The Numerical Analysis of Milvio Capovani

Paolo Zellini
Dipartimento di Matematica, Università di Roma "Tor Vergata"
zellini@mat.uniroma2.it

Cortona, September 2008

Scientific Computation vs. Computer Science

Smale, 1990: Schism or conflict between Scientific Computation and Computer Science.

	Scientific Computation	Computer Science
Mathematics	continuous	discrete
Problems	classical	newer
Goals	practical, immediate	long range
Foundations	none	developed
Complexity	undeveloped	developed
Machine, model	none	Turing

Blum, Shub, Smale, 1989: Theory of computation and complexity over the real numbers, NP-completeness, Recursive Functions, Universal Machines.

Possible links

Milvio Capovani: computational complexity, infomational content, models of computation (bilinear programs), algebraic theory of matrices

Analytical approach $\longrightarrow$ Combinatorial, algebraic approach

Arithmetizing analysis:
1. Foundations: all analysis could be based logically on a combination of ordinary arithmetic and passage to the limit (Weierstrass, Dedekind, Poincaré, Cantor)
2. Fredholm's theory of integral equations, whose kernels $K(x, y)$ can be treated as limits of matrices
3. Variational methods: Rayleigh, 1873; Ritz, 1906. Dirichlet problem: proof of a constructive existence theorem
4. Arithmetizing Analysis in principle $\longrightarrow$ Arithmetizing practically, effective procedures (complexity, error)

Arithmetizing: Goldstine, von Neumann, Strang

H. Goldstine, J. von Neumann, 1946: "Our problems are usually given as continuous-variable analytical problems, frequently wholly or partly of an implicit character. For the purposes of digital computing they have to be replaced, or rather approximated, by purely arithmetical "finitistic" explicit (usually step-by-step or iterative) procedures." (Compare to Hilbert's foundational program)

G. Strang, 1994: "For engineers and social and physical scientists, *linear algebra* now fills a place that is often more important than calculus. My generation of students, and certainly my teachers, did not see this change coming. It is partly the move from analog to digital; functions are replaced by vectors. Linear algebra combines the insight of $n-$dimensional space with the applications of matrices"

Arithmetizing $\longrightarrow$ matrix computation

Informational content

Numerical work is often concerned with operations on matrices belonging to special classes. Within a class the generic matrix is often specified by a number k of parameters less than the numbers of elements. $\rightarrow$ Informational content of a matrix

Measure of informational content: amount of memory required to store the matrix as compactly as possible in a computer (Forsythe, 1967)

1. Representation of a matrix in a computer (Forsythe)
2. Computational complexity (Capovani, Capriz, Bini, Bevilacqua, Zellini)

Compare to Chaitin, 1974: complexity of a string of bits as the minimum length of a program that generates the string

Capriz, Capovani, 1976: $\mathcal{C}_n^k$ = class of matrices $n \times n$ of informational content k = manifold of dimension k, $k \leq n^2$, in the space of dimension n^2 of all real $n \times n$ matrices.

Informational content and computational complexity

The case when $\mathcal{C}_n^k =$ is an algebra spanned by k linearly independent matrices $J_i, i = 1, 2, \ldots, k$

Let $A = \sum_{i=1}^k a_i J_i, \quad B = \sum_{j=1}^k b_j J_j, \quad J_i J_j = \sum_{h=1}^k t_{hij} J_h$
$t_{hij} =$ multiplication table

$$AB = \sum_{i,j=1}^k a_i b_j J_i J_j = \sum_{h=1}^k [\sum_{i,j=1}^k t_{hij} a_i b_j] J_h = \sum_{h=1}^k f_h(a,b) J_h$$

where $f_h(a,b) = \sum_{i,j=1}^k t_{hij} a_i b_j$ = bilinear form in the indeterminates a, b.

the last formula exhibits possible reductions in computational complexity rk(t_{hij}) = rank of the tensor t_{hij} in a field $\mathcal{F}$ = minimum integer q such that

$$t_{hij} = \sum_{r=1}^{q} u_{hr} v_{ir} w_{jr}$$

for $3q$ vectors $u_h, v_h, w_h, h = 1, 2, \ldots, q$ with elements in $\mathcal{F}$.

