Chapter 2: Convergence (behavior) in exact arithmetic

2.3 Non-Hermitian, but normal

Normal matrices have full set of eigenvectors forming the basis which can be chosen orthonormal. Therefore the change to (orthonormal) eigenvector coordinates does not involve any distortion of geometry.

Substantial difference which causes enormous technical difficulties in proofs and in deriving bounds - the eigenvalues are not real.

In the remaining part of Chapter 2 we restrict ourselves to the GMRES method.

Minimal residual methods

$$\|r_n\| \;=\; \min_{u \in x_0 + K_n(A,r_0)} \|b - Au\| \;=\; \min_{z \in AK_n(A,r_0)} \|r_0 - z\|$$

$$\Leftrightarrow \;\; r_n \perp AK_n(A, r_0)\,.$$

(Hermitian) MINRES [Paige, Saunders - 75] and its general **simplification** GMRES [Saad, Schultz - 86]; mathematically equivalent to GCR analyzed in [Elman - 1982], and to many other (mostly numerically inferior) methods.

MINRES IS NOT a symmetric variant of GMRES!

Implementation of GMRES [Saad, Schultz - 86]

- Arnoldi basis $\{v_1 = r_0/\|r_0\|, v_2, \ldots, v_n\}$, $\quad AV_n = V_{n+1}H_{n+1,n}$.

- $x_n = x_0 + V_n y_n$,
  $\|\|r_0\|e_1 - H_{n+1,n}\, y_n\| = \min_y \|\|r_0\|e_1 - H_{n+1,n}\, y\|$.

In the normal case still [Joubert -93], [Gurvits, Greenbaum - 93], [Trefethen - 93]

$$\frac{\|r_n\|}{\|r_0\|} = \min_{p\in\Pi_n} \| p(A)q_1 \| \leq \max_{\|q\|=1} \min_{p\in\Pi_n} \| p(A)q \| = \min_{p\in\Pi_n} \| p(A) \| .$$

Chapter 2: Convergence (behavior) in exact arithmetic

2.4 Non-normal case

1. Matrix polynomial and worst case bound

2. Operator approach

3. Another look

4. Pathological initial residuals?

5. Convection-diffusion model problem

2.4.1 Matrix polynomial and worst case bound

For a general matrix it can happen

$$\frac{\|r_n\|}{\|r_0\|} = \min_{p\in\Pi_n} \| p(A)q_1 \| \leq \max_{\|q\|=1} \min_{p\in\Pi_n} \| p(A)q \| \neq \min_{p\in\Pi_n} \| p(A) \| .$$

Moreover, there are matrices for which the gap between the worst case bound and the norm of the minimal operator polynomial is large.

[Toh - 96], [Joubert, Faber, Knill, Manteuffel - 94]

For such A and an arbitrary r_0 GMRES performes much better than any analysis based only on A would suggest. This fact represents a general warning for the operator approach. Its significance is, however, unclear yet.

Another issue is, that the GMRES behavior for some particular initial residual (determined, e.g., by the outer forces and the boundary conditions of the physical problem) can dramatically differ from the worst case behavior.

2.4.2 Operator approach

Eigenvalues?

A is diagonalizable: $\dfrac{\|r_n\|}{\|r_0\|} \leq \|X\| \, \|X^{-1}\| \min_{p \in \Pi_n} \max_i |p(\lambda_i)| \, .$

When $\|X\| \, \|X^{-1}\|$ is reasonably bounded, then eigenvalues tell the story.

Diagonalizable matrices with simple eigenvalues form a dense open set in the space of all matrices (a consequence of the Schur theorem). Can this fact be used in GMRES convergence analysis for defective marices? In general, it can not. We can restrict the change of the eigenvalues, but the condition number of the eigenvector matrix may grow to infinity.

Jordan form : $\dfrac{\|r_n\|}{\|r_0\|} \leq \|S\| \, \|S^{-1}\| \min\limits_{p \in \Pi_n} \|p(J)\|$

The identities and bounds will contain derivatives of p_n at the defective eigenvalues with degrees determined by the size of the particular Jordan blocks.

A natural idea how to include non-normality: **Cauchy integral representation**

$$f(A) = \frac{1}{2\pi i} \int_\Gamma (zI - A)^{-1} \, f(z) \, dz$$

$\Gamma \ldots$ a simple closed curve (or union of the closed curves) containing the spectrum of A. Focus on the growth of the rezolvent $(zI - A)^{-1}$ in the neighborhood of the spectrum.

ε - **Pseudospectra:**

$\| (zI - A)^{-1} \| = 1/\varepsilon$ on the boundary Γ_ε. Then

$$\| f(A) \| \leq \frac{\mathcal{L}(\Gamma_\varepsilon)}{2\pi\varepsilon} \max_{z \in \Gamma_\varepsilon} |f(z)| .$$

The GMRES bound

$$\| r^n \| / \| r^0 \| \leq \frac{\mathcal{L}(\Gamma_\varepsilon)}{2\pi\varepsilon} \min_{p \in \Pi_n} \max_{z \in \Gamma_\varepsilon} |p(z)|$$

however, may give a large overestimate.

