

Lecture Notes in Computer Science

Edited by G. Goos and J. Hartmanis

179

Victor Pan

How to
Multiply Matrices Faster



Springer-Verlag
Berlin Heidelberg New York Tokyo 1984

Editorial Board

D. Barstow W. Brauer P. Brinch Hansen D. Gries D. Luckham
C. Moler A. Pnueli G. Seegmüller J. Stoer N. Wirth

Author

Victor Pan
State University of New York at Albany
Department of Computer Science
1400 Washington Avenue, Albany, NY 12222, USA

CR Subject Classification (1982): F.2.1, G.1.3

ISBN 3-540-13866-8 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-13866-8 Springer-Verlag New York Heidelberg Berlin Tokyo

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machine or similar means, and storage in data banks. Under § 54 of the German Copyright Law where copies are made for other than private use, a fee is payable to "Verwertungsgesellschaft Wort", Munich.

© by Springer-Verlag Berlin Heidelberg 1984
Printed in Germany

Printing and binding: Beltz Offsetdruck, Hemsbach / Bergstr.
2146 / 3140-543210

CONTENTS

SOME NOTATION AND ABBREVIATIONS	VII
INTRODUCTION	1
PART 1. THE EXPONENT OF MATRIX MULTIPLICATION	7
Summary	7
1. The Power of Recursive Algorithms for Matrix Multiplication	7
2. Bilinear Algorithms for MM	9
3. The Search for a Basis Algorithm and the History of the Asymptotic Acceleration of MM	18
4. The Basic Algorithm and the Exponent 2.67	20
5. The Dependence of the Exponent of MM on the Class of Constants Used	23
6. λ -algorithms and Their Application to MM. Accumulation of the Accelerating Power of λ -algorithms via Recursion	28
7. Strassen's Conjecture. Its Extended and Exponential Versions	33
8. Recursive Algorithms for MM and for Disjoint MM (Definitions, Notation, and Two Basic Facts)	36
9. Some Applications of the Recursive Construction of Bilinear Algorithms	43
10. Trilinear Versions of Bilinear Algorithms and of Bilinear λ -algorithms. Duality. Recursive Trilinear Algorithms	50
11. Trilinear Aggregating and Some Efficient Basis Designs	56
12. A Further Example of Trilinear Aggregating and Its Refinement via a Linear Transformation of Variables	58
13. Aggregating the Triplets of Principal Terms	61
14. Recursive Application of Trilinear Aggregating	69
15. Can the Exponent Be Further Reduced?	77
16. The Exponents Below 2.5	80

17. How Much Can We Reduce the Exponent?	89
PART 2. CORRELATION BETWEEN MATRIX MULTIPLICATION AND OTHER COMPUTATIONAL PROBLEMS.	
BIT-TIME, BIT-SPACE, STABILITY, and CONDITION	95
Summary	95
18. Reduction of Some Combinatorial Computational Problems to MM	96
19. Asymptotic Arithmetical Complexity of Some Computations in Linear Algebra	103
20. Two Block-Matrix Algorithms for the QR-factorization and QR-type Factorization of a Matrix	107
21. Applications of the QR- and QR-type Factorization to the Problems MI, SLE, and Det	113
22. Storage Space for Asymptotically Fast Matrix Operations	115
23. The Bit-Complexity of Computations in Linear Algebra. The Case of Matrix Multiplication	117
24. Matrix Norms and Their Application to Estimating the Bit-Complexity of Matrix Multiplication	124
25. Stability and Condition of Algebraic Problems and of Algorithms for Such Problems	128
26. Estimating the Errors of the QR-factorization of a Matrix	133
27. The Bit-Complexity and the Condition of the Problem of Solving a System of Linear Equations	140
28. The Bit-Complexity and the Condition of the Problem of Matrix Inversion	143
29. The Bit-Complexity and the Condition of the Problem of the Evaluation of the Determinant of a Matrix	145
30. Summary of the Bounds on the Bit-Time of Computations in Linear Algebra; Acceleration of Solving a System of Linear Equations Where High Relative Precision of the Output Is Required	148

PART 3. THE SPEED-UP OF THE MULTIPLICATION OF MATRICES OF A FIXED SIZE	153
Summary	153
31. The Currently Best Upper Bounds on the Rank of the Problem of MM of Moderate Sizes	154
32. Commutative Quadratic Algorithms for MM	162
33. λ -algorithms for the Multiplication of Matrices of Small and Moderate Sizes	166
34. The Classes of Straight Line Arithmetical Algorithms and λ -algorithms and Their Reduction to Quadratic Ones	171
35. The Basic Active Substitution Argument and Lower Bounds on the Ranks of Arithmetical Algorithms for Matrix Multiplication	175
36. Lower Bounds on the λ -rank and on the Commutative λ -rank of Matrix Multiplication	183
37. Basic Active Substitution Argument and Lower Bounds on the Number of Additions and Subtractions	187
38. Nonlinear Lower Bounds on the Complexity of Arithmetical Problems Under Additional Restrictions on the Computational Schemes	190
39. A Trade-off between the Additive Complexity and the Asynchronicity of Linear and Bilinear Algorithms	193
40. An Attempt of Practical Acceleration of Matrix Multiplication and of Some Other Arithmetical Computations	196
APPENDIX	199
INDEX OF SOME CONCEPTS	201
REFERENCES	205

