UNIVERSITY OF MINNESOTA TWIN CITIES

Iterative linear algebra for large scale computations

Yousef Saad

University of Minnesota

ICIAM 2023, Tokyo

August 22nd, 2023

Outline

- Aims: an overview of “iterative linear algebra” [linear systems mostly] with a historical perspective
- Follow flow of ideas over the years [‘big ideas’ or ’breakthroughs’]
- A few thoughts on where the field may be heading

Introduction

- In 1953, George Forsythe published this survey paper:

SOLVING LINEAR ALGEBRAIC EQUATIONS CAN BE INTERESTING

GEORGE E. FORSYTHE[1]

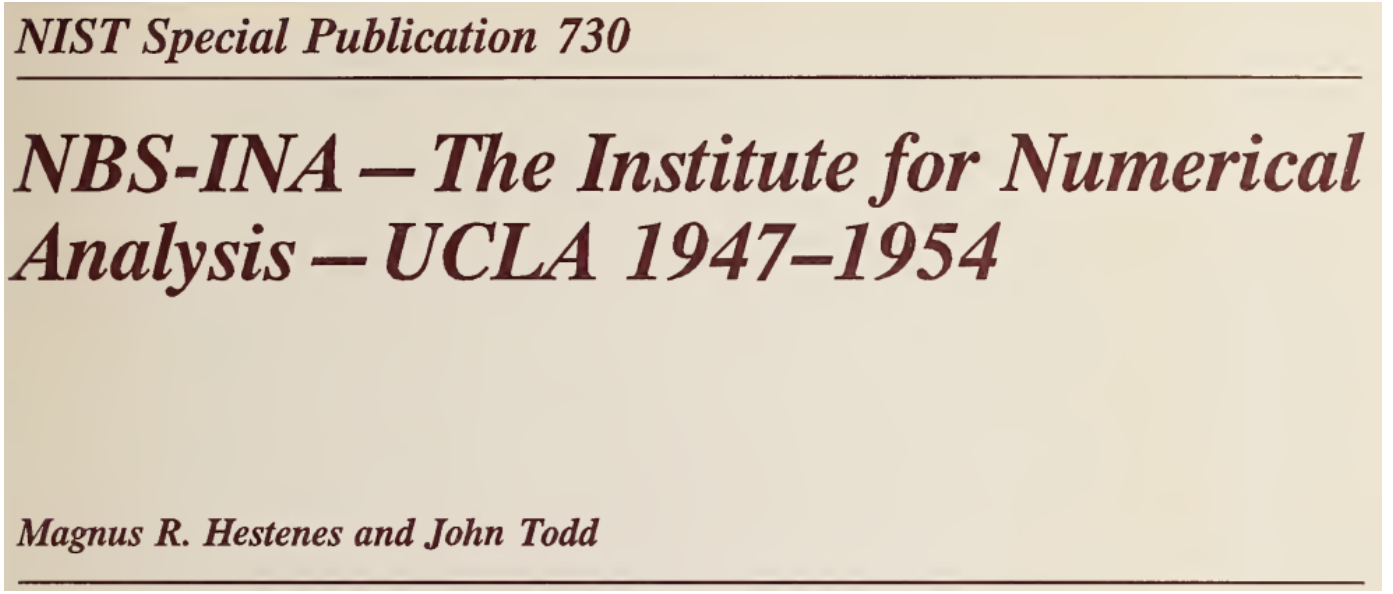
1. **Introduction.** The subject of this talk is mathematically a lowly one. Consider a system of n linear algebraic equations in n unknowns, written

(1) $$Ax = b.$$

Here A is a square matrix of order n, whose elements are given real numbers a_{ij} with a determinant $d(A) \neq 0$; x and b denote column

- A fascinating article with extraordinary vision
- Author urged researchers to start looking at linear systems

- Topics discussed: iterative methods, Conjugate Gradient, condition numbers, arithmetic, ...
- Mysterious footnote: *'original title was Solving linear equations is not trivial'*
- First line says: *The subject of this talk is mathematically a lowly one*
- One must realize: in 1950's Numerical Analysis was new [advent of computers] and not widely accepted yet as a new field by mathematicians



Hestenes and Todd say this in their history of the INA:

"It was Hartree's experience that, by and large, the mathematical community was inexperienced in numerical analysis and showed little interest in the subject. Some even belittled it. In 1948 in an INA symposium (...) Hartree said, 'One of the unsolved problems of Numerical Analysis is how to overcome the attitude of the Mathematical Fraternity on this subject'.

Why iterative methods? A quick look at direct methods

- Oldest method for solving linear systems: Gaussian Elimination
- Starting in the 1960: huge effort and progress on 'sparse direct solvers'
- 1961: link between sparse elimination and graphs [Seymour Parter]
- 1975: Fill-ins and paths [Rose-Tarjan Theorem]
- Late 70s: Elimination Trees [YSMP group at Yale]
- Solution Packages: YSMP, SPARSPAK (70's, 80's)
- 1981: Book by George and Liu

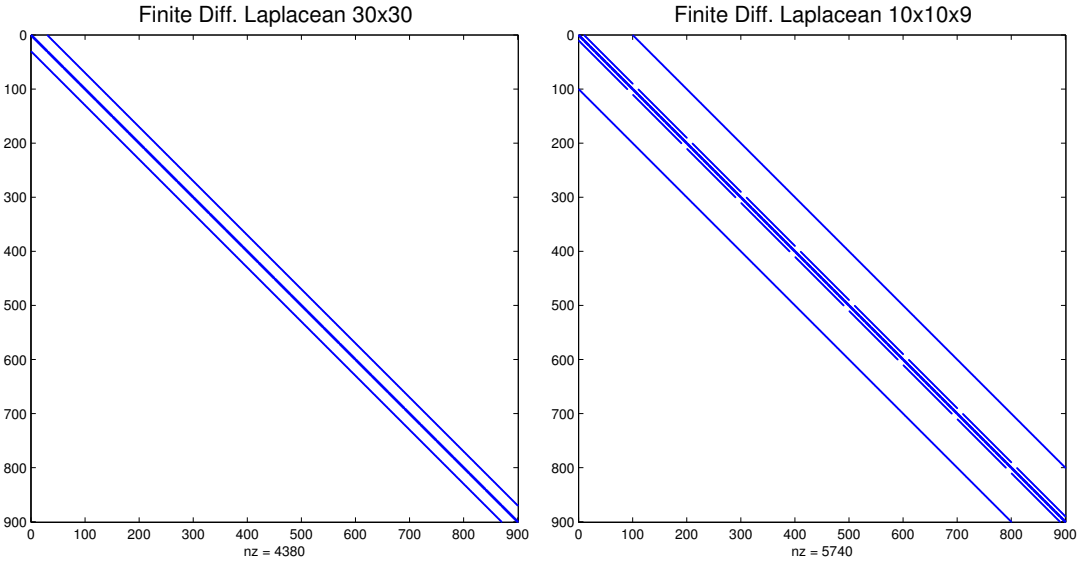
Direct solution of two systems of size $N = 122,500$

Very common misconception: *3-D problems are harder just because they are bigger.* In fact they are intrinsically harder.

