For any algebraic algorithm computing the product of two $n \times n$ matrices using $S$ space and $T$ time we have $ST = \Omega(n^3)$.

Any subcubic-time algorithm for matrix multiplication has superlogarithmic space complexity.

Trivial algorithm: $T = O(n^3)$  $S = O(\log n)$

Strassen algorithm: $T = O(n^{2.81})$  $S = O(n^2)$

same quadratic space complexity for all the other fast algorithms we studied
Overview of the Lectures

- Fundamental techniques for fast matrix multiplication (1969~1987)
  - Basics of bilinear complexity theory: exponent of matrix multiplication, Strassen’s algorithm, bilinear algorithms
  - First technique: tensor rank and recursion
  - Second technique: border rank
  - Third technique: the asymptotic sum inequality
  - Fourth technique: the laser method

- Recent progress on matrix multiplication (1987~)
  - Laser method on powers of tensors
  - Other approaches
  - Lower bounds
  - Rectangular matrix multiplication

- Applications of matrix multiplications, open problems
Rectangular Matrix Multiplication

Compute the product of an $n \times m$ matrix $A$ and an $m \times n$ matrix $B$

\[
\begin{align*}
&\begin{bmatrix}
    \vdots & \vdots & \vdots \\
    a_{ij} & & \\
    \vdots & \ddots & \vdots \\
    \end{bmatrix} \times \\
&\begin{bmatrix}
    \vdots & \vdots & \vdots \\
    b_{ij} & & \\
    \vdots & \ddots & \vdots \\
    \end{bmatrix} = \\
&\begin{bmatrix}
    \vdots & \vdots & \vdots \\
    c_{ij} & & \\
    \vdots & \ddots & \vdots \\
    \end{bmatrix}
\end{align*}
\]

$m$ multiplications and $(m-1)$ additions

\[
c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}
\]

for all $1 \leq i \leq n$ and $1 \leq j \leq n$

Trivial algorithm: $n^2(2m-1)=O(mn^2)$ arithmetic operations
Rectangular Matrix Multiplication

Compute the product of an $n \times m$ matrix $A$ and an $m \times n$ matrix $B$

Compute the product of an $m \times n$ matrix $A$ and an $n \times n$ matrix $B$

Compute the product of an $n \times n$ matrix $A$ and an $n \times m$ matrix $B$

same algebraic complexity (as seen yesterday)

✓ The problem (with $m \neq n$) appears as the bottleneck in many applications:

- linear algebra problems
- all-pairs shortest path problems
- dynamic computation of the transitive closure of a graph
- detecting directed cycles in a graph
- computational geometry (colored intersection searching)
- computational complexity (circuit lower bounds)
Compute the product of an \( n \times n^k \) matrix \( A \) and an \( n^k \times n \) matrix \( B \) for any fixed \( k \geq 0 \)

Exponent of rectangular matrix multiplication

\[
\omega(1,1,k) = \inf \{ \tau \mid \text{this product can be computed using } O(n^\tau) \text{ arithmetic operations} \}
\]

Exponent of rectangular matrix multiplication

\[
\omega(1,1,k) = \inf \{ \tau \mid R(\langle n, n, nk \rangle) = O(n^\tau) \}
\]

trivial algorithm: \( O(n^{2+k}) \) arithmetic operations

\( \omega(1,1,k) \leq 2 + k \)

square matrices: \( \omega(1,1,1) = \omega \leq 2.38 \)

trivial lower bounds: \( \omega(1,1,k) \geq 2 \)

\( \omega(1,1,k) \geq 1 + k \)
Exponent of Rectangular Matrix Multiplication

Property [Lotti 83]

\( \omega(1,1,k) \) is a convex function

upper bounds on \( \omega(1,1,k) \)

trivial algorithm: \( O(n^{2+k}) \) arithmetic operations

\( \omega(1,1,k) \leq 2 + k \)

square matrices: \( \omega(1,1,1) = \omega \leq 2.38 \)

trivial lower bounds: \( \omega(1,1,k) \geq 2 \)

\( \omega(1,1,k) \geq 1 + k \)
Exponent of Rectangular Matrix Multiplication

Property [Lotti 83]