If the rank of t_{hij} is q, then the coefficients c_h of AB are

$$c_h = \sum_{i,j=1}^{k} a_i b_j \sum_{r=1}^{q} u_{hr} v_{ir} w_{jr} = \sum_{r=1}^{q} u_{hr} (\sum_{i=1}^{k} a_i v_{ir}) \cdot (\sum_{j=1}^{k} b_j v_{jr})$$

i.e. q non-scalar multiplications are sufficient (necessary when commutativity is not assumed) to compute c_h. Then the rank of the tensor t_{hij} of the multiplication table of $\mathcal{C}_n^k$ defines the multiplicative complexity of the product of two elements of $\mathcal{C}_n^k$.

Let $\mathcal{F}$ be a field with infinite elements and $T = t_{hij}$ a tensor on $\mathcal{F}$. rkb(T)= border rank of T = minimum integer t such that, for every $\varepsilon > 0$ we have a tensor $E = e_{hij}$, with $|e_{hij}| < \varepsilon$, such that $\mathrm{rk}(T + E) = t$.

We have $\mathrm{rkb}(T) \leq \mathrm{rk}(T)$ and sometimes $\mathrm{rkb}(T) < \mathrm{rk}(T)$. For

$$T = (\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix})$$

we have $\mathrm{rkb}(T) = 2$ and $\mathrm{rk}(T) = 3$.

Bini, Capovani, Lotti, Romani, 1979-1980, 1981: Complexity of approximate algorithms, main applications to:
1. band Toeplitz matrices
2. matrix multiplication: algorithm of complexity $O(n^w)$, $w \leq 2.7798\ldots$ for solving a system of n linear equations, improving Strassen's limit $w \leq log_2 7 = 2.807\ldots$

Bevilacqua, Capovani, 1972: algebra $\mathcal{C}_n^k = \tau$ of informational content $k = n$

$$\tau_5 = \begin{bmatrix} t_1 & t_2 & t_3 & t_4 & t_5 \\ t_2 & t_1 + t_3 & t_2 + t_4 & t_3 + t_5 & t_4 \\ t_3 & t_2 + t_4 & t_1 + t_3 + t_5 & t_2 + t_4 & t_3 \\ t_4 & t_3 + t_5 & t_2 + t_4 & t_1 + t_3 & t_2 \\ t_5 & t_4 & t_3 & t_2 & t_1 \end{bmatrix}$$

cross-sum condition: $t_{i-1,j} + t_{i+1,j} = t_{i,j-1} + t_{i,j+1}$

τ generated over R by the matrix

$$H = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The class τ is now used (like the class of circulant matrices) in many problems in numerical linear algebra: matrix displacement decompositions, optimal preconditioning, complexity of Toeplitz matrices.
τ = example of class $\mathcal{C}_n^k$ with $k = n$ obtained by choosing an orthogonal matrix Q of order n and taking all matrices $G = QDQ^T$ where D = arbitrary real diagonal matrix. Compare to circulant matrices and to Hartley algebra (Bini, Favati, 1993).

Bini, Capovani, 1983: "We try to separate what is related to the structure of the class from what is related to the specific values (informational content) of the matrix"

$Q \longrightarrow$ structure
$D \longrightarrow$ informational content

$A = QD_AQ^T$, $\quad A$ generated by $H = QDQ^T$

For all classes of matrices which are algebras generated by one matrix H it is possible to accomplish completely such a separation between the structure and the informational content. In fact, if $H = QDQ^T$ and the class is generated by H, then all matrices A of the class have the form $A = QD_AQ^T$ (and commute with H).

Structure of $n-$dimensional commutative spaces $\sum_{k=1}^{n} a_k J_k$ of minimal informational content and minimal complexity, where J_k are $(0, 1)$ matrices with prescribed sum.
Zellini, 1979 and 1985; Grone, Hoffman, Wall, 1982; Bevilacqua, Zellini, 1989 and 1996, Bevilacqua, Di Fiore, Zellini, 1996;

This theoretical study has inspired numerical research: preconditioning tecniques, representations of a matrix A as sums of products of matrices belonging to spaces $\sum_{k=1}^{n} a_k J_k$, using displacement rank.