First: Laplacean on a 350×350 grid (2D); $N = 122,500$

Second: Laplacean on a $50 \times 50 \times 49$ grid (3D); $N = 122,500$

Patterns of similar [much smaller] matrices



```
>> demo_2Dvs3D
 n = 122500, nnz/n =   4.99
Elapsed time is 0.282512 seconds.
 n = 122500, nnz/n =   6.88
Elapsed time is 2.696554 seconds.
>>
>>
```

First big idea: Relaxation

- Idea goes back to Gauss ($\sim$ 1823).
- Term 'relaxation' likely introduced by Southwell (1940s)
- To solve: $Ax = b$ View as a collection of equations:

  $a_i^T x = \beta_i$ ($a_i^T = i$-th row) for $i = 1, 2, \cdots, n$
- Notation: $r = b - Ax$ (residual)
- *Relaxation:* Modify i-th component of x into $x_i^{(new)} := x_i + \delta_i$ so that: $r_i^{(new)} = 0$. ➤ Do this in a certain order for i
- $$a_i^T(x + \delta e_i) = \beta_i \quad \longrightarrow \quad \delta_i = \frac{r_i}{a_{ii}}$$

Gauss letter to Gerling (1823)

In order to eliminate indirectly,[6] I note that, if 3 of the quantities a, b, c, d are set equal to 0, the fourth gets the largest value when d is chosen as the fourth.[8] Naturally every quantity must be determined from its own equation, and hence d from the fourth. I therefore set $d = -201$ and substitute this value. The absolute terms then become: +5232, −6352, +1074, +46; the other terms remain the same.

Now I let b take its turn, find $b = +92$, substitute, and find the absolute terms: +4036, −4, −3526, −506. And thus I continue until there is nothing more to correct. Of this whole calculation I actually write only the following table:[9]

	$d=-201$	$b=+92$	$a=-60$	$c=+12$	$a=+5$	$b=-2$	$a=-1$
+ 6	+5232	+4036	+ 16	−320	+ 15	+41	−26
− 7558	−6352	− 4	+ 776	+176	+111	−27	−14
−14604	+1074	−3526	−1846	+ 26	−114	−14	+14
+22156	+ 46	− 506	+1054	+118	− 12	0	+26.

Insofar as I carry the calculation only to the nearest 2000-th of a second, I see that now there is nothing more to correct. I therefore collect

Ends letter with:

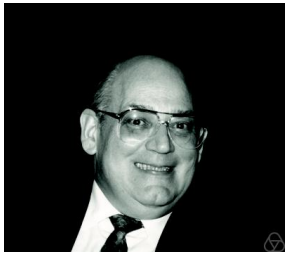
... Almost every evening I make a new edition of the tableau, wherever there is easy improvement. Against the monotony of the surveying business, this is always a pleasant entertainment; one can also see immediately whether anything doubtful has crept in, what still remains to be desired, etc. I recommend this method to you for imitation. You will hardly ever again eliminate directly, at least not when you have more than 2 unknowns.

Ends letter with:

... Almost every evening I make a new edition of the tableau, wherever there is easy improvement. Against the monotony of the surveying business, this is always a pleasant entertainment; one can also see immediately whether anything doubtful has crept in, what still remains to be desired, etc. I recommend this method to you for imitation. You will hardly ever again eliminate directly, at least not when you have more than 2 unknowns. The indirect procedure can be done while half asleep, or while thinking about other things.

- Recommends this iterative scheme *(indirect elimination)* over Gaussian elimination for systems of order >2 (!)

➤ Fast-forward to 1960s:



R. Varga

As an example of the magnitude of problems that have been successfully solved on digital computers by cyclic iterative methods, the Bettis Atomic Power laboratory of the Westinghouse Electric Corporation had in daily use in 1960 a two-dimensional program which would treat as a special case, Laplacean-type matrix equations of order 20,000. Richard Varga, 1962

➤ State of the art in early 1960s was a 20,000 $\times$ 20,000 Laplace equation.

➤ Today: Millions to Tens of Millions is common: CFD, MHD, ...

➤ Models are 3D - so we need iterative methods.

Second big idea: Projection

Goal of projection methods: to extract an approximate solution to a problem from a subspace.

- We define a subspace of approximants of dimension m and a set of m conditions (orthogonality constraints) to extract the solution.

Problem:	(P)
To solve for	$x \in \mathbb{R}^n$

Project $\longrightarrow$

Problem:	$(\widetilde{P})$
To solve for	$\tilde{x} \in K \subseteq \mathbb{R}^n$

- Typically: dimension m of K much smaller than n.
- Related ideas: Model Order Reduction (in Control); Proper Order Decomposition (Solving PDEs); Principal Component Analysis (Stats.)

Projection methods for linear systems

➤ Initial Problem: $Ax = b$

• Given subspaces K and L of dim. m, define Projected problem:

Find $\tilde{x} \in K$ such that $b - A\tilde{x} \perp L$

➤ m degrees of freedom (K) + m constraints (L) $\rightarrow m \times m$ system

➤ Basic projection step. Typically a sequence of such steps are applied

➤ With a nonzero initial guess x_0, approximate problem is

Find $\tilde{x} = x_0 + \delta,\ \delta \in K$ such that $b - A\tilde{x} \perp L \rightarrow$

Find $\delta \in K$ such that $r_0 - A\delta \perp L$

Two Important Particular Cases

❶ $L = K$ When A is SPD then $\|x^* - \tilde{x}\|_A = \min_{z \in K} \|x^* - z\|_A$.

➤ Class of Galerkin or Orthogonal projection methods *(OP) methods*

➤ Important member of this class: Conjugate Gradient (CG) method

❷ $L = AK$ In this case $\|b - A\tilde{x}\|_2 = \min_{z \in K} \|b - Az\|_2$

➤ Class of Minimal Residual–type Methods *(MR) methods*

➤ CR, GCR, ORTHOMIN, GMRES, CGNR, ...

Example: One-dimensional projection processes

Notation: x == current iterate; $r = b - Ax$ = current residual.
$\tilde{x}$ = new iterate

(OP) *Steepest Descent* $\begin{vmatrix} K = \mathbf{span}\{r\} \\ L = K \end{vmatrix}$ → Iteration:

$$\begin{aligned} \alpha &:= (r, r)/(Ar, r) \\ \tilde{x} &:= x + \alpha r \end{aligned}$$

➤ When A is SPD each step minimizes $f(x) = \|x - x_*\|_A^2$ in direction $-\nabla f$.

(MR) *Min. Residual iteration:* $\begin{vmatrix} K = \mathbf{span}\{r\} \\ L = AK \end{vmatrix}$ → Iteration:

$$\begin{aligned} \alpha &:= (Ar, r)/(Ar, Ar) \\ \tilde{x} &:= x + \alpha r \end{aligned}$$

➤ Each step minimizes $f(x) = \|b - Ax\|_2^2$ in direction r.