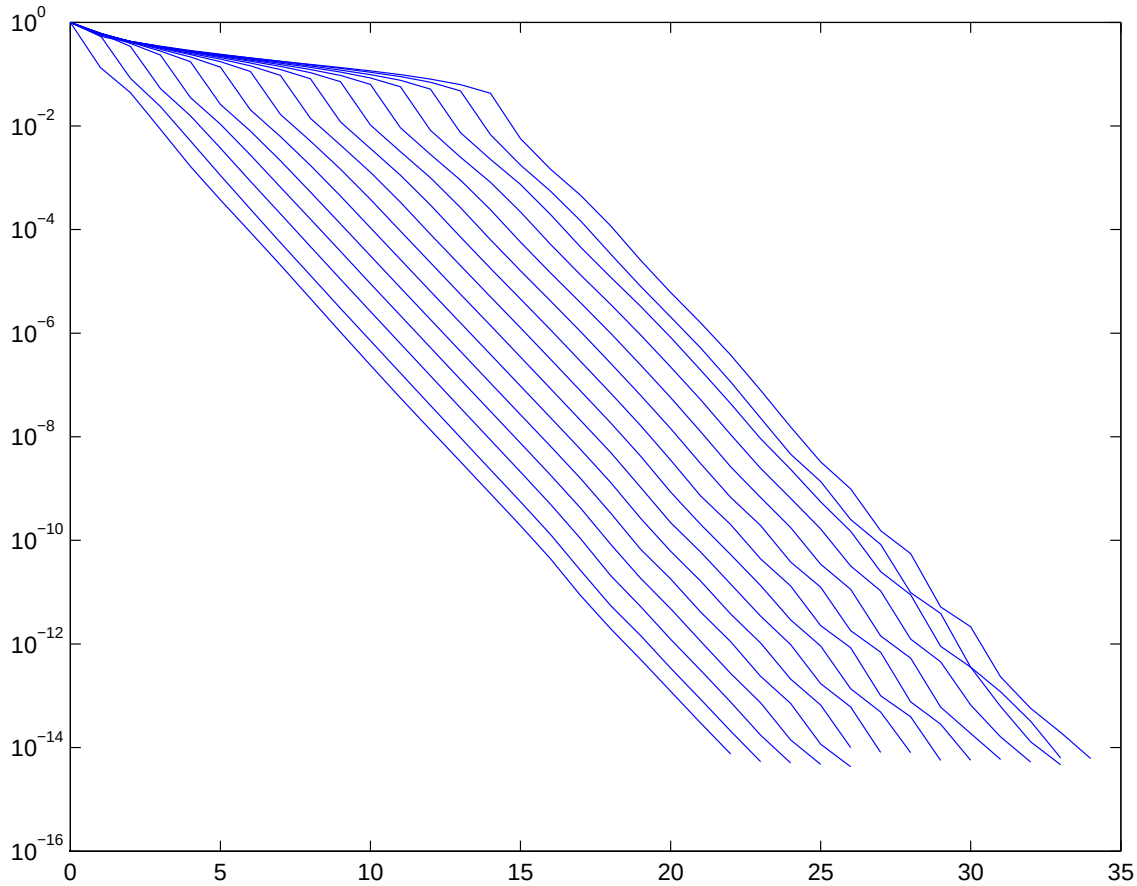
[Trefethen - 91], [Trefethen - 95], [Greenbaum, S - 94]

Enclosures, asymptotic convergence factors, conformal mapping, potential theory, polynomial numerical hull of degree k ...
Recent GMRES convergence results:

Greenbaum, Eiermann, Ernst, Liesen, Nevanlinna, Huhtanen, Trefethen, Embree, Knizhnerman, Nachtigal, Toh, Starke, Hochbruck, Lubich, Frommer, Morgan, Meyer, Ipsen, Van der Vorst, Fischer, Reichel, Calvetti, Simoncini, Bertaccini, Ng, Serra-Capizzano, ...

Any operator bound separating A from b neglects the dependence of the convergence behavior on b. In the best case it therefore represents bound for the worst case behavior.

Are we interested in the worst case?



2.4.3 Another look

For a given fixed r_0 , try to find B such that GMRES$(A, r_0) \equiv$ GMRES(B, r_0) and GMRES(B, r_0) can be analyzed, e.g. B normal whose eigenvalues can be related to some simple properties of A. Then we can analyze GMRES(A, r_0) in terms of these simple properties.

- There is always an equivalent unitary matrix.
- If zero is outside the field of values of A then there is an equivalent HPD matrix. The given behavior of GMRES(A, r_0) can be generated by a HPD matrix B if and only if $\| r_0 \|, \| r_1 \|, \ldots, \| r_n \|$ is monotonically decreasing.

- The given behavior of GMRES(A, r_0) is generated by a Hermitian matrix B if and only if $\| r_n \|$ decreases every two consecutive steps.

But, relationship between the eigenvalues of B and some special properties of A?

Moreover, there is an example of A diagonalizable, eigenvalues "essentially" determine the GMRES(A, r), and for some r_0 there can be no normal B, GMRES$(A, r_0) \equiv$GMRES(B, r_0), whose eigenvalues are close to those of A.

[Greenbaum, S - 93]

Eigenvalues and convergence?

Consider any nonzero eigenvalues and a sequence (desired convergence curve)

$$f(0) \geq f(1) \geq \cdots \geq f(n-1) > f(N) = 0\,.$$

Then the size of the component of the initial residual eliminated in the jth GMRES step is $\phi_j = (f(j-1)^2 - f(j)^2)^{\frac{1}{2}}$. Let R_{N-1} be nonsingular upper triangular, $h = (\phi_1, \ldots, \phi_N)^T$, be the first column of

$$\Phi = \left(\begin{array}{c|c} h & \\ \hline & R_{n-1} \end{array} \right),$$

Theorem [Greenbaum], [Arioli, Pták, S - 97]

The following two assertions are equivalent:

1° *The spectrum of* A *is* $\{\lambda_1, \ldots, \lambda_N\}$ *and* *GMRES*(A, r_0) *yields residuals such that* $\|r_k\| = f(k)\,, \quad k = 0, 1, \ldots, N\,.$

2° $A = U\,(\Phi\, C\, \Phi^{-1})\, U^*$ *and* $r_0 = Uh$ *where* C *is the companion matrix corresponding to the spectrum of* A *and* U *is unitary.*

Theorem gives a complete parametrization of the set of **all pairs** $\{A, r_0\}$ for which GMRES gives the prescribed convergence curve while the matrix A has the prescribed eigenvalues.

Bound by Elman step by step for A normal:

$$\|r_n\| = \|p_n(A)r_0\| = \min_{p\in\Pi_n} \|p(A)r_0\| = \min_{p\in\Pi_n} \| Y\, [p(\Lambda)\, Y^* r_0]\|$$

$$= \min_{p\in\Pi_n} \|p(\Lambda)\, Y^* r_0\| = \min_{p\in\Pi_n} \{ \sum_i |\, (y_i^* r_0)\, p(\lambda_i)\, |^2 \}^{\frac{1}{2}}$$

$$\leq \|r_0\| \min_{p\in\Pi_n} \max_i |p(\lambda_i)|\, .$$

$p_n(\lambda_i)$ represents a multiplicative correction to the values of the individual components of r_0 in the orthonormal basis $\{y_1, \ldots, y_N\}$ in order to minimize the sum of squares.