$\omega(1,1,k)$ is a convex function

Upper bounds on $\omega(1,1,k)$

$\omega(1,1,0.172) = 2$

$\omega(1,1,0.294) = 2$

The product of an $n \times n^{0.172}$ matrix by an $n^{0.172} \times n$ matrix can be computed using $O(n^{2+\varepsilon})$ arithmetic operations for any $\varepsilon > 0$

$[\text{Coppersmith 1982}]: \omega(1,1,0.172) = 2$

$[\text{Coppersmith 1997}]: \omega(1,1,0.294) = 2$
[Coppersmith 1982]: \( \omega(1,1,0.172) = 2 \)

The product of an \( n \times n^{0.172} \) matrix by an \( n^{0.172} \times n \) matrix can be computed using \( O(n^{2+\varepsilon}) \) arithmetic operations for any \( \varepsilon > 0 \).

**Proof outline**

Remember Tuesday’s exercise

Consider the computation of the product of two matrices \( A \) and \( B \) of the following form:

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
0 & a_{22} & a_{23}
\end{pmatrix}, \quad B = \begin{pmatrix}
b_{11} & b_{12} \\
b_{21} & 0 \\
b_{31} & 0
\end{pmatrix}
\]

(i) Write the tensor corresponding to this computational task.

(ii) Show that the border rank of this tensor is at most 5.

Taking power \( N \) of this tensor and combining it with power \( N \) of the same tensor with permuted variables we can show:

\[
R(\langle M, 4^{N/5}, M \rangle) \leq 5^{2N} \text{ with } M \approx 5^N
\]

\[
4^{N/5} = M^\alpha \text{ with } \alpha = \frac{1}{5} \log_5 4 = 0.1722 ...
\]
Exponent of Rectangular Matrix Multiplication

[Coppersmith 1997]: \( \omega(1,1,0.294) = 2 \)

**Idea:** Analyze the first power of the CW tensor in an asymmetric way

**Tool:** Rectangular version of the asymptotic sum inequality

As in the square case, it is enough to create a direct sum of matrix products of the desired format
Exponent of Rectangular Matrix Multiplication

Idea: Analyze the first power of the CW tensor in an asymmetric way

\[ T_{CW} = T_{CW}^{011} + T_{CW}^{101} + T_{CW}^{110} + T_{CW}^{002} + T_{CW}^{020} + T_{CW}^{200} \]

\[ T_{CW}^{011} \cong \langle 1, 1, q \rangle \]
\[ T_{CW}^{101} \cong \langle q, 1, 1 \rangle \]
\[ T_{CW}^{110} \cong \langle 1, q, 1 \rangle \]

\[ T_{CW}^{002} \cong \langle 1, 1, 1 \rangle \]
\[ T_{CW}^{020} \cong \langle 1, 1, 1 \rangle \]
\[ T_{CW}^{200} \cong \langle 1, 1, 1 \rangle \]

square case

rectangular case

Analysis of the second construction

Theorem 5

For any \( 0 \leq \alpha \leq 1/3 \) and for \( N \) large enough, the tensor \( T_{CW}^{\otimes N} \) can be converted into a direct sum of

\( 2(\frac{2}{3} - \alpha, 2\alpha, \frac{1}{3} - \alpha) - o(1) \) terms, each isomorphic to

\[ \langle q^{\alpha N}, q^{\alpha N}, q^{\alpha' N} \rangle \]

more variables

recalculate

adjust

adjust

adjust

conjecture
Exponent of Rectangular Matrix Multiplication

Upper bounds on $\omega(1,1,k)$

- $\omega(1,1,0) = 2$
- $\omega(1,1,0.172) = 2$
- $\omega(1,1,0.294) = 2$
- $\omega(1,1,0.5356) < 2.0712$
- $\omega(1,1,0.8) < 2.2356$
- $\omega(1,1,2) < 3.2699$

This curve has been used in most applications of rectangular matrix multiplication.

Property [Lotti 83]: $\omega(1,1,k)$ is a convex function.

Curve obtained by doing the same analysis for any value of $k$.

[Coppersmith 1982]: $\omega(1,1,0.172) = 2$
[Coppersmith 1997]: $\omega(1,1,0.294) = 2$

[Ke, Zeng, Han, Pan 2008]: $\omega(1,1,0.5356) < 2.0712$
- $\omega(1,1,0.8) < 2.2356$
- $\omega(1,1,2) < 3.2699$

(slightly improving a bound from [Huang, Pan 1998]).

