(An initial attempt at) a history of "large-scale" linear algebra and optimization

Jack Poulson

Department of Mathematics
Stanford University

Stanford University Mathematical Organization
January 15, 2015

Outline

Disclaimer

This is an informal, self-indulgent, and very preliminary, step towards expanding on the style of quotations in Nicholas Higham's "Accuracy and stability of numerical algorithms" to a more general class of "direct" algorithms.

Outline

Pivoted LU factorizations

$$PAQ^T = LU$$

P,Q permutations, L lower-triangular, U upper-triangular

When I joined N.P.L. in 1946 the mood of pessimism about the stability of elimination methods for solving linear systems was at its height and was a major talking point. Bounds had been produced which purported to show that the error in the solution would be proportional to 4^n *and this suggested that it would be impractical to solve systems even of quite modest order.*

James H. Wilkinson, 1970 Turing Award Speech

I think it was true to say that at that time (1946) it was the more distinguished mathematicians who were most pessimistic, the less gifted being perhaps unable to appreciate the full severity of the difficulties.

James H. Wilkinson, 1970 Turing Award Speech

However, it happened that some time after my arrival, a system of 18 equations arrived in Mathematics Division [...]. A system of 18 is surprisingly formidable [...]. The operation was manned by Fox, Goodwin, Turing, and me, and we decided on Gaussian elimination with complete pivoting. Turing was not particularly enthusiastic, [...] partly because he was convinced that it would be a failure.

James H. Wilkinson, 1970 Turing Award Speech

It is interesting that [...] the first solution had had almost six correct figures. I suppose this must be regarded as a defeat for Turing [...]. About a year later he produced his famous paper "Rounding-off errors in matrix processes" which together with the paper of J. von Neumann and H. Goldstine did a great deal to dispel the gloom.

James H. Wilkinson, 1970 Turing Award Speech

By 1949 the major components of the Pilot ACE were complete and undergoing trials [...] 26th June, 1951 was a landmark in the history of the machine, for on that day it first rivalled alternative computing methods by yielding by 3 p.m. the solution to a set of 17 equations submitted the same morning.

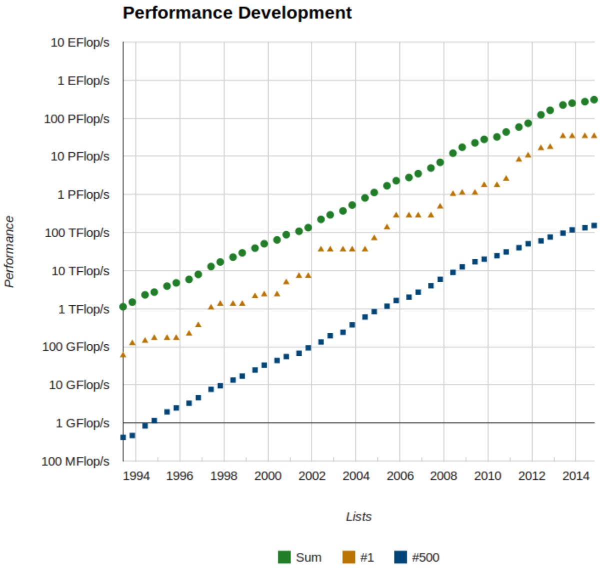
Michael Woodger, The History and Present Use of Digital Computers at the National Physical Laboratory (1958)

“Top 500” List

> *The benchmark used in the LINPACK Benchmark is to solve a dense system of linear equations. For the TOP500, we used that version of the benchmark that allows the user to scale the size of the problem and to optimize the software in order to achieve the best performance for a given machine.*

Dongarra et al., top500.org

“Top 500” List



Dongarra et al., top500.org

"Top 500" List

RANK	SITE	SYSTEM	CORES	RMAX (TFLOP/S)	RPEAK (TFLOP/S)	POWER (KW)
1	National Super Computer Center in Guangzhou China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
5	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945
6	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint** - Cray XC30, Xeon E5-2670 8C 2.600GHz, Aries interconnect , NVIDIA K20x Cray Inc.	115,984	6,271.0	7,788.9	2,325
7	Texas Advanced Computing Center/Univ. of Texas United States	**Stampede** - PowerEdge C8220, Xeon E5-2680 8C 2.700GHz, Infiniband FDR, Intel Xeon Phi SE10P Dell	462,462	5,168.1	8,520.1	4,510
8	Forschungszentrum Juelich (FZJ) Germany	**JUQUEEN** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	458,752	5,008.9	5,872.0	2,301
9	DOE/NNSA/LLNL United States	**Vulcan** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	393,216	4,293.3	5,033.2	1,972
10	Government United States	Cray CS-Storm, Intel Xeon E5-2660v2 10C 2.2GHz, Infiniband FDR, Nvidia K40 Cray Inc.	72,800	3,577.0	6,131.8	1,499

Dongarra et al., top500.org

Outline

Schur decompositions

$$A = QTQ^H$$

Q unitary, T upper-triangular

Using a modern version of the QR algorithm one can expect to produce an accurate eigensystem of a dense matrix of order 100 in a time which is of the order of a minute. [In 1957] we did not appear to be within hailing distance of such an achievement.

James H. Wilkinson, 1970 Turing Award Speech

John Francis's implicitly shifted QR algorithm turned the problem of matrix eigenvalue computation from difficult to routine almost overnight about fifty years ago. [...] it deserves to be more widely known and understood by the general mathematical community.

