GMINRES or GLSQR?

Michael Saunders
Systems Optimization Laboratory (SOL)
Institute for Computational Mathematics and Engineering (ICME)
Stanford University

Workshop on Matrix Computations
in Memory of Professor Gene Golub

Institute for Computational Mathematics (ICM)
Hong Kong Baptist University

Outline

Outline

Outline

1. Orthogonal matrix reductions
2. MINRES-type solvers
3. Bi-tridiagonalization of general A

Outline

1. Orthogonal matrix reductions
2. MINRES-type solvers
3. Bi-tridiagonalization of general A
4. Numerical results

Outline

Abstract

Given a general matrix A and starting vectors b, c we can construct orthonormal matrices U_k, V_k that reduce A to tridiagonal form: $AV_k \approx U_k T_k$ and $A^T U_k \approx V_k T_k^T$.

Saunders, Simon, and Yip (1988) proposed methods for solving square systems $Ax = b$ and $A^T y = c$ simultaneously. The solver USYMQR becomes equivalent to MINRES in the symmetric case with $b = c$.

The method was rediscovered by Reichel and Ye (2008) with emphasis on rectangular systems. For implementation reasons it was regarded as a generalization of LSQR (although it does not reduce to LSQR in any special case). The method has been applied to two square systems by Golub, Stoll, and Wathen (2008) with focus on estimating $c^T x$ and $b^T y$.

Orthogonal matrix reductions

Direct:	$V =$ product of Householder transformations	$n \times n$
Iterative:	$V_k = \begin{pmatrix} v_1 & v_2 & \dots & v_k \end{pmatrix}$	$n \times k$
	Mostly short-term recurrences	

Tridiagonalization of symmetric A

Direct:

$$V^TAV = \begin{pmatrix} x & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \end{pmatrix}$$

Tridiagonalization of symmetric A

Direct:

$$V^T A V = \begin{pmatrix} x & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \end{pmatrix} \qquad V^T \begin{pmatrix} 0 & b^T \\ b & A \end{pmatrix} V = \begin{pmatrix} 0 & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \end{pmatrix}$$

Tridiagonalization of symmetric A

Direct:

$$V^TAV = \begin{pmatrix} x & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \end{pmatrix} \qquad V^T\begin{pmatrix} 0 & b^T \\ b & A \end{pmatrix} V = \begin{pmatrix} 0 & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \end{pmatrix}$$

Iterative: Lanczos process $\qquad \begin{pmatrix} b & AV_k \end{pmatrix} = V_{k+1} \begin{pmatrix} \beta e_1 & T_{k+1,k} \end{pmatrix}$

Bidiagonalization of rectangular A

Direct:

$$U^TAV = \begin{pmatrix} x & x & & & \\ & x & x & & \\ & & x & x & \\ & & & x & x \\ & & & & x \\ & & & & \\ & & & & \end{pmatrix}$$

Bidiagonalization of rectangular A

Direct:

$$U^T A V = \begin{pmatrix} x & x & & & \\ & x & x & & \\ & & x & x & \\ & & & x & x \\ & & & & x \\ & & & & \\ & & & & \end{pmatrix} \qquad U^T \begin{pmatrix} b & A \end{pmatrix} V = \begin{pmatrix} x & x & & & \\ & x & x & & \\ & & x & x & \\ & & & x & x \\ & & & & x \\ & & & & \\ & & & & \end{pmatrix}$$

Bidiagonalization of rectangular A

Direct:

$$U^T A V = \begin{pmatrix} x & x & & & \\ & x & x & & \\ & & x & x & \\ & & & x & x \\ & & & & x \\ & & & & \\ & & & & \end{pmatrix} \qquad U^T \begin{pmatrix} b & A \end{pmatrix} V = \begin{pmatrix} x & x & & & \\ & x & x & & \\ & & x & x & \\ & & & x & x \\ & & & & x \\ & & & & \\ & & & & \end{pmatrix}$$

Iterative: Golub-Kahan process $\begin{pmatrix} b & AV_k \end{pmatrix} = U_{k+1} \begin{pmatrix} \beta e_1 & B_k \end{pmatrix}$