One-dimensional projection processes

- Steepest descent: Cauchy [1847]. But Kantorovitch [1945] introduced it in the form we know today for linear systems for SPD matrices:

$$\min_x \ \tfrac{1}{2} x^T A x - b^T x$$

- Cimmino's method [1938] and Kaczmarz's method [1937] were also 'Line-search' type methods in the direction of a row^T or a column of A.
- Kaczmarz algorithm played a big role in computer tomography - (Known as ART & Implemented in first medical scanners in 1970s)

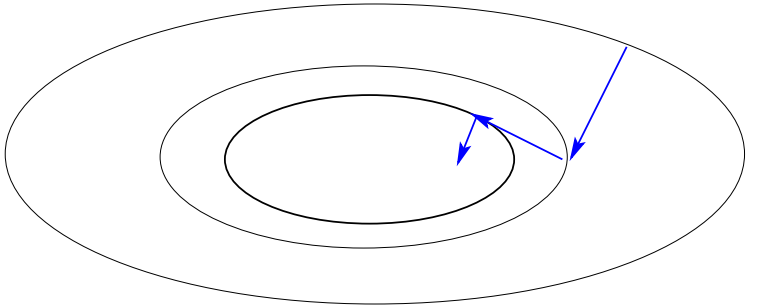
Third big idea: Polynomial acceleration

➤ One-dimensional projection methods are greedy, 'short-sighted', methods

Example:

In Steepest Descent: New direction of search $\tilde{r}$ is $\perp$ to old direction of search r.

$$
\begin{aligned}
r &\leftarrow b - Ax, \\
\alpha &\leftarrow (r, r)/(Ar, r) \\
x &\leftarrow x + \alpha r
\end{aligned}
$$



Question: can we do better by combining successive iterates?

➤ Yes: Polynomial iteration, acceleration, Krylov subspace methods..

Polynomial Acceleration

➤ Consider MR (or steepest descent): $x_{k+1} = x_k + \alpha_k r_k \rightarrow$

$$\begin{aligned} r_{k+1} &= b - A(x_k + \alpha_k r_k) \\ &= r_k - \alpha_k A r_k = (I - \alpha_k A) r_k \end{aligned}$$

- In the end: $r_{k+1} = (I - \alpha_k A)(I - \alpha_{k-1} A) \cdots (I - \alpha_0 A) r_0 = p_{k+1}(A) r_0$

➤ $p_{k+1}(t)$ is a polynomial of degree $k+1$ of the form $p_{k+1}(t) = 1 - t q_k(t)$

➤ Note that: $x_{k+1} = x_0 + q_k(A) r_0$, with deg $(q_k) = k$

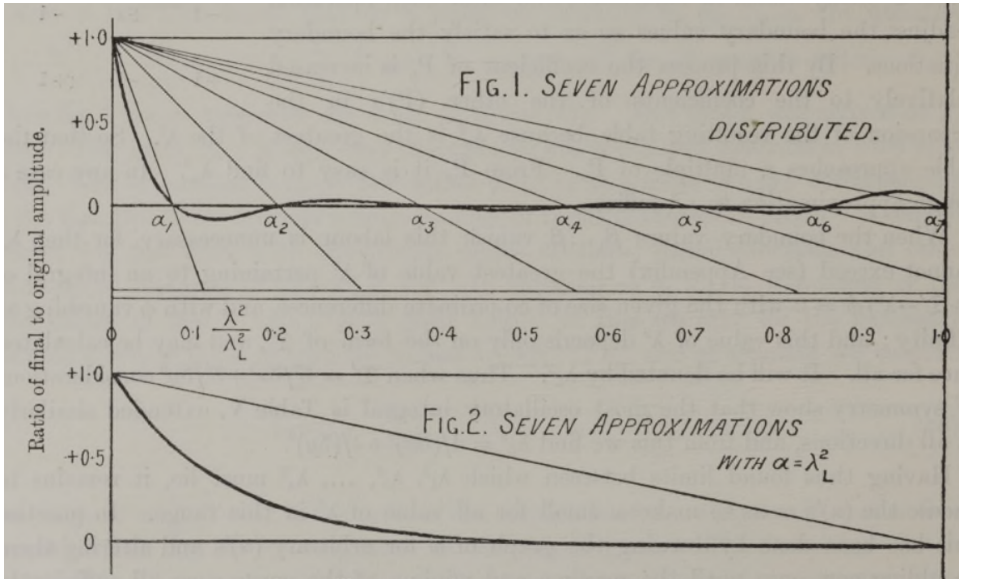
Idea: Acceleration. Build new sequence = 'Best' linear combination of k first iterates ($\sum \eta_{j,k} = 1$)

$$y_k = \sum_{j=0}^{k} \eta_{j,k} x_j$$

➤ Amounts to selecting residual polynomial p_{k+1} that is 'best' in some sense

1910: L. F. Richardson Introduced polynomial acceleration.
Coefficients: Ad-hoc.

➤ Seems unaware of Chebyshev pols.



1952: C. Lanczos mentions Chebyshev polynomials as pre-processing in his MR paper. Different from Chebyshev acceleration.

1953: G. Shortley invokes Chebyshev acceleration. But method impractical [3-term recurrence not exploited]

1954: D. M. Young uses Chebyshev again. Goal: complete what Richardson did in 1910. Uses roots of Chebyshev pol. (No 3-term recurrence)

➤ First Chebyshev acceleration scheme that exploits 3-term recurrence:

1959: A. Blair, N. Metropolis, J. von Neumann, A. H. Taub,& M. Tsingou "A study of a numerical solution to a two-dimensional hydrodynamical problem"

> This problem was discussed and formulated for machine computation by John von Neumann and others. His own original draft of a discussion of the differential and difference equations is given in Appendix I, and an iteration scheme for solving systems of linear equations is given in Appendix II. In the main body of this paper

➤ And 2 years later in:

1961: G. H. Golub & R. Varga 'semi-iterative methods' - Footnote in 3rd page of paper:

> ** This is called "Linear Acceleration" by Forsythe [5]. Professor A. H. Taub has kindly pointed out to us that these results were known much earlier to von Neumann, see [2].*

Chebyshev Acceleration

- Iteration amounts to: $x_{k+1} = x_0 + q_k(A)r_0$ where $q_k =$ pol. of deg. k
- $r_{k+1} = b - Ax_{k+1} = [I - Aq_k(A)]r_0 \equiv p_{k+1}(A)r_0 \quad \longrightarrow$

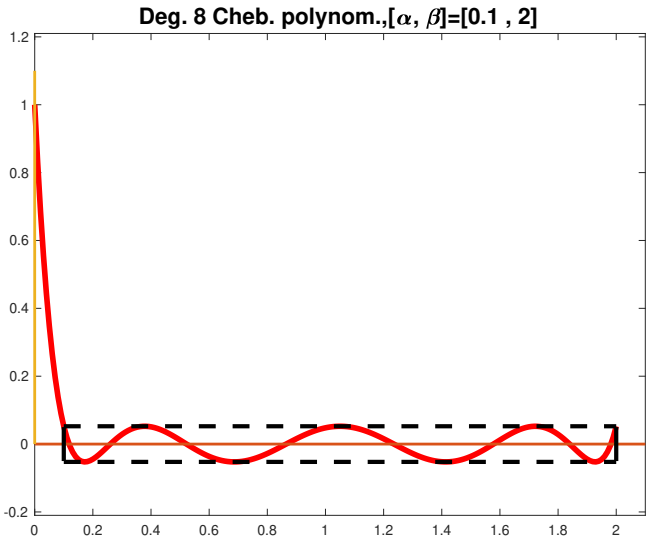
$$p_{k+1}(t) = 1 - tq_k(t)$$

Note: $p_{k+1} \in \mathbb{P}_{k+1,0}$ = set of Pol.s p of degree $k+1$ s.t. $p(0) = 1$.