Bound by Elman step by step for A diagonalizable:

$$\begin{aligned}
\|r_n\| &= \|p_n(A)r_0\| = \min_{p\in\Pi_n} \|p(A)r_0\| = \min_{p\in\Pi_n} \| Y\, [\, p(\Lambda)\, Y^{-1} r_0]\| \\
&\leq \|Y\| \min_{p\in\Pi_n} \|p(\Lambda)\, Y^{-1} r_0\| = \|Y\| \min_{p\in\Pi_n} \{\sum_i |\, [Y^{-1}r_0]_i\, p(\lambda_i)\,|^2\}^{\frac{1}{2}} \\
&\leq \|Y\|\, \|Y^{-1}r_0\| \min_{p\in\Pi_n} \max_i |p(\lambda_i)| \\
&\leq \|r_0\|\, \kappa(Y) \min_{p\in\Pi_n} \max_i |p(\lambda_i)|\,.
\end{aligned}$$

For a general Y , some of the components $Y^{-1}r_0$ can become very large. In such case $Y\,[\,p(\Lambda)\,Y^{-1}r_0]$ represents a significant cancelation. The minimization problem

$$\|r_n\| \;=\; \min_{p\in\Pi_n} \|\,Y\,[\,p(\Lambda)\,Y^{-1}r_0]\|$$

reflects that, while the term in the bound

$$\|Y\|\;\min_{p\in\Pi_n} \|\,p(\Lambda)\,Y^{-1}r_0\,\|$$

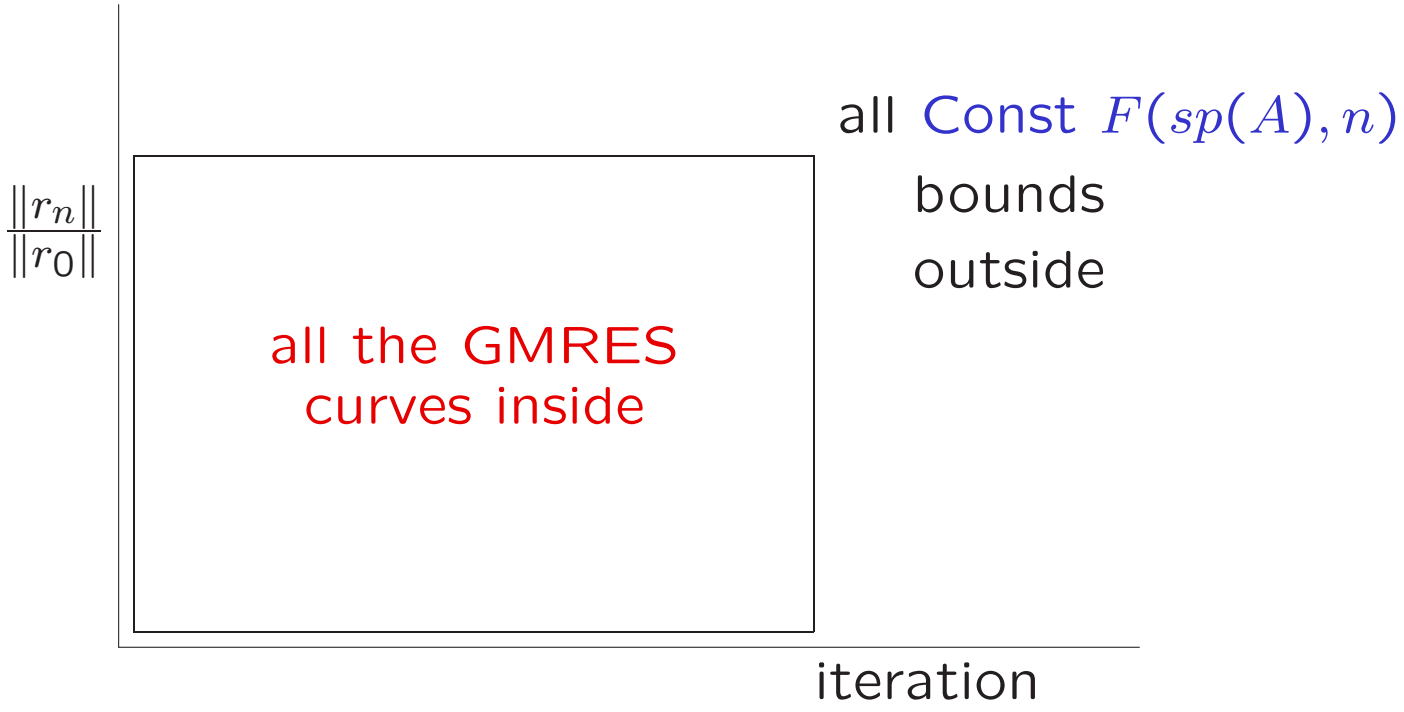
does not (cf. [Trefethen-97]).

The problem of "constants" in the bounds of the type

$$\| r_n \| \leq C(A, r_0)\, F(sp(A), n)\,.$$

If conclusion is based only on $F(sp(A), n)$ and the dependence of $C(A, r_0)$ on the data is not included, then the bound must hold for any data. Consequently, the bound is for any finite dimensional problem irrelevant, otherwise we get a contradiction with the Theorem.

The bound Const $F(sp(A), n)$ does not intersect the rectangle $(1, 0) - (1, n) - (0, n) - (0, 0)$.

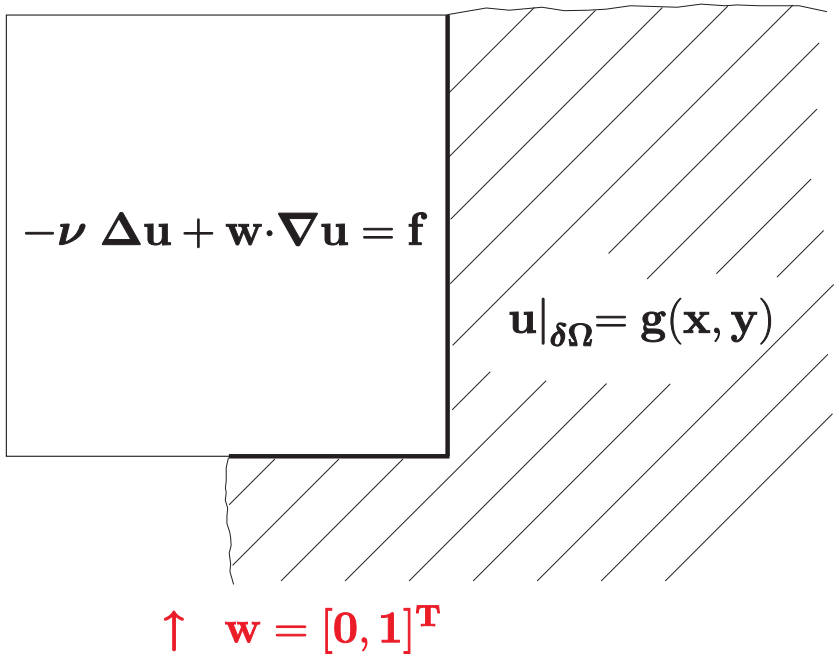


2.4.4 Pathological initial residuals?

This skeptical view seems to be in conflict with the common wisdom – convergence is commonly related to eigenvalue distribution even for general matrices without examining eigenvectors. The proved facts should not be ignored (even a common knowledge can be wrong), but they should be understood and interpreted correctly! There are good reasons for linking convergence to eigenvalues in many cases, but the reasons **must be given and examined** (contrary to common practice).