Some Notation and Abbreviations

Notation	Meaning, Comments	Defined or First Used In Section(s) (see also Index)
A	algorithm	19,23
ar(A);ar(P)	number of arithmetical operations involved in A; required in order to solve a problem P	19,23
as(A)	number of additions/subtractions involved in A	32,33
AAPR	accumulation of the accelerating power via recursion	6
$b_y(X,Y)$	bilinear form in X,Y	2
BA(n)	bilinear algorithm for $n \times n$ MM	2,22,23
BA(n, λ)	bilinear λ -algorithm for $n \times n$ MM	23
BBM	Boolean MM	18
bs(A)	bit-space used by A	23
bt(A)	bit-time used by A	23
bt(s),bt(*,s),bt(\pm ,s)		18
bs(P),bt(P)	bit-time and bit-space of a computational problem P	23
C	the field of complex numbers	2
φ	commutative rank	32
$c\varphi$	commutative λ -rank	33
$C(g,h)$	$g!/(h!(g-h)!)$	8,9
cond	condition	25
D	domain of definition of problem or algorithm	Part 2. (Summary); 23
d	degree of λ -algorithm	6
d	shortest distance	18 only

VIII

$\det(W)$	determinant of a matrix W	19
$\text{Det}(n)$	the problem of the evaluation of the determinant of an $n \times n$ matrix	Part 2 (Summary); 19
$\text{DFT}(n)$	discrete Fourier transform,	38,39
E	extension of a ring (field)	5
$E, E(n), e(n), E(A, D, h),$ $E(Z(V), D, h)$	error bounds	Part 2 (Summary); 23-30
F	ring, field	2
$F[\lambda]$	ring of polynomials over F	6
$f(i, j, q), f'(j, k, q),$ $f''(k, i, q),$ $f(\alpha, q), f'(\beta, q),$ $f''(\gamma, q)$	constant coefficients (from F) of bilinear algorithms	2
$f(i, j, q, \lambda),$ $f'(j, k, q, \lambda),$ $f''(k, i, q, \lambda),$ $f(\alpha, q, \lambda),$ $f'(\beta, q, \lambda),$ $f''(\gamma, q, \lambda)$	coefficients (from $F[\lambda]$) of bilinear λ -algorithms	6
$f, f', f'', \tilde{f}, \tilde{f}', \tilde{f}''$		23
FFT	fast Fourier transform	Intr., 2, 38
$h(s)$	2^{1-s}	23
u^H, W^H	complex conjugate of number u , conjugate transpose of matrix W	19
I (also I_n)	identity matrix (of size $n \times n$)	19
$l_h(l_2, l_\infty)$	l_h -norm of a matrix or of a vector	24
$\log u$	logarithm to the base 2 of u	1

L_q, L'_q		2
L''_q		10
M	rank of algorithm, λ -rank of λ -algorithm	2,4,6
MA,MS	matrix addition, subtraction	20
MI	matrix inversion	Part 2 (Summary); 19
MM	matrix multiplication	Intr., 1
(m,n,p); also $m \times n \times p$ MM	the problem of $m \times n$ by $n \times p$ MM	2
$O(g(s)), o(g(s))$	see Notation 18.1	Intr., 1, 18
O, O_n	null matrix	19, 20
PM	polynomial multiplication	2
Q	field of rational numbers	2
Q	unitary matrix (a QR-factor)	20
$Q(s)$	computed approximation to Q	26-30
QR, \tilde{QR}, QR^*		20
R	upper triangular matrix (a QR-factor)	20
$R(s)$	computed approximation to R	26-30
R	field of real numbers	2
\underline{R}	set of vectors in the proof of Theorem 7.2	9 only
SLE	the problem of solving a system of linear equations	Part 2 (Summary); 19
$sm(A)$	number of scalar multiplications in A	32, 33
T	trilinear form	10
TA	trilinear aggregating	Intr., 3, 11

$\text{Tr}(W)$	trace of a matrix W	10
TMI	triangular matrix inversion	21
t	tensor	2,10
U, V, W, X, Y, Z	matrices	1,2,4,6,10
\mathbb{Z}	ring of integers	2,5
$\mathbb{Z}(\mathbb{N})$	ring of integers modulo \mathbb{N}	2,5
$Z(V)$	output matrix	24-30
$\delta(i,j)$	$\delta(i,j)=0, i \neq j; \delta(i,i)=1$	2
Δ, Δ'	error value, error matrix	23-30
λ	see λ -algorithms	4,6
ρ, ρ_F	rank, rank over a ring F	2
$\rho(m,n,p)$	rank of $m \times n \times p$ MM	2
ρ_λ	λ -rank	36
ω	exponent of MM	2
ω_F	exponent of MM over a ring F of constants	2
Σ, Π	symbols of sums, products	
Σ	diagonal matrix	20 only
$\lfloor u \rfloor, \lceil u \rceil$	see Notation 18.1	18
\oplus	direct sum of disjoint problems	8
\odot	direct sum of identical disjoint problems	2,5,8
\otimes	(tensor) product of bilinear problems	2,5,8
\boxtimes	direct (Kronecker) product of vectors, matrices, tensors, and of linear, bilinear, or polylinear forms	10,14,16

$\#$	generalized MM	18 only
$\ v\ , \ W\ $	norms of vector v , matrix W	24
$t \leftarrow t'$	mapping (algorithm)	5,8
$ S $	cardinality of a set S	
$ u $	absolute value (modulus) of a number u	
C, \subseteq	inclusion of one set into another	5
\in	inclusion of an element into a set	9
\cup	union of sets	5
\blacksquare	end of clause, of proof, of algorithm	