David S. Watkins, "Francis's Algorithm", 2011

Proper explanation of the “Francis algorithm” [Francis-1959] as it relates to the elementary QR algorithm [Francis-1961,Kublanovskaya-1961],

$$QR := A - \rho I \quad \hat{A} := RQ + \rho I$$

is long and nuanced (see “Francis’s Algorithm” by David S. Watkins). Subspace Rayleigh quotients are crucial for modern interpretations.

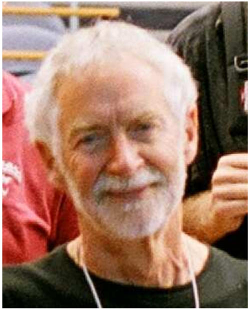


Photo source: David Watkins / Frank Uhlig

Interestingly, Francis implemented the algorithm on a very primitive computer and then moved on from linear algebra, unaware of his contribution

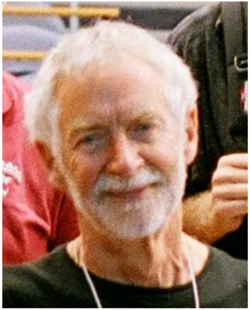
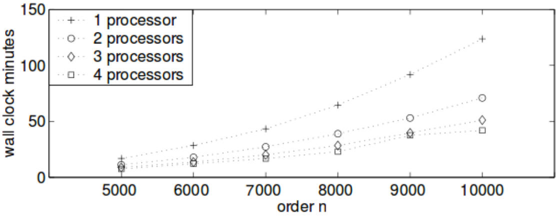


Photo source: David Watkins / Frank Uhlig



Aggressive early deflation recognizes converged eigenvalues before classical small-subdiagonal deflation would. [...] Sometimes, aggressive early deflation reduces execution time from hours down to minutes.

Braman, Byers, and Mathias, *The Multishift QR algorithm. Part II: Aggressive Early Deflation*, SIAM J. Matrix Anal. Appl., Vol. 23, No. 4, 2002.

Table VIII: Execution time (in seconds) on *Bellatrix* for `fullrand` matrices.

$p =$ $p_r \times p_c$	$n = 4000$ S-v180	SISC	NEW	$n = 8000$ S-v180	SISC	NEW	$n = 16000$ S-v180	SISC	NEW	$n = 32000$ S-v180	SISC	NEW
1×1	637	73	50	5377	441	252						
2×2	192	24	21	1594	137	91						
4×4	68	16	13	498	79	63	4505	552	271			
6×6	47	12	11	294	44	39	1886	247	165	∞	1267	901
8×8	36	16	12	204	42	37	1347	184	129	∞	1362	714
10×10	37	14	9	181	39	40	961	140	110	∞	726	525

[...] we believe to have come to a point, where it will be difficult to attain further dramatic performance improvements for parallel nonsymmetric eigensolvers, without leaving the classical framework of QR algorithms.

Granat et al., *Parallel library software for the multishift QR algorithm with Aggressive Early Deflation*, (Preprint, 2013)

Outline

Disclaimer

Dense systems of equations

Dense eigenvalue problems

Interior Point Methods

Interior Point Methods for Quadratic Programs

The *quadratic program*

$$\begin{aligned}&\min_x \frac{1}{2}x^T Qx + c^T x\\ &\text{such that } Ax = b, x \geq 0,\end{aligned}$$

where Q is symmetric positive-semidefinite, is transformed to a sequence of problems

$$\begin{aligned}&\min_x \frac{1}{2}x^T Qx + c^T x - \mu \sum_j \ln x_j\\ &\text{such that } Ax = b,\end{aligned}$$

with $\mu \downarrow 0$.

- Formulation for *linear programs* informally proposed by von Neumann in a 1948 conversation with George Dantzig (the inventor of the simplex algorithm).
- Dantzig followed up by proving that the algorithm converged, albeit with a worst-case convergence rate of 10^{2q} iterations for q digits of accuracy (see Dantzig and Thapa's *Linear Programming: 2: Theory and Extensions*)

- Karmarkar introduced a practical polynomial-time algorithm for solving linear programs in 1984 (Khachian previously proposed the polynomial *ellipsoidal algorithm* in 1979)
- Claims that the method was typically faster than the simplex method are at best controversial
- AT&T patented it in 1985.
- Despite the debate over the validity of the patent, Karmarkar's algorithm reinvigorated Interior Point Methods.

- In 1989, Sanjay Mehrotra introduced his "predictor-corrector" technique, which often drastically reduces the runtime of IPMs
- The basic idea is to use second-order information from the relaxed complementary slackness condition, $x_i z_i = \mu$
- The only computational price is an additional solve with the factored first-order optimality matrix (which is asymptotically faster than the factorization).

- In a Tech Report in 1994, Robert Vanderbei introduced a now widely-used technique for avoiding the need for sophisticated and expensive symmetric-indefinite sparse factorization techniques (essentially, if an entry is too small to divide by, make it larger).
- The commercial code that resulted (LOQO) is apparently still licensed...

In 1998, Anna Altman and Jacek Gondzio proposed a modified version of Vanderbei's scheme which could be interpreted as dynamically adding regularization to the original problem when the factorization algorithm would have otherwise broken down.

In 2006, Gondzio and Grothey reported that a parallelization of the method of Altman and Gondzio had been used to solve a (non-trivial) quadratic program with one billion variables.

Closing

Apologies

- No discussion of sparse-direct or Krylov subspaces
- No discussion of simplex, first-order methods, or cone programming for convex optimization

Acknowledgments

- Nicholas J. Higham's "Accuracy and Stability of Numerical Algorithms" for being a gold-mine of historical quotes.
- Sven Hammarling for suggesting Turing's award speech over his more technical 1980 memoir.
- Stephen Boyd and Emmanuel Candès for having let me repeatedly bug them with questions about convex optimization

Questions?