The role of "pathological initial residuals"; just academic examples ? Not true. Convection-diffusion examples were described by Trefethen long ago, see also [Ernst - 00].

2.4.5 Convection-diffusion model problem



Convection dominated: $\nu \ll \|w\|$

Discretization

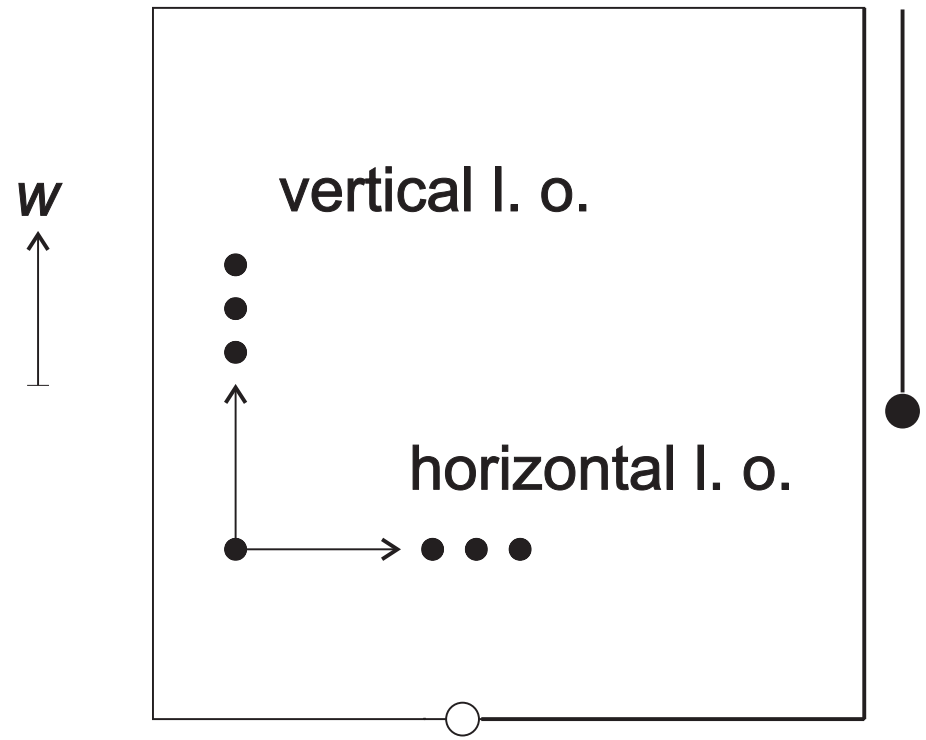
- regular $h \times h$ grid, $h = 1/(N+1)$, bilinear finite elements, mesh Peclet number $P_h \equiv (h\|w\|)/(2\nu)$;

- $P_h > 1$, then Galerkin discretization produces wiggles (non-physical oscillations near the boundary layers);

- Streamline Upwind Petrov Galerkin (SUPG) equivalent to adding stabilizing diffusion in the direction of the flow (wind);

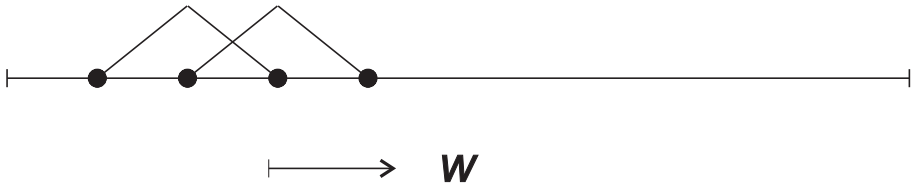
- wind parallel to the mesh; here the vertical wind

$$w = [0, 1]^T .$$



With our choice of w, the differential equation is separable, and the eigendecompozition of the discretized operator is known analytically.

Consider the mass (M) , stiffness (K) and gradient (C) matrices of the corresponding $1D$ convection-diffusion model problem discretized using linear elements with the mesh size h,



$$M = \frac{h}{6}\,\text{tridiag}\,(1,4,1)\,,\ K = \frac{1}{h}\,\text{tridiag}\,(-1,2,-1)\,,$$

$$C = \frac{1}{2}\,\text{tridiag}\,(-1,0,1)\,.$$

Let ' $\otimes$ ' denote the Kronecker product of matrices.

Then the $2D$ SUPG discretized $N^2 \times N^2$ (operator is for the horizontal line ordering of unknowns