Problem: Find $p_{k+1} \in \mathbb{P}_{k+1,0}$ such that $p_{k+1}(\lambda_i)$ small for $\lambda_i \in \Lambda(A)$

- When A Symmetric Positive Definite (SPD): $\Lambda(A) \subset [\alpha,\ \beta]$ with $\alpha > 0$
- Idea: Replace $\Lambda(A)$ by $[\alpha,\ \beta]$

Chebyshev Polynomials



Problem: $\min_{p\in\mathbb{P}_{k,0}} \max_{t\in[\alpha,\beta]} |p(t)|$

Solution: Let: $\theta \equiv \frac{\beta+\alpha}{2}, \quad \delta \equiv \frac{\beta-\alpha}{2}.$

➤ Then Solution is:

$T_k(t) \equiv \frac{1}{\sigma_k} C_k\left(\frac{\theta-t}{\delta}\right)$ with $\sigma_k \equiv C_k\left(\frac{\theta}{\delta}\right).$

➤ Three-term recurrence for the Chebyshev polynomials leads to

$\sigma_{k+1} = 2\,\frac{\theta}{\delta}\sigma_k - \sigma_{k-1},\ k = 1, 2\ldots,$ with: $\sigma_1 = \frac{\theta}{\delta}, \quad \sigma_0 = 1,$ and:

$$T_{k+1}(t) \equiv \frac{1}{\sigma_{k+1}}\left[2\,\frac{\theta - t}{\delta}\sigma_k T_k(t) - \sigma_{k-1}T_{k-1}(t)\right] \qquad k \geq 1, \qquad (R)$$
$$= \frac{\sigma_k}{\sigma_{k+1}}\left[2\,\frac{\theta - t}{\delta}T_k(t) - \frac{\sigma_{k-1}}{\sigma_k}T_{k-1}(t)\right], \text{ with } T_1(t) = 1 - \frac{t}{\theta},\ T_0(t) = 1$$

Define $\rho_k \equiv \dfrac{\sigma_k}{\sigma_{k+1}}, \quad k = 1, 2, \ldots.$ Then above recurrences yield:

$$\begin{cases} \rho_k &= \frac{1}{2\sigma_1 - \rho_{k-1}} \\ T_{k+1}(t) &= \rho_k\left[2\left(\sigma_1 - \frac{t}{\delta}\right)T_k(t) - \rho_{k-1}T_{k-1}(t)\right], \quad k \geq 1. \end{cases}$$

➤ Above formulas can start at $k = 0$ if we set $T_{-1} \equiv 0$ and $\rho_{-1} \equiv 0$

Goal: to obtain an iteration that produces residual vector $r_{k+1} = T_{k+1}(A)r_0$

Idea: Exploit difference $r_{k+1} - r_k = -A(x_{k+1} - x_k) = (T_{k+1}(A) - T_k(A))r_0.$

➤ Then use recurrence (R) above. Calculations lead to algorithm:

ALGORITHM : 1 ▪ *Chebyshev Acceleration*

1. $r_0 = b - Ax_0; \sigma_1 = \theta/\delta;$
2. $\rho_0 = 1/\sigma_1; d_0 = \frac{1}{\theta} r_0;$
3. *For* $k = 0, \ldots,$ *until convergence Do:*
4. $\quad x_{k+1} = x_k + d_k$
5. $\quad r_{k+1} = r_k - Ad_k$
6. $\quad \rho_{k+1} = (2\sigma_1 - \rho_k)^{-1};$
7. $\quad d_{k+1} = \rho_{k+1}\rho_k d_k + \frac{2\rho_{k+1}}{\delta} r_{k+1}$
8. *EndDo*

Lines 7 and 4 can be recast into udate:

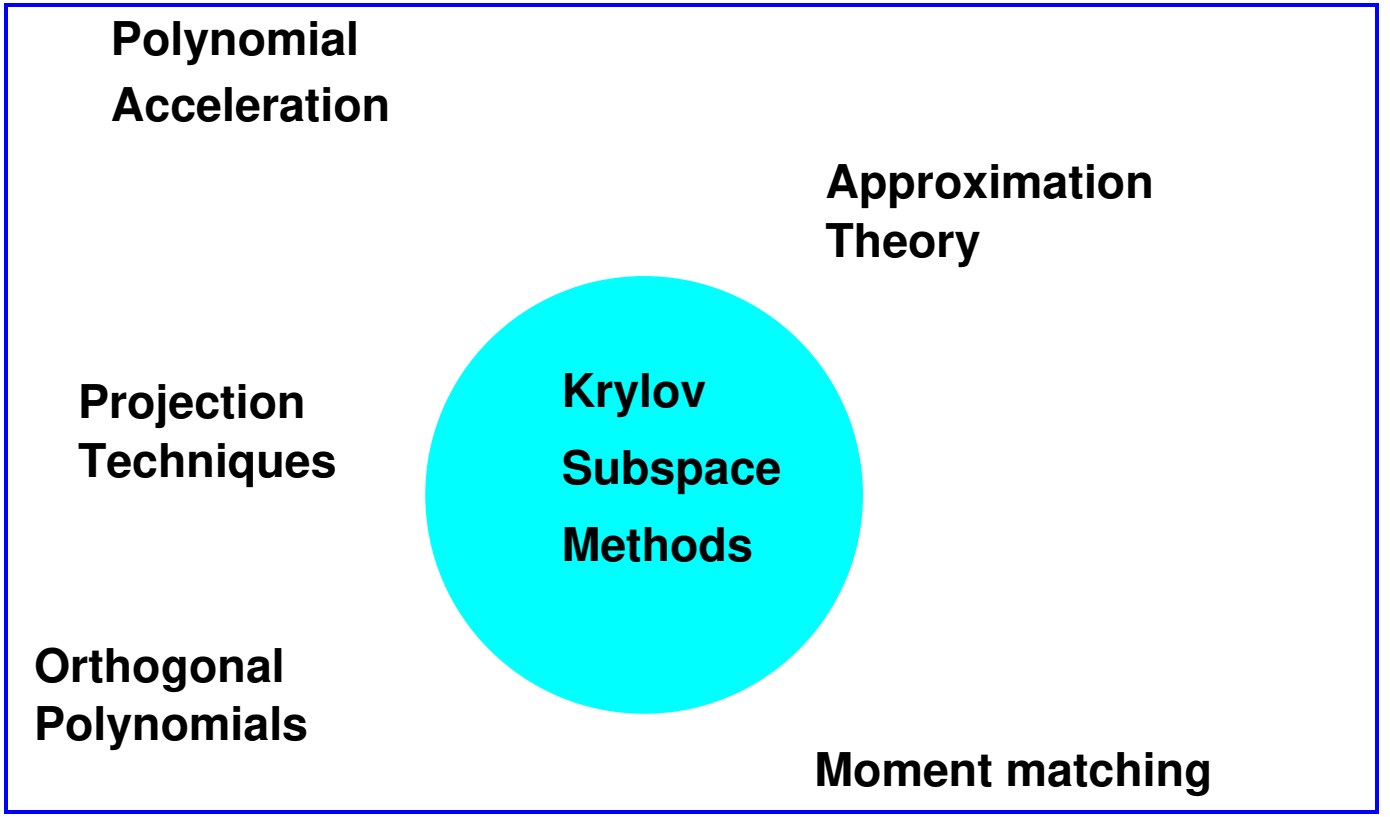
$$x_{k+1} = x_k + \rho_k \left[\rho_{k-1}(x_k - x_{k-1}) + \frac{2}{\delta}(b - Ax_k) \right]$$