Tridiagonalization of rectangular A

Direct:

$$U^T \begin{pmatrix} 0 & c^T \\ b & A \end{pmatrix} V = \begin{pmatrix} 0 & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \\ & & & & x \\ & & & & \end{pmatrix}$$

Tridiagonalization of rectangular A

Direct:

$$U^T \begin{pmatrix} 0 & c^T \\ b & A \end{pmatrix} V = \begin{pmatrix} 0 & x & & & \\ x & x & x & & \\ & x & x & x & \\ & & x & x & x \\ & & & x & x \\ & & & & x \\ & & & & \end{pmatrix}$$

Iterative: S-Simon-Yip (1988), Reichel-Ye (2008)

$$\begin{pmatrix} b & AV_k \end{pmatrix} = U_{k+1} \begin{pmatrix} \beta e_1 & T_{k+1,k} \end{pmatrix}$$
$$\begin{pmatrix} c & A^T U_k \end{pmatrix} = V_{k+1} \begin{pmatrix} \gamma e_1 & T_{k+1,k}^T \end{pmatrix}$$

MINRES-type solvers

based on

Lanczos, Arnoldi, Golub-Kahan, bi-tridiag

MINRES-type solvers for $Ax \approx b$

A	Process			Solver
symmetric	Lanczos	Paige-S	1975	MINRES
rectangular	Golub-Kahan	Paige-S	1982	LSQR
		Fong-S	2011	LSMR
unsymmetric	Arnoldi	Saad-Schultz	1986	GMRES
unsymmetric	bi-tridiag	S-Simon-Yip	1988	USYMQR
rectangular	bi-tridiag	Reichel-Ye	2008	GLSQR

MINRES-type solvers for $Ax \approx b$

A	Process			Solver
symmetric	Lanczos	Paige-S	1975	MINRES
rectangular	Golub-Kahan	Paige-S	1982	LSQR
		Fong-S	2011	LSMR
square	Arnoldi	Saad-Schultz	1986	GMRES
square	bi-tridiag	S-Simon-Yip	1988	GMINRES
rectangular	bi-tridiag	Reichel-Ye	2008	GLSQR

MINRES-type solvers for $Ax \approx b$

A	Process			Solver
symmetric	Lanczos	Paige-S	1975	MINRES
rectangular	Golub-Kahan	Paige-S	1982	LSQR
		Fong-S	2011	LSMR
square	Arnoldi	Saad-Schultz	1986	GMRES
square	bi-tridiag	S-Simon-Yip	1988	GMINRES
rectangular	bi-tridiag	Reichel-Ye	2008	GLSQR

All methods:

$$\begin{pmatrix} b & AV_k \end{pmatrix} = U_{k+1} \begin{pmatrix} \beta e_1 & H_k \end{pmatrix}$$

$$b - AV_k w_k = U_{k+1}(\beta e_1 - H_k w_k)$$

$$\|b - AV_k w_k\| \le \|U_{k+1}\| \|\beta e_1 - H_k w_k\|$$

MINRES-type solvers for $Ax \approx b$

A	Process			Solver
symmetric	Lanczos	Paige-S	1975	MINRES
rectangular	Golub-Kahan	Paige-S	1982	LSQR
		Fong-S	2011	LSMR
square	Arnoldi	Saad-Schultz	1986	GMRES
square	bi-tridiag	S-Simon-Yip	1988	GMINRES
rectangular	bi-tridiag	Reichel-Ye	2008	GLSQR

All methods:

$$\begin{pmatrix} b & AV_k \end{pmatrix} = U_{k+1} \begin{pmatrix} \beta e_1 & H_k \end{pmatrix}$$
$$b - AV_k w_k = U_{k+1}(\beta e_1 - H_k w_k)$$
$$\|b - AV_k w_k\| \le \|U_{k+1}\| \|\beta e_1 - H_k w_k\|$$

$$\Rightarrow \ x_k = V_k w_k \quad \text{where} \quad \min \|\beta e_1 - H_k w_k\|$$