Representation of a band symmetric Toeplitz matrix (BST)
$a_i - a_j \rightarrow i - j$

$$B = \begin{bmatrix} 1-3 & 2-4 & 3 & 4 & 0 & 0 & 0 \\ 2-4 & 1 & 2 & 3 & 4 & 0 & 0 \\ 3 & 2 & 1 & 2 & 3 & 4 & 0 \\ 4 & 3 & 2 & 1 & 2 & 3 & 4 \\ 0 & 4 & 3 & 2 & 1 & 2 & 3 \\ 0 & 0 & 4 & 3 & 2 & 1 & 2-4 \\ 0 & 0 & 0 & 4 & 3 & 2-4 & 1-3 \end{bmatrix} \in \tau_{n+2},\ n = 5$$

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 0 \\ 2 & 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 2 & 3 \\ 4 & 3 & 2 & 1 & 2 \\ 0 & 4 & 3 & 2 & 1 \end{bmatrix} = \text{7-diagonal BST matrix}$$

informational content and bordering 2

If μ_i are the eigenvalues of B, $\mu_1 \geq \mu_2 \ldots \geq \mu_{n+2}$, then a representation of B in the basis $I, H, \ldots, H^{n+1}$ gives the following representation of a $n \times n$ (n even) 7-diagonal BST Toeplitz matrix A, with elements a_1, a_2, a_3, a_4:

$$A = VP^T \begin{bmatrix} D_1 + \mu_1 v_1 v_1^T & 0 \\ 0 & D_2 + \mu_{n+2} v_2 v_2^T \end{bmatrix} PV$$

where $D_1 = diag(\mu_3, \mu_5, \ldots, \mu_{n+1})$, $D_2 = diag(\mu_2, \mu_4, \ldots, \mu_n)$, P = permutation matrix, and $V, v_i, i = 1, 2$, do not depend on a_k.

Bini, Capovani, 1983: The eigenvalues λ_i of A satisfy
$\mu_{i+2} \leq \lambda_i \leq \mu_i$

$V, v_i \rightarrow$ structure
μ_i = linear functions of $a_1, a_2, a_3, a_4 \rightarrow$ informational content

Milvio Capovani, fundamental idea: the error in approximation is not always a cause of failure; by approximating a problem by a "better" one - where matrix algebras and fast transforms are involved - we can improve efficiency.
In quasi-Newton methods for unconstrained minimization in R^n an analogous idea is used to reduce complexity. In fact, in the BFGS iterative step for $\min f(x), x \in R^n$(B_k positive definite)

$$
\begin{aligned}
&d_k = -B_k^{-1}\nabla f(x_k),\\
&x_{k+1} = x_k + \lambda_k d_k\\
&B_{k+1} = \Phi(B_k, s_k, y_k)\\
&s_k = x_{k+1} - x_k \text{ and } y_k = \nabla f(x_{k+1}) - \nabla f(x_k)
\end{aligned}
$$

B_k can be approximated, in Frobenius norm, by a matrix with strong structure (τ, circulant or others), reducing the informational content sufficient for convergence and leading to $O(nlogn)$ arithmetic operations per step, instead of $O(n^2)$ of BFGS (Di Fiore, Fanelli, Lepore, Zellini, 2003)

Winograd-Parlett: FFT via circulants

Rader, 1968; McClellan, Rader, 1979: for n prime, the nontrivial part of a Fourier transform $F_n x$ is the computation of Cy, where C is a special circulant of order $n-1$ and y's elements are a subset of x's elements.

Ciclic convolution on n points as product of two polynomials mod $u^{n-1}-1$ (Winograd, 1978) $\rightarrow$ real spectral factorization of C (Parlett, 1982)

$$C = GDG^T$$

D is block diagonal with 2×2 and 1×1 blocks and G's elements are small integers, so G and G^T act via additions, and only the application of D involves genuine multiplications. For $n=5$

$$D = -\frac{1}{4} \oplus \frac{1}{2}(\cos\frac{1}{5}\pi + \cos\frac{2}{5}\pi) \oplus \begin{bmatrix} \sin\frac{2}{5}\pi & -\sin\frac{1}{5}\pi \\ \sin\frac{1}{5}\pi & \sin\frac{2}{5}\pi \end{bmatrix} i$$