Obtained by a similar asymmetric analysis of the first power of the CW tensor.
Exponent of Rectangular Matrix Multiplication

**Upper bounds on \( \omega(1,1,k) \)**

\[ \omega(1,1,0.294) = 2 \]
\[ \alpha > 0.294 \]
proving that \( \alpha = 1 \) is equivalent to proving that \( \omega = 2 \)

- Coppersmith 1997: \( \omega(1,1,0.294) = 2 \)
- LG 2012: \( \omega(1,1,0.302) = 2 \)

**Dual exponent of matrix multiplication**

\[ \alpha = \sup \{ k \mid \omega(1,1,k) = 2 \} \]

- from the analysis of the first power of the CW tensor
- from the analysis of the second power of the CW tensor

This curve has been used in most applications of rectangular matrix multiplication.
from [LG 2012]

<table>
<thead>
<tr>
<th>k</th>
<th>upper bound on ω(1, 1, k)</th>
<th>k</th>
<th>upper bound on ω(1, 1, k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.30298</td>
<td>2</td>
<td>0.60</td>
<td>2.096571</td>
</tr>
<tr>
<td>0.31</td>
<td>2.000063</td>
<td>0.65</td>
<td>2.125676</td>
</tr>
<tr>
<td>0.32</td>
<td>2.000371</td>
<td>0.70</td>
<td>2.156959</td>
</tr>
<tr>
<td>0.33</td>
<td>2.000939</td>
<td>0.75</td>
<td>2.190087</td>
</tr>
<tr>
<td>0.34</td>
<td>2.001771</td>
<td>0.80</td>
<td>2.224790</td>
</tr>
<tr>
<td>0.35</td>
<td>2.002870</td>
<td>0.85</td>
<td>2.260830</td>
</tr>
<tr>
<td>0.40</td>
<td>2.012175</td>
<td>0.90</td>
<td>2.298048</td>
</tr>
<tr>
<td>0.45</td>
<td>2.027102</td>
<td>0.95</td>
<td>2.336306</td>
</tr>
<tr>
<td>0.50</td>
<td>2.046681</td>
<td>1.00</td>
<td>2.375477</td>
</tr>
<tr>
<td>0.60</td>
<td>2.086936</td>
<td>1.05</td>
<td>2.456151</td>
</tr>
<tr>
<td>0.55</td>
<td>2.070063</td>
<td>1.20</td>
<td>2.539392</td>
</tr>
</tbody>
</table>

Table 1: Our upper bounds on the exponent of the multiplication matrix.

- exactly the same bound as the one obtained by Coppersmith and Winograd for square matrix multiplication
- better than all previous bounds for k≠1
- curve of the same shape, but slightly below the previous curve

[LG 2012]: ω(1,1,0.302) = 2 → α > 0.302

from the analysis of the second power of the CW tensor
### Exponent of Rectangular Matrix Multiplication

The dual exponent of matrix multiplication is defined as:

\[ \alpha = \sup \{ k \mid \omega(1,1,k) = 2 \} \]

<table>
<thead>
<tr>
<th>Power of the CW Tensor</th>
<th>Upper Bound of (\omega)</th>
<th>Lower Bound of (\alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>(\omega &lt; 2.3872)</td>
<td>(\alpha &gt; 0.294)</td>
</tr>
<tr>
<td>Second</td>
<td>(\omega &lt; 2.3755)</td>
<td>(\alpha &gt; 0.302)</td>
</tr>
<tr>
<td>Fourth</td>
<td>(\omega &lt; 2.3730)</td>
<td>(\alpha &gt; 0.313)</td>
</tr>
<tr>
<td>Eighth</td>
<td>(\omega &lt; 2.3729)</td>
<td>???</td>
</tr>
</tbody>
</table>

The gap increases with higher powers. What will happen with higher powers?
Overview of the Lectures

- Fundamental techniques for fast matrix multiplication (1969~1987)
  - Basics of bilinear complexity theory: exponent of matrix multiplication, Strassen’s algorithm, bilinear algorithms
  - First technique: tensor rank and recursion
  - Second technique: border rank
  - Third technique: the asymptotic sum inequality
  - Fourth technique: the laser method

- Recent progress on matrix multiplication (1987~)
  - Laser method on powers of tensors
  - Other approaches
  - Lower bounds
  - Rectangular matrix multiplication

- Applications of matrix multiplications, open problems