$$A_H = \nu M \otimes K + ((\nu + \delta h)K + C) \otimes M \,,$$

for the vertical line ordering of unknowns

$$A_V = \nu K \otimes M + M \otimes ((\nu + \delta h)K + C) \,.$$

A_H and A_V are orthogonally similar,

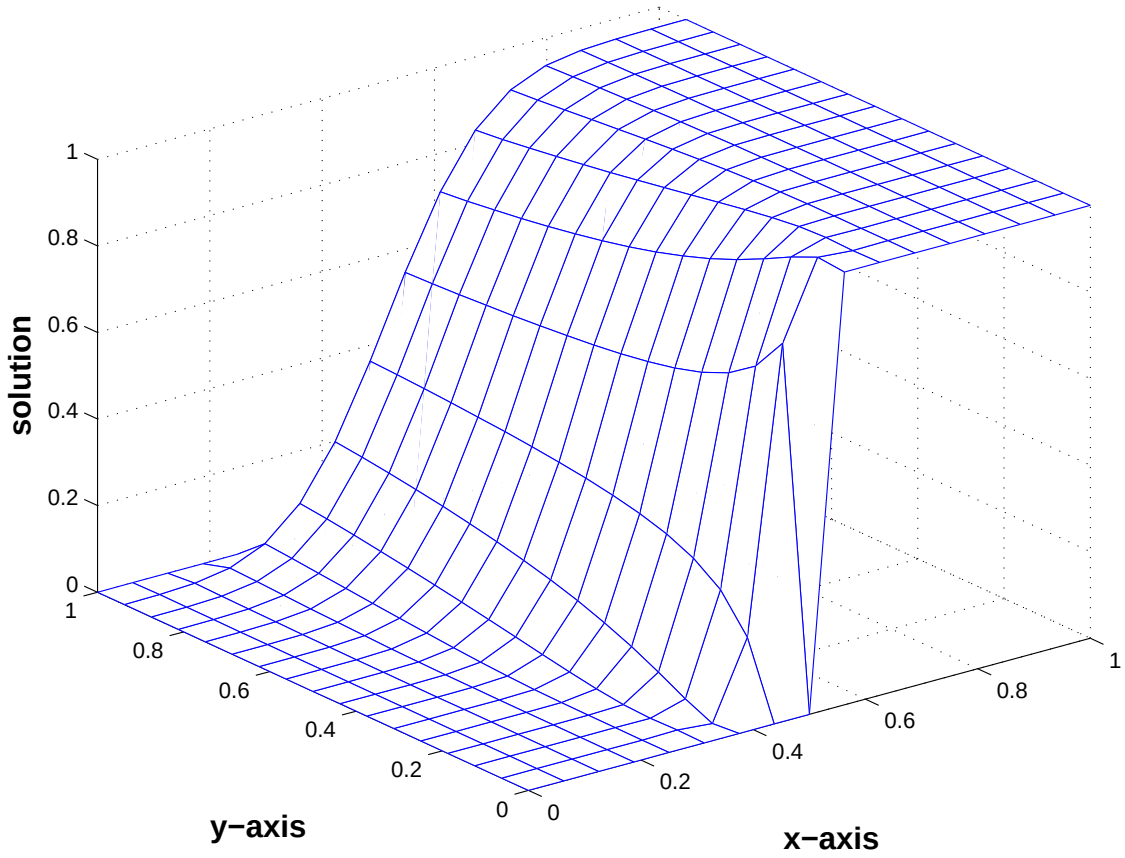
$$A_V = P\, A_H\, P \,,\; P = [I \otimes e_1, \ldots, I \otimes e_n] \,,\; P = P^T \,, \quad P^2 = I \,.$$

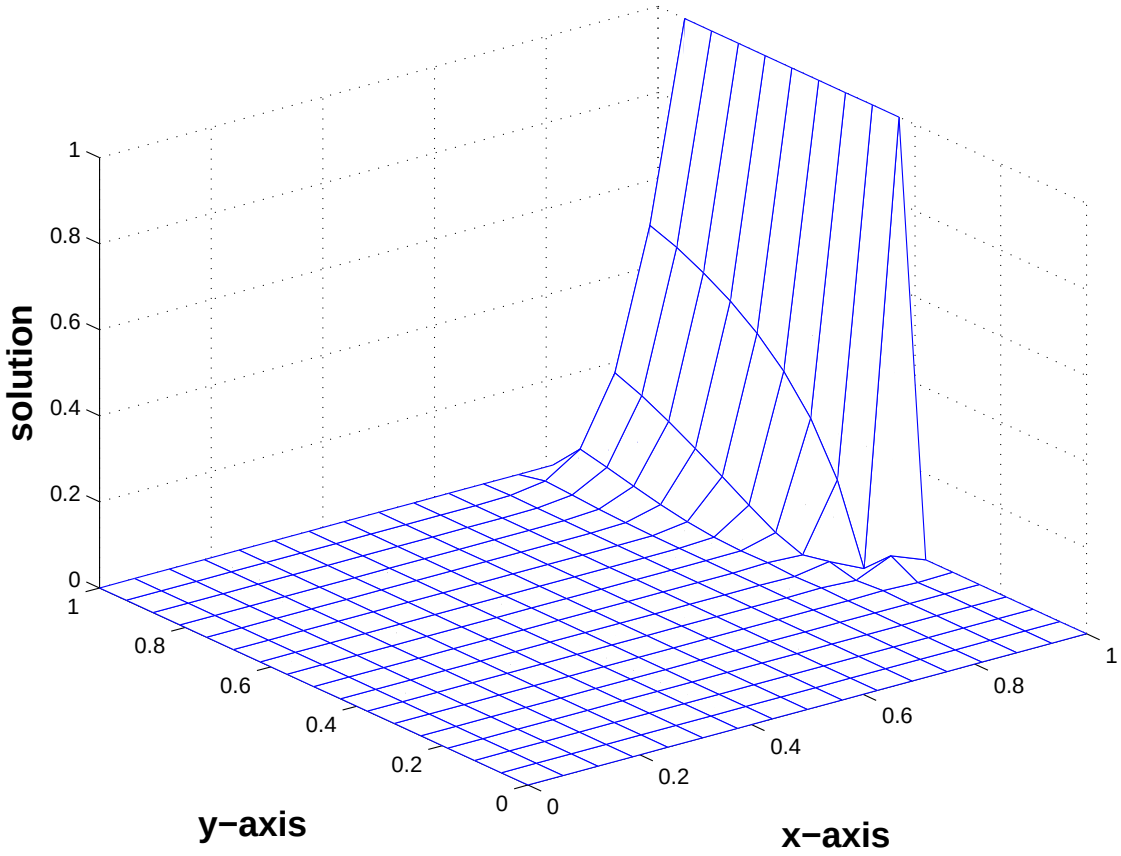
$\approx$ optimal stabilization parameter $\delta_* \equiv \frac{1}{2}\left(1 - \frac{1}{P_h}\right)$ affects

- smoothing of the discretized solution,
- behavior of the linear algebraic solver (convergence behavior of GMRES).

Examples of boundary conditions:

- Raithby (discontinuous inflow),
- partial right side of the domain.


solution
1
0.8
0.6
0.4
0.2
0
1
0.8
0.6
0.4
0.2
0
y−axis
0
0.2
0.4
0.6
0.8
1
x−axis


solution
1
0.8
0.6
0.4
0.2
0
1
0.8
0.6
0.4
0.2
0
y−axis
0
0.2
0.4
0.6
0.8
1
x−axis

Long list of authors, papers and books

Brooks, Hughes, Raithby, Roos, Stynes, Tobiska, Morton, Axelsson, . . .