Symmetric methods for unsymmetric $Ax \approx b$

Lanczos on $\begin{pmatrix} I & A \\ A^T & -\delta^2 I \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}$ (general A)

gives Golub-Kahan

CG-type subproblem gives LSQR

MINRES-type subproblem gives LSMR

Symmetric methods for unsymmetric $Ax \approx b$

Lanczos on $\begin{pmatrix} I & A \\ A^T & -\delta^2 I \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}$ (general A)
gives Golub-Kahan

CG-type subproblem gives LSQR
MINRES-type subproblem gives LSMR

Lanczos on $\begin{pmatrix} & A \\ A^T & \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} b \\ c \end{pmatrix}$ (square A)
is not equivalent to bi-tridiagonalization (but seems worth trying!)

Tridiagonalization of general A using orthogonal matrices

Some history of bi-tridiagonalization

Bi-tridiagonalization

- 1988 Saunders, Simon, and Yip, SINUM 25

  "Two CG-type methods for unsymmetric linear equations"

  Focus on square A

  USYMLQ and USYMQR (GSYMMLQ and GMINRES)

Bi-tridiagonalization

- 1988 Saunders, Simon, and Yip, SINUM 25

  "Two CG-type methods for unsymmetric linear equations"

  Focus on square A

  USYMLQ and USYMQR (GSYMMLQ and GMINRES)

- 2008 Reichel and Ye

  "A generalized LSQR algorithm"

  Focus on rectangular A

  GLSQR

Bi-tridiagonalization

- 1988 Saunders, Simon, and Yip, SINUM 25

  "Two CG-type methods for unsymmetric linear equations"

  Focus on square A

  USYMLQ and USYMQR (GSYMMLQ and GMINRES)

- 2008 Reichel and Ye

  "A generalized LSQR algorithm"

  Focus on rectangular A

  GLSQR

- 2008 Golub, Stoll, and Wathen

  "Approximation of the scattering amplitude"

  Focus on $Ax = b$, $A^T y = c$ and estimation of $c^T x$, $b^T y$ (without x, y)

Bi-tridiagonalization

- 1988 Saunders, Simon, and Yip, SINUM 25
  "Two CG-type methods for unsymmetric linear equations"
  Focus on square A
  USYMLQ and USYMQR (GSYMMLQ and GMINRES)
- 2008 Reichel and Ye
  "A generalized LSQR algorithm"
  Focus on rectangular A
  GLSQR
- 2008 Golub, Stoll, and Wathen
  "Approximation of the scattering amplitude"
  Focus on $Ax = b$, $A^T y = c$ and estimation of $c^T x$, $b^T y$ (without x, y)
- 2012 Patrick Küschner, Max Planck Institute, Magdeburg
  Eigenvalues
  Need to solve $Ax = b$ and $A^T y = c$

Original motivation (S 1981)

- CG, SYMMLQ, MINRES work well for symmetric $Ax = b$

Original motivation (S 1981)

- CG, SYMMLQ, MINRES work well for symmetric $Ax = b$
- Bi-tridiagonalization of unsymmetric A is no more than twice the work and storage per iteration

Original motivation (S 1981)

- CG, SYMMLQ, MINRES work well for symmetric $Ax = b$
- Bi-tridiagonalization of unsymmetric A is no more than twice the work and storage per iteration
- If A is symmetric, we get Lanczos and MINRES etc

Original motivation (S 1981)

- CG, SYMMLQ, MINRES work well for symmetric $Ax = b$
- Bi-tridiagonalization of unsymmetric A is no more than twice the work and storage per iteration
- If A is symmetric, we get Lanczos and MINRES etc
- If A is nearly symmetric, total itns should be not much more

Elizabeth Yip's SIAM conference abstract (1982)

CG method for unsymmetric matrices applied to PDE problems

We present a CG-type method to solve $Ax = b$, where A is an arbitrary nonsingular unsymmetric matrix. The algorithm is equivalent to an orthogonal tridiagonalization of A.

Each iteration takes more work than the orthogonal bidiagonalization proposed by Golub-Kahan, Paige-Saunders for sparse least squares problems (LSQR).