➤ Compare this with momentum-type methods

Fourth big idea: Krylov methods

➤ In essence: Optimal polynomial acceleration via projection

Can be viewed from different angles:



Krylov subspace methods for solving $Ax = b$

Definition: A projection method on the subspace

$$K \equiv K_m(A, b) = \mathbf{span}\{b, Ab, \cdots, A^{m-1}b\}$$

- Approximate solution $\rightarrow$ $\tilde{x} = p(A)b$ where p == pol. of deg. $m - 1$

(OP) Case: A is Symmetric Positive Definite. Find $\tilde{x} \in K$ such that $\boxed{b - A\tilde{x} \perp K_m(A, b)}$. Note: $f(x) = \frac{1}{2}x^T Ax - b^T x$ minimized

- One implementation yields the Conjugate Gradient method
- Major breakthrough in early 1950s

Note: We assume initial guess x_0 is 0. Otherwise $\tilde{x} \in x_0 + K$ and $K = K_m(A, r_0)$

Krylov methods take off: The CG algorithm

➤ Magnus Hestenes [UCLA] and Eduard Stiefel [ETH, Zürich] developed the method of Conjugate Gradient independently

M. Hestenes

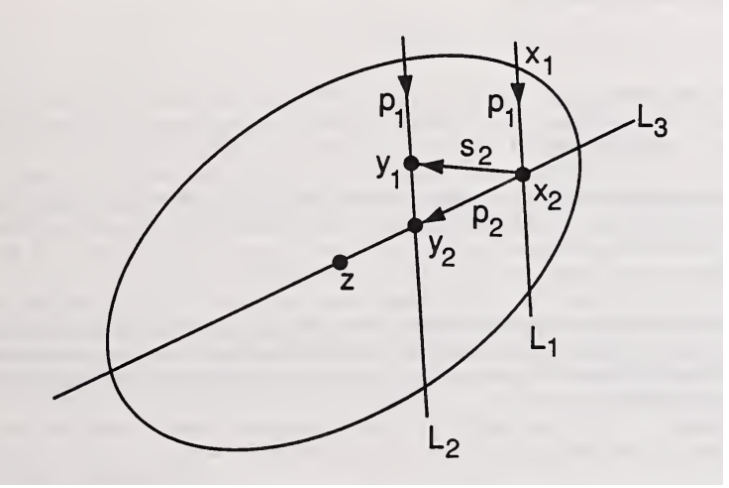
Article:

Methods of conjugate gradients for solving linear systems, Nat. Bur. Standards, 1952.

E. Stiefel

The CG algorithm and 'Conics'

- Hestenes & Stiefel used purely geometric arguments: Ellipses, conics, ...
- Goal: find min. of $f(x) = \frac{1}{2}(Ax, x) - (b, x)$
- The Min. is at the center of the ellipsoids $\{x | f(x) = K\}$



Source: NIST spec. pub. 730

- Min.of $f(x)$ on a chord reached in its middle
- Take 2 parallel chords (L_1, L_2)
- Line L_3 joining minima is a conjugate direction
- Line L_3 passes through center of ellipse
- Min. of $f(x)$ on that line is at center
- $x_2 = x_1 + \alpha_1 p_1$, $y_1 = x_2 + s_2$, $y_2 = y_1 + \beta p_1$
- $p_2 = y_2 - x_2$ == conjugate direction

- C. Lanczos developed a similar method [different notation and viewpoint:]
  Solution of systems of linear equations by minimized iterations, Nat. Bur. Standards (1952)

- A minimal residual (MR) method, implemented with ...
- ... *the Lanczos procedure for eigenvalue problems* [Lanczos '50]
- Lanczos article came out in July '52, Hestenes and Stiefel in Dec. '52
- CG: Not too well received initially, viewed as an unstable, direct method...
- Engeli [1959]: CG 'as iterative process' takes $2n$ or $3n$ to 'converge'
- ... until the early 1970s : paper by John Reid + analysis by Kaniel

Krylov subspace methods for solving $Ax = b$, Nonsymmetric case

Definition: A projection method on the subspace

$$K_m(A, b) = \mathbf{span}\{b, Ab, \cdots, A^{m-1}b\}$$

➤ Approximate solution $\rightarrow$ $\tilde{x} = p(A)b$ where p == pol. of deg. $m - 1$

(MR) Case: Find $\tilde{x}$ such that $\boxed{b - A\tilde{x} \perp AK_m(A, b)}$.
Note $\|b - A\tilde{x}\|_2$ is minimized

➤ Implementation with orthonormal basis of K_m [Arnoldi] $\rightarrow$ GMRES (1)

➤ Other implementations:

- Axelsson's CGLS(2)
- ORTHOMIN (3)
- ORTHODIR (4)
- GCR (5)

---

(1) YS and Schultz '86; (2) O. Axelsson '80; (3) Vinsome, '76;
(4) Young & Jea '80; (5) Eisenstat, Elman, Schultz '83

➤ Sample from 3 decades of prolific activity in Krylov subspace methods:

P.K.W. Vinsome '76 *(ORTHOMIN)*; P. Concus, G.H. Golub, D.P. O'Leary '76 *(Generalized CG)*; C.C. Paige & M. Saunders '82 *(LSQR)*; V. Faber and T. Manteuffel '84 *(Faber-Manteuffel theorem)*; P. Sonneveld, '89 *(CGS)*; P.N. Brown & YS '90 *(Inexact-Newton GMRES)*; R. W. Freund & N.M. Nachtigal '91 *(QMR)*; R. Bramley & A. Sameh '92 *(Projection methods)* ; H. A. Van der Vorst '92 *(Bi-CGSTAB)*; D.Y. Hu, L. Reichel '92 *(Krylov for Sylvester equations)*; G.H. Golub and G. Meurant '93 *(Moments, quadrature)*; N.M. Nachtigal, S.C. Reddy, L.N. Trefethen '92 *(comparing Krylov methods)*; R. W. Freund '93 *(TFQMR)*; R.W. Freund, M.H. Gutknecht, N.M. Nachtigal '93 *(Look-ahead Lanczos)* , L. Zou, H. F. Walker '94 *(residual smoothing)*; X.C. Cai, W.D. Gropp, D.E. Keyes, M.D. Tidriri '94 *(Newton-Krylov)*; D.L. Boley '94 *(Krylov for control)*; P. Feldmann, R.W. Freund' 95 *(Pade via Lanczos)*; O. Axelsson '96*(Book, Generalized versions of CG, CGLS,..)*; A Greenbaum, V Pták, Z Strakoš '96 *(Prescribed convergence of GMRES)*; Greenbaum '97 *(book)*; M. Hochbruck, C. Lubich '97 *(Krylov subs. integrators)*; C.C. Paige, M Rozloznik, Z Strakos '06 *(backward stability)*; ...

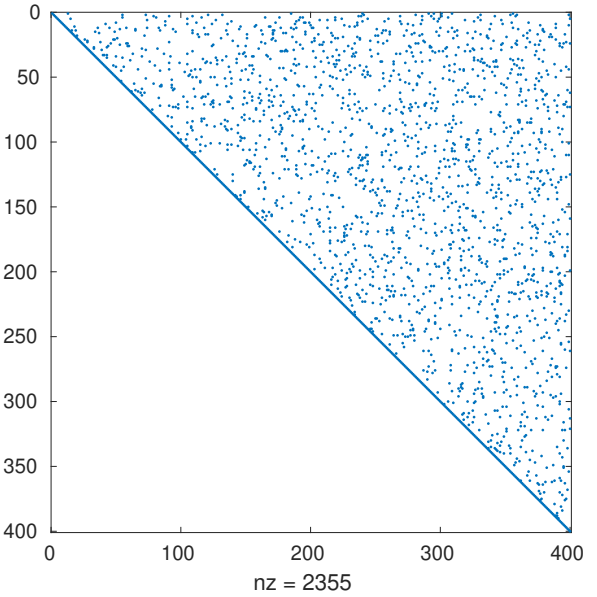
Excellent reference:

has $>$ 1000 refs.

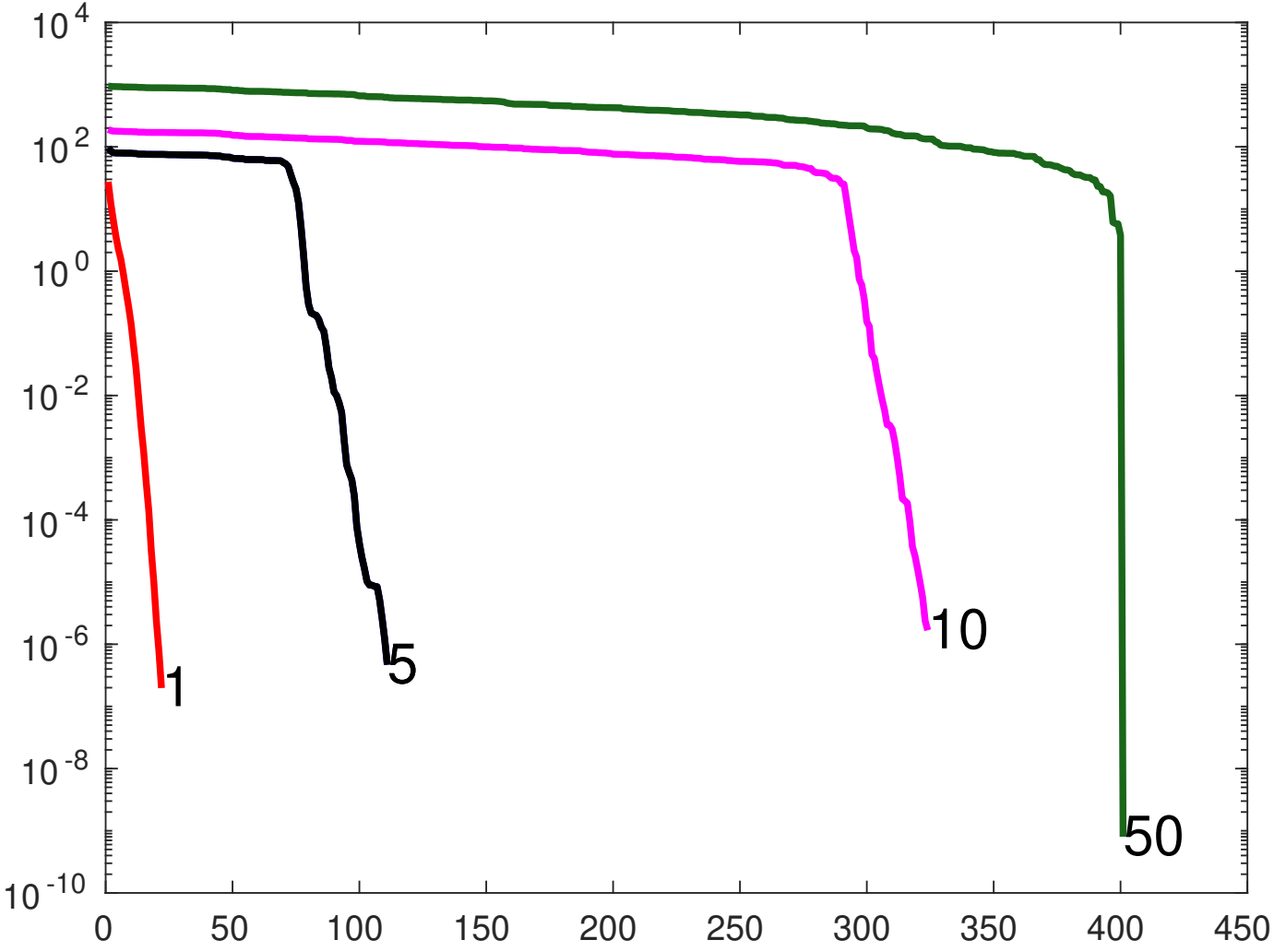
Why is it so hard to analyze convergence in non-normal case?