GMRES convergence studied using the field of values and the eigendecompozition of the system matrix in particular by

[Eiermann - 89], [Ernst - 00], [Eiermann, Ernst - 02], [Fisher, Ramage, Silvester and Wathen - 99], [Elman, Ramage - 01, 02].

Different approach suggested in [Liesen, S - 04], [Liesen, S - 05].

Eigendecompozition of A_H (A_V) does not lead to useful bounds due to the ill-conditioned eigenvectors and cancelation. Instead: The matrices K and M are symmetric tridiagonal Toeplitz. The matrix of eigenvectors

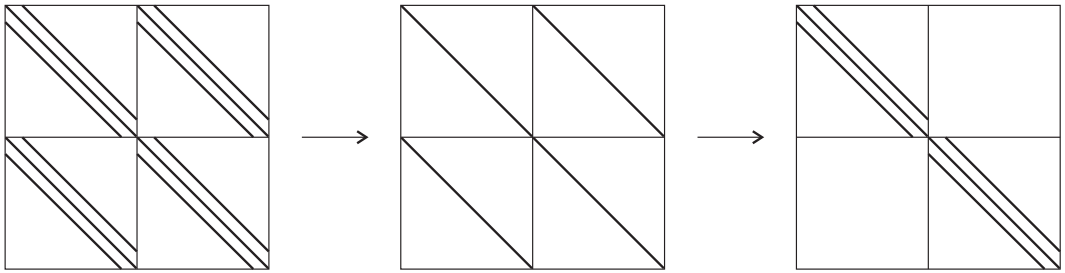
$$U = [u_1, \ldots, u_N],\ U = U^T,\ U^2 = I$$

$$u_j = (2h)^{1/2} [\sin(jh\pi), \ldots, \sin(Njh\pi)]^T,\ j = 1, \ldots, N,$$

represents the Fourier basis. Consider the Fourier transformation of unknowns in the direction perpendicular to the wind. Subsequent reordering of the transformed unknowns by vertical lines gives

$$P(I \otimes U)\, A_H\, (I \otimes U) P\, P(I \otimes U)\, x_H = P(I \otimes U)\, b_H\,.$$

The transformation above corresponds to the simultaneous diagonalization of the symmetric tridiagonal Toeplitz blocks in the block tridiagonal matrix A_H , with the subsequent permutation of the rows and columns.



The approach using $A_V x_V = b_V$ is even more straight, A_H was used here for historical reasons. Resulting system:

$$\mathbf{T\,y = f}$$

$$T = \text{diag}\,(T_j)\,,\ T_j = \text{tridiag}\,(\gamma_j, \lambda_j, \mu_j)\,,\ j = 1, \ldots, N\,.$$

Thus, the original discretized system transformes to N non-symmetric tridiagonal Toeplitz systems

$$\boxed{\mathbf{T_j}\,\mathbf{y_j} = \mathbf{f_j}\,, \quad j = 1, \ldots, N}$$

representing N discretized one-dimensional convection-diffusion problems (in the vertical direction of the original mesh, but accounting for the diffusion in the horizontal direction).

GMRES convergence behavior:

$$\|r_n\|^2 = \min_{p\in\Pi_n} \|\,p(A)\,b\,\|^2 = \min_{p\in\Pi_n} \|\,p(T)\,f\,\|^2 = \min_{p\in\Pi_n} \sum_{j=1}^{N} \|\,p(T_j)\,f_j\,\|^2 .$$

GMRES for non-symmetric tridiagonal Toeplitz systems? Interesting case: the superdiagonal (μ_j) substantially smaller in magnitude than the two others, $|\gamma_j| \approx \lambda_j \gg \mu_j$. Relating the problem for T_j, f_j to convergence of GMRES for scaled Jordan blocks, we proved (and quantified)

Theorem

Let l be the index of the first significant nonzero entry in f_j. Let $|\gamma_j| \approx \lambda_j \gg \mu_j$. Then GMRES for $T_j y_j = f_j$ must converge slowly for at least $N - l$ steps.

Slow initial convergence:

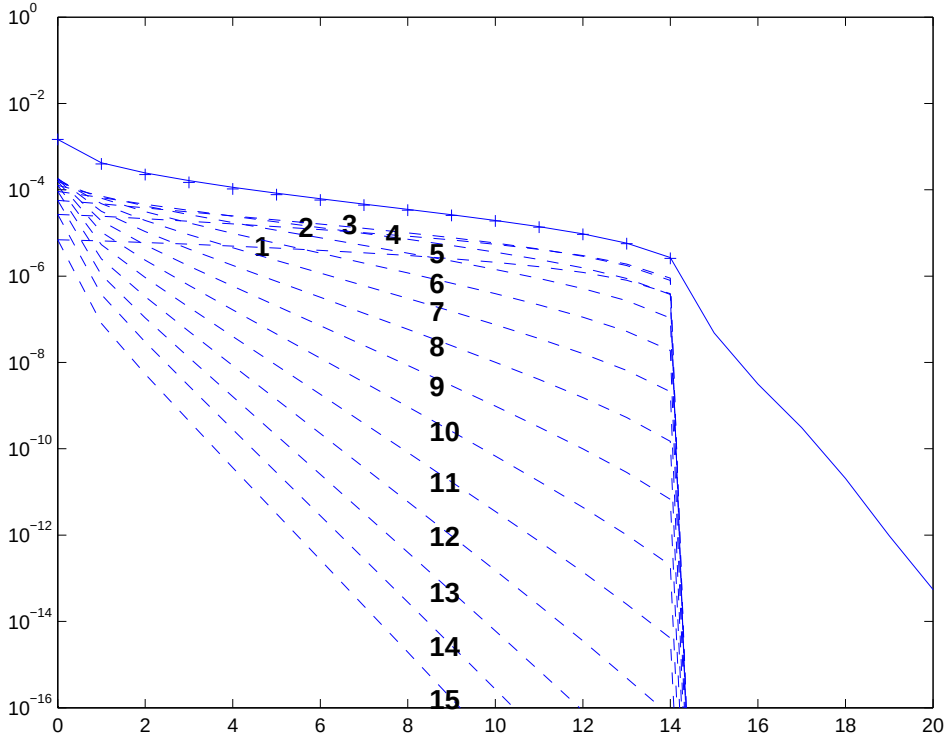
$$\|r_n\|^2 = \min_{p \in \Pi_n} \sum_{j=1}^{N} \| p(T_j) f_j \|^2 \geq \sum_{j=1}^{N} \min_{p \in \Pi_n} \| p(T_j) f_j \|^2 .$$

If the theorem applies at least for one j, then the convergence of GMRES for $A x = b$ must be slow for at least $N - l$ steps.