Elizabeth Yip's SIAM conference abstract (1982)

CG method for unsymmetric matrices applied to PDE problems

We present a CG-type method to solve $Ax = b$, where A is an arbitrary nonsingular unsymmetric matrix. The algorithm is equivalent to an orthogonal tridiagonalization of A.

Each iteration takes more work than the orthogonal bidiagonalization proposed by Golub-Kahan, Paige-Saunders for sparse least squares problems (LSQR).

However, ... the condition number for our tridiagonalization is the square root of that for the bidiagonalization.

Elizabeth Yip's SIAM conference abstract (1982)

CG method for unsymmetric matrices applied to PDE problems

We present a CG-type method to solve $Ax = b$, where A is an arbitrary nonsingular unsymmetric matrix. The algorithm is equivalent to an orthogonal tridiagonalization of A.

Each iteration takes more work than the orthogonal bidiagonalization proposed by Golub-Kahan, Paige-Saunders for sparse least squares problems (LSQR).

However, ... the condition number for our tridiagonalization is the square root of that for the bidiagonalization.

We apply a preconditioned version (Fast Poisson) to the difference equation of unsteady transonic flow with small disturbances. (Compared with ORTHOMIN(5))

Numerical results with bi-tridiagonalization

Numerical results (SSY 1988)

$$A = \begin{pmatrix} B & -I & & & \\ -I & B & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & B & -I \\ & & & -I & B \end{pmatrix}$$

400×400

$$B = \text{tridiag}\begin{pmatrix} -1-\delta & 4 & -1+\delta \end{pmatrix}$$

20×20

Numerical results (SSY 1988)

$$A = \begin{pmatrix} B & -I & & & \\ -I & B & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & B & -I \\ & & & -I & B \end{pmatrix} \qquad B = \text{tridiag}\begin{pmatrix} -1-\delta & 4 & -1+\delta \end{pmatrix}$$

400×400 $\qquad$ 20×20

Megaflops to reach $\|r\| \le 10^{-6}\|b\|$:

δ	0.0	0.01	0.1	1.0	10.0	100.0
ORTHOMIN(5)	0.31	0.57	0.75	0.83	2.55	2.11
LSQR	0.28	1.38	1.48	0.80	0.57	0.27
USYMQR	0.30	1.88	1.98	1.41	0.99	0.64

Numerical results (SSY 1988)

$$A = \begin{pmatrix} B & -I & & & \\ -I & B & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & B & -I \\ & & & -I & B \end{pmatrix} \qquad B = \text{tridiag}\begin{pmatrix} -1-\delta & 4 & -1+\delta \end{pmatrix}$$

400×400 $\qquad$ 20×20

Megaflops to reach $\|r\| \le 10^{-6}\|b\|$:

δ	0.0	0.01	0.1	1.0	10.0	100.0
ORTHOMIN(5)	0.31	0.57	0.75	0.83	2.55	2.11
LSQR	0.28	1.38	1.48	0.80	0.57	0.27
USYMQR	0.30	1.88	1.98	1.41	0.99	0.64

Bottom line:
ORTHOMIN sometimes good, can fail. LSQR always better than USYMQR

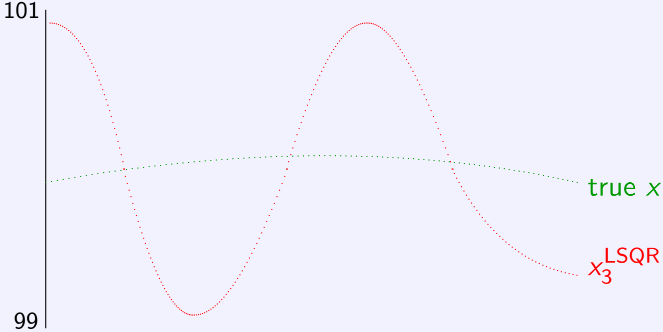
Numerical results (Reichel and Ye 2008)

- Focused on rectangular A and least-squares
  (Forgot about SSY 1988 and USYMQR — hence GLSQR)
- Three numerical examples (all square!)

Numerical results (Reichel and Ye 2008)

- Focused on rectangular A and least-squares
  (Forgot about SSY 1988 and USYMQR — hence GLSQR)
- Three numerical examples (all square!)