- First analyses of convergence relied on the eigenvalues: $\rightarrow$ weak results
- What we now know: *"there can be no sharp convergence results that depend on eigenvalues alone"* ... because :
- *For any given spectrum, one can build a system (A and b) that will yield a prescribed convergence of the residual norms* [A. Greenbaum, Z. Strakos'94, A. Greenbaum, V. Ptak, Z. Strakos,'96,...]
- Interesting counter-example-type negative result but ...
- ... We still don't fully understand how Krylov methods converge
- Let us take a look at upper triangular matrices

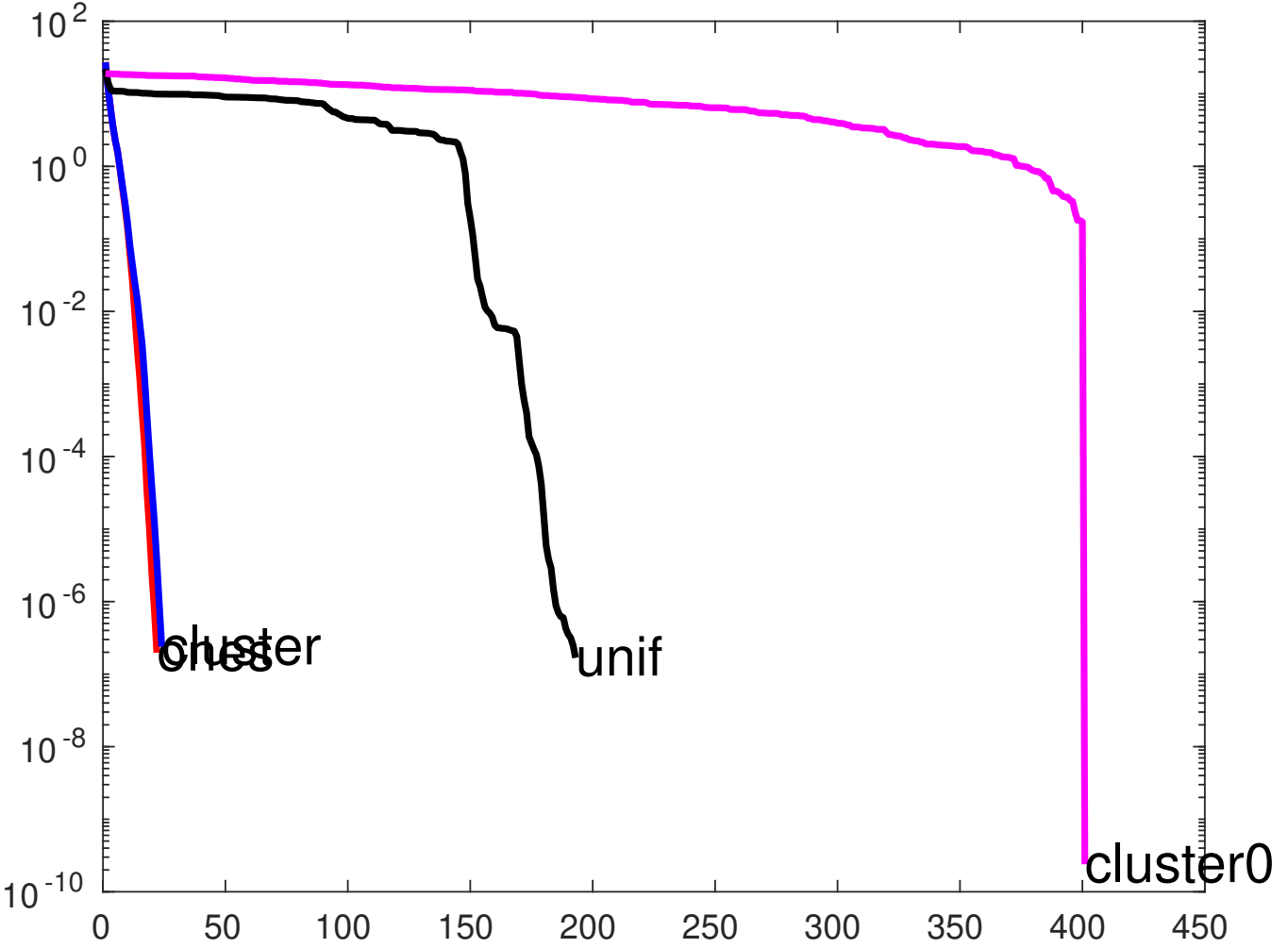
- Let $A = D - \mu E_0$ where:
  - E_0 random strict upper triangular matrix
  - μ selected to vary difficulty (non-normality)
  - Can select E_0 to be sparse or full
  - D (eigenvalues) = either the identity or has some prescribed distribution