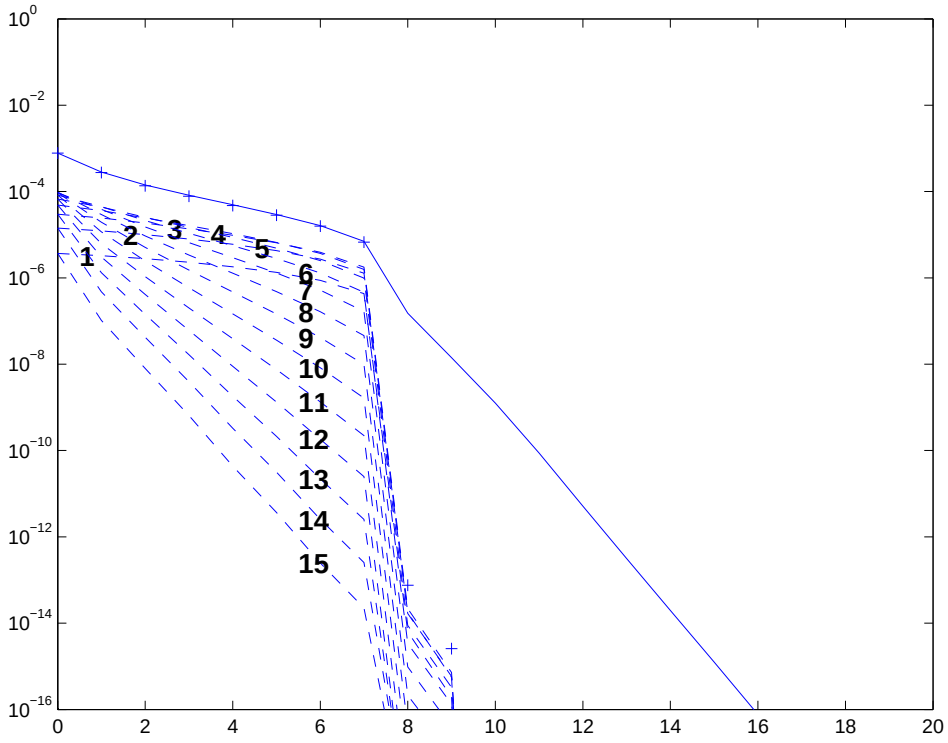
Acceleration of convergence:

The technique developed in [Greenbaum, Duintjer Tebbens and S - 05?] leads to tight upper and lower bounds which capture the sharp convergence acceleration.

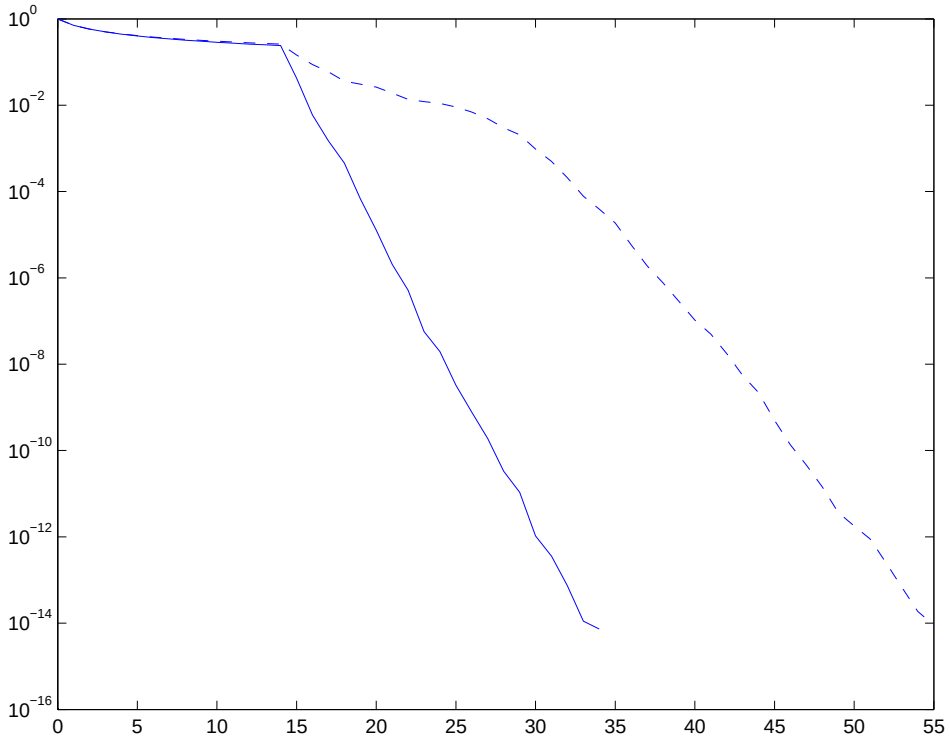
Nonzero boundary conditions on the full right side of the domain, GMRES convergence for the whole system (solid line) and for the individual tridiagonal Toeplitz blocks.



Nonzero boundary conditions on the part of the right side of the domain, GMRES convergence for the whole system (solid line) and for the individual tridiagonal Toeplitz blocks.



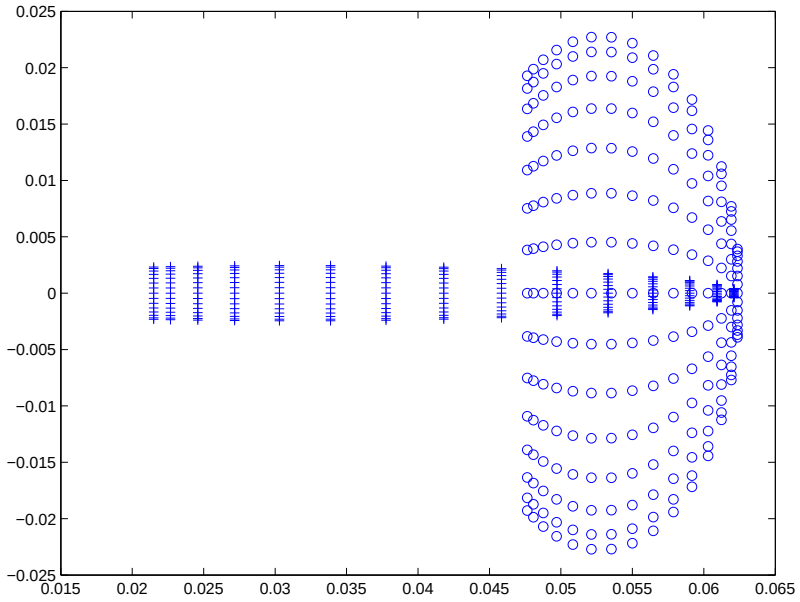
Discontinuous inflow boundary conditions (Raithby), two different values of the diffusion coefficient $\nu = 0.01$ and $\nu = 0.0001$ correspond to the solid and to the dashed line, respectively.