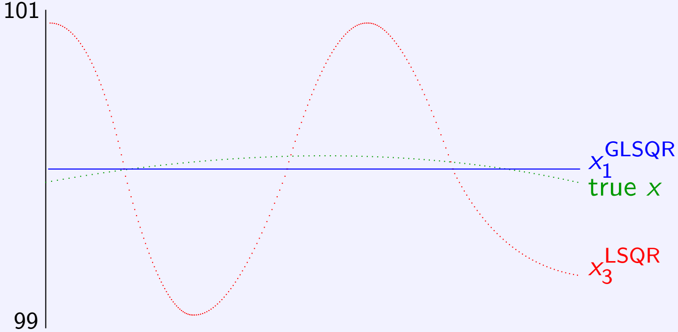
Example 1 ($Ax \approx b$ from Fredholm integral eqn of first kind)



Numerical results (Reichel and Ye 2008)

- Focused on rectangular A and least-squares
  (Forgot about SSY 1988 and USYMQR — hence GLSQR)
- Three numerical examples (all square!)

Example 1 ($Ax \approx b$ from Fredholm integral eqn of first kind)



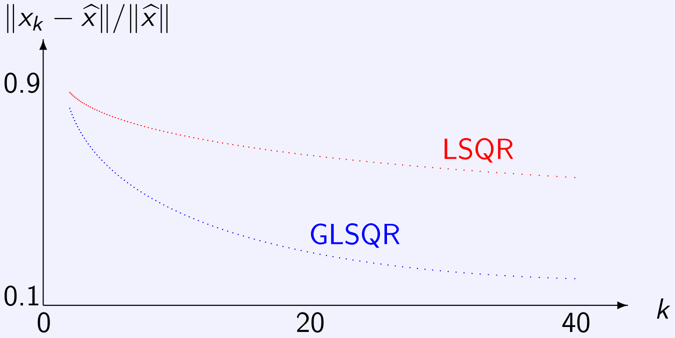
For GLSQR, choose $c = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^T$ because true $x \approx 100c$

Example 2 (Star cluster)

- 256×256 pixels ($n = 65536$), 470 stars
- Square $Ax \approx b$, choose $c = b$
- Compare error in x_k^{LSQR} and x_k^{GLSQR} for 40 iterations

Example 2 (Star cluster)

- 256×256 pixels ($n = 65536$), 470 stars
- Square $Ax \approx b$, choose $c = b$
- Compare error in x_k^{LSQR} and x_k^{GLSQR} for 40 iterations



Conclusions

Subspaces

- Unsymmetric Lanczos generates two Krylov subspaces:

$$\begin{aligned} U_k &\in \operatorname{span}\{b \quad Ab \quad A^2b \quad \dots \quad A^{k-1}b\} \\ V_k &\in \operatorname{span}\{c \quad A^Tc \quad (A^T)^2c \quad \dots \quad (A^T)^{k-1}c\} \end{aligned}$$

Subspaces

- Unsymmetric Lanczos generates two Krylov subspaces:

$$\begin{aligned} U_k &\in \operatorname{span}\{b \;\; Ab \;\; A^2b \;\; \dots \;\; A^{k-1}b\} \\ V_k &\in \operatorname{span}\{c \;\; A^Tc \;\; (A^T)^2c \;\; \dots \;\; (A^T)^{k-1}c\} \end{aligned}$$

- Bi-tridiagonalization generates

$$\begin{aligned} U_{2k} &\in \operatorname{span}\{b \;\; AA^Tb \;\; \dots \;\; (AA^T)^{k-1}b \;\; Ac \;\; (AA^T)Ac \;\; \dots\} \\ V_{2k} &\in \operatorname{span}\{c \;\; A^TAc \;\; \dots \;\; (A^TA)^{k-1}c \;\; A^Tb \;\; (A^TA)A^Tb \;\; \dots\} \end{aligned}$$

Reichel and Ye 2008:
Richer subspace for ill-posed $Ax \approx b$ (can choose $c \approx x$)