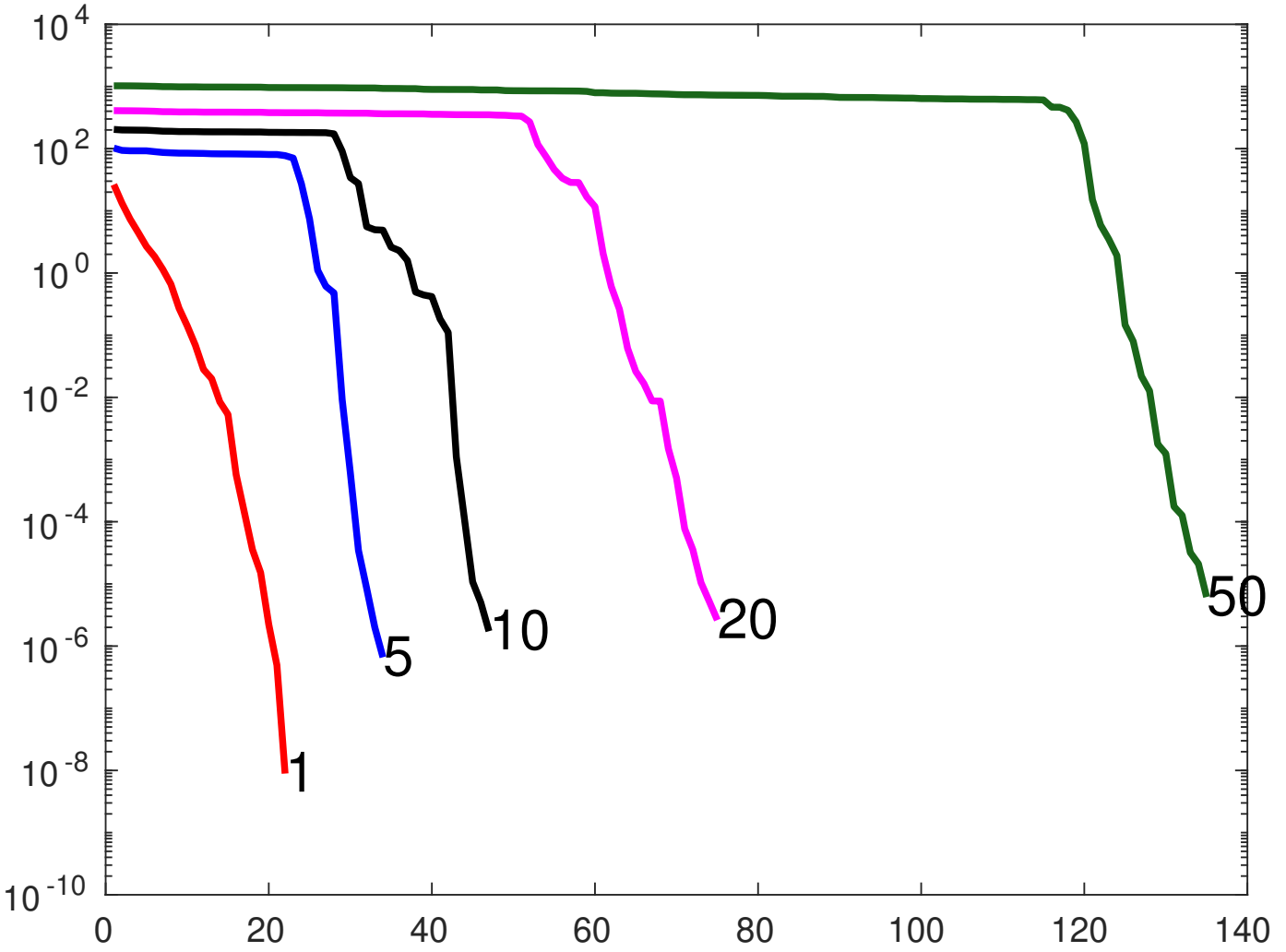
Changing μ, fix $D = I$



Changing spectrum - $\mu = 1$



Changing μ, D = clustered, Sparse case



Observations:

- When $\mu \gg 1$, even $D = I$ can lead to 'non-convergence'
- but spectrum clustered around one seems similar to case $D = I$
- Eigenvalue distribution does play a role - but not the only factor, e.g., ...
- ... Structure plays a role, as does degree of non-normality, etc.
- Difficulty of analysis boils down to this:
  When D and E_0 do not commute not much can be said of $p(D - \mu E_0)$.
- In contrast no issue with $p(I - \mu E_0)$ or $p(D)$ (Hermitian case)

➤ *There can be no 'sharp results' without some assumptions on normality*

➤ Can show a *basic* result by assuming:

❶ departure from normality not large ❷ clustering of spectrum around 1

Main idea: Assume $A = D - E_0$ where E_0 = strict Upper Triang.  Then rewrite A as:

$$A = I - [(I - D) + E_0] \equiv I - E \quad \text{with} \quad E = (I - D) + E_0$$

➤ E now has entries on diagonals (small under 'clustering' assumption.) Take 'test vector:' $x = x_0 + q(A)r_0$ with $\boxed{q(A) = I + E + \cdots + E^{k-1}.}$

Then GMRES residual r_k at step k satisfies:

$\|r_k\| \leq \|(I - Aq(A))r_0\| \equiv \|E^k r_0\| \leq \|E^k r_0\|_1 \longrightarrow 0$ provided: $\|E\|_1 \leq \delta < 1$.

👍 Good news: $\|E_0\|$ small + clustered spectrum $\longrightarrow$ good convergence

➤ This leads to next big idea.

Fifth big idea: Preconditioning

- Idea: use Krylov subspace method to solve $M^{-1}Ax = M^{-1}b$ preconditioner M is close to A in some sense
- e.g., $M = LU$ = incomplete LU factorization of A
- In effect: calling direct methods (or other methods) for help!
- Many other preconditioning ideas, e.g., AMG, Physics-based, Fast-Poisson solvers, ...

1977: IC *(Incomplete Cholesky)* + CG *(ICCG)* [Meijerink and Van der Vorst]

- Idea of preconditioning was old – but '77 ICCG paper was a major catalyst
- Preconditioned Krylov subspace methods gained enormous popularity

Research in iterative linear algebra today

- Active: Preconditioners for some types of problems [Helmholtz, Maxwell, Structures, Data related applications, ..] Parallel implementations, ...
- Not as active: accelerators ... Except: adaptation to new hardware
- Impact of floating point arithmetic
- Some ideas are finding their way to Data Science context, e.g.
  - Peaceman-Rachford's ADI [1955] $\longrightarrow$ ADMM
  - Relaxation techniques $\longrightarrow$ Coordinate descent
  - Steepest Descent $\longrightarrow$ 'Stochatic Gradient Descent' (SGD)
  - Kaczmarz $\longrightarrow$ Randomized Kaczmarz [T. Stohmer, R. Vershyin, '09; D. Needell, R. Ward, N. Srebro '14,..]

- Important new consideration: randomness + *statistical analysis*
- In this context: Standard 'optimal' methods (e.g. CG, GMRES) not as useful
- Instead: Big activity in randomized numerical linear algebra e.g. Projection on randomly generated subspaces *N. Halko, P.G. Martinsson, J.A. Tropp, '11*

So: What will be the next 'Big Idea' in NLA?

- INA experience from late 40s suggest that big ideas result from:

❶ Pressing need to solve well defined problems

❷ Bright researchers with exceptional training and vision

- For ❶ : Right now demand is all about ML & data-related methods
- What about ❷ ? We must strive to disseminate our work with a goal of inspiring readers, one of whom may become the next big star
- That means: efforts to present ideas in an insightful way to spark interest
- + Make resources available, contribute to making science more open, ..
- Problem: Not so easy in current fast-paced & competitive environment

Acknowledgments:

➤ Long list of people to whom I am deeply indebted. Among them:

- My doctorate advisor: Françoise Chatelin (1941-2020)

- My mentors:

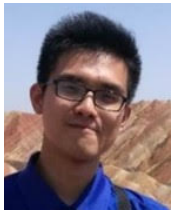
Ahmed Sameh

Bill Gear (1935-2022)

Beresford Parlett

Martin Schultz

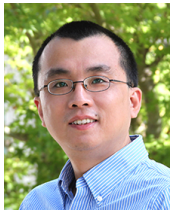
- Students, & *Post-docs*:


Tianshi

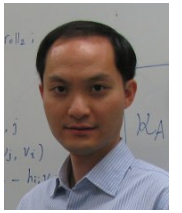

Vasilis

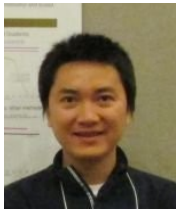

Shashanka

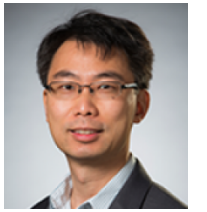

Daniel


Ruipeng


Abdelkader


Jie


Thanh


Edmond


Kesheng


Bernie


Irene


Sangback


Yuanzhe


Agnieszka


Mohammed


Pierre

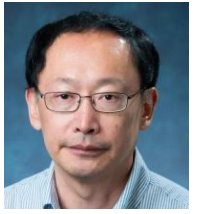

Haw-ren


Suzanne


Yok


Matthias


Emmanuel


Xiao-Chuan


Andreas


Scott


Costas

. . .

+ *24*

more

- You made this (continuing) journey truly amazing. **THANK YOU!**

and ...

Thank you

... for your attention