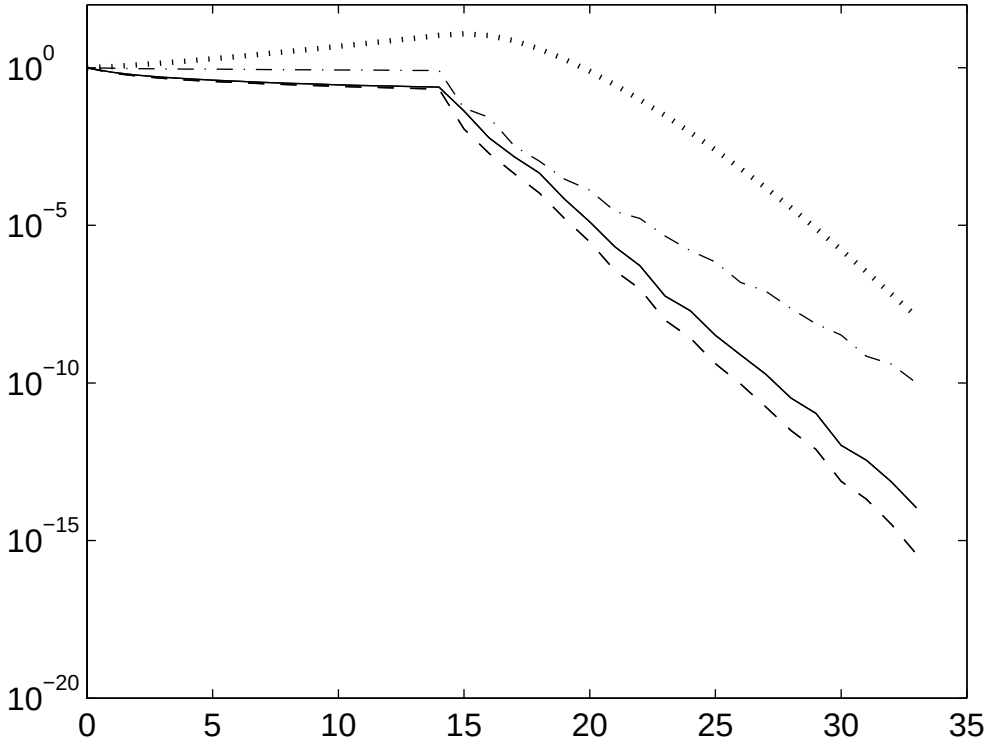
$$\sigma_{jk} \;=\; \lambda_j \,+\, (\gamma_j\,\mu_j)^{1/2}\,\omega_k\,,\; \omega_k = 2\cos\left(kh\pi\right),\; k = 1,\ldots,N\,.$$

Which spectrum corresponds to which convergence curve?

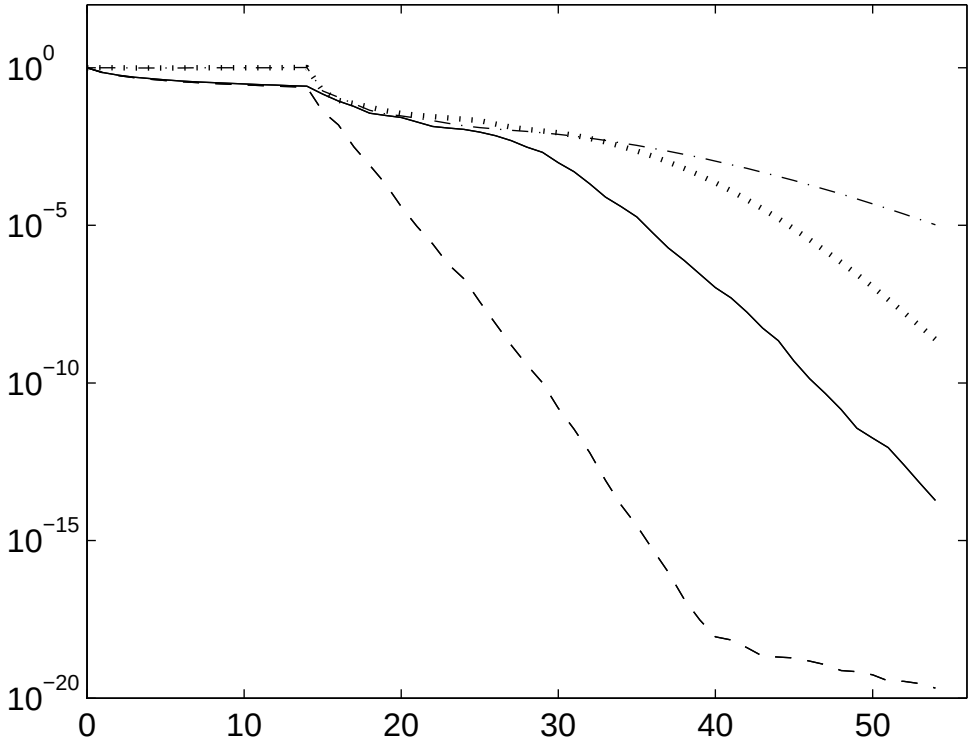
$$\lambda_j \;>\; 0\,, \quad \gamma_j\,\mu_j \;<\; 0\,.$$



GMRES Convergence curve, upper and lower bounds for $\nu = 0.01$.



GMRES Convergence curve, upper and lower bounds for $\nu = 0.0001$.



Concluding remarks

- initial phase is important, it depends on the right hand side!
- technique: orthonormal transformation to Jordan-like-structure (the problem is diagonalizable!)
- generalizations? Many ways . . . ?
- analytical study of preconditioning?