Functionals $c^T x$, $b^T y$

- Lu and Darmofal (SISC 2003) use unsymmetric Lanczos with QMR to solve $Ax = b$ and $A^T y = c$ *simultaneously* and to estimate $c^T x$ and $b^T y$ at a *superconvergent rate*:

$$|c^T x_k - c^T x| \approx |b^T y_k - b^T y| \approx \frac{\|b - Ax_k\| \|c - A^T y_k\|}{\sigma_{\min}(A)}$$

Functionals $c^T x$, $b^T y$

- Lu and Darmofal (SISC 2003) use unsymmetric Lanczos with QMR to solve $Ax = b$ and $A^T y = c$ *simultaneously* and to estimate $c^T x$ and $b^T y$ at a *superconvergent rate*:

$$|c^T x_k - c^T x| \approx |b^T y_k - b^T y| \approx \frac{\|b - Ax_k\| \|c - A^T y_k\|}{\sigma_{\min}(A)}$$

- Golub, Stoll and Wathen (2008) use bi-tridiagonalization with GLSQR to do likewise

Functionals $c^T x$, $b^T y$

- Lu and Darmofal (SISC 2003) use unsymmetric Lanczos with QMR to solve $Ax = b$ and $A^T y = c$ *simultaneously* and to estimate $c^T x$ and $b^T y$ at a *superconvergent rate*:

$$|c^T x_k - c^T x| \approx |b^T y_k - b^T y| \approx \frac{\|b - Ax_k\| \|c - A^T y_k\|}{\sigma_{\min}(A)}$$

- Golub, Stoll and Wathen (2008) use bi-tridiagonalization with GLSQR to do likewise
  - Matrices, moments, and quadrature

Functionals $c^T x$, $b^T y$

- Lu and Darmofal (SISC 2003) use unsymmetric Lanczos with QMR to solve $Ax = b$ and $A^T y = c$ *simultaneously* and to estimate $c^T x$ and $b^T y$ at a *superconvergent rate*:

$$|c^T x_k - c^T x| \approx |b^T y_k - b^T y| \approx \frac{\|b - Ax_k\| \|c - A^T y_k\|}{\sigma_{\min}(A)}$$

- Golub, Stoll and Wathen (2008) use bi-tridiagonalization with GLSQR to do likewise
  - Matrices, moments, and quadrature
  - Golub, Minerbo, and Saylor
    Nine ways to compute the scattering cross-section
    (1): Estimating $c^T x$ iteratively

Block Lanczos

The bi-tridiagonalization process is equivalent to

- block Lanczos on $A^T A$ with starting block $(c \;\; A^T b)$
  Parlett 1987

Block Lanczos

The bi-tridiagonalization process is equivalent to

- block Lanczos on A^TA with starting block $\begin{pmatrix} c & A^Tb \end{pmatrix}$
  Parlett 1987
- block Lanczos on $\begin{pmatrix} & A \\ A^T & \end{pmatrix}$ with starting block $\begin{pmatrix} & b \\ c & \end{pmatrix}$
  Golub, Stoll, and Wathen 2008

Block Lanczos

The bi-tridiagonalization process is equivalent to

- block Lanczos on A^TA with starting block $\begin{pmatrix} c & A^Tb \end{pmatrix}$
  Parlett 1987
- block Lanczos on $\begin{pmatrix} & A \\ A^T & \end{pmatrix}$ with starting block $\begin{pmatrix} & b \\ c & \end{pmatrix}$
  Golub, Stoll, and Wathen 2008

There are two ways of spreading light.
To be the candle
or the mirror that reflects it.
– Edith Wharton

References

- M. A. Saunders, H. D. Simon, and E. L. Yip (1988).
  Two conjugate-gradient-type methods for unsymmetric linear equations,
  SIAM J. Numer. Anal. 25:4, 927–940.
- L. Reichel and Q. Ye (2008).
  A generalized LSQR algorithm,
  Numer. Linear Algebra Appl. 15, 643–660.
- G. H. Golub, M. Stoll, and A. Wathen (2008).
  Approximation of the scattering amplitude and linear systems,
  ETNA 31, 178–203.

Gene is with us every day