Spectral information
and Krylov subspace methods

Zdeněk Strakoš,

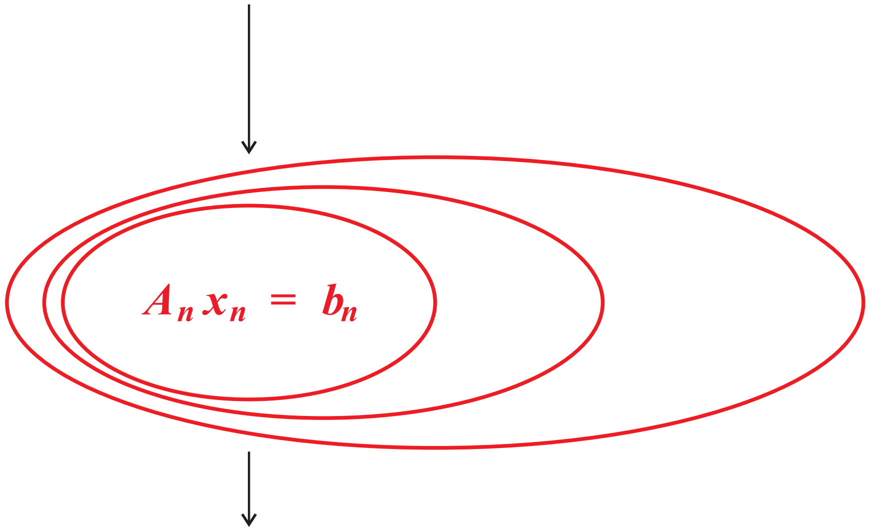
Academy of Sciences and Charles University, Prague

http://www.cs.cas.cz/˜strakos

Oberwolfach, August 2009.

Projections onto Krylov subspaces

$$A\,x = b, \quad A \in \mathbb{C}^{N \times N}, \quad r_0 = b - Ax_0$$



Here x_n approximates the solution x using the restriction and projection onto low dimensional subspaces

$$\mathcal{K}_n(A, r_0) \equiv span\,\{r_0, Ar_0, \cdots, A^{n-1}r_0\}$$

.

Nonlinearity and moments

The idea of restrictions and projections using Krylov subspaces is in a fundamental way linked with the problem of moments.

The story goes back to Gauss (1814), Jacobi (1826), Christoffel (1858, 1877), Chebyshev (1875), Stieltjes (1883-84) ...

And continues within the computational mathematics

by the works of Krylov (1931), Hestenes and Stiefel (1952), Gordon (1968), Golub and many co-workers (1968 -) ...

Outline

1. Krylov subspace methods as matching moments model reduction
2. Spectral information and GMRES

1 : Moment problem

Consider a non-decreasing distribution function $\omega(\lambda),\ \lambda \geq 0$ with the moments given by the Riemann-Stieltjes integral

$$\xi_k = \int_0^\infty \lambda^k \, d\omega(\lambda) \,, \quad k = 0, 1, \ldots \,.$$

Find the distribution function $\omega^{(n)}(\lambda)$ with n points of increase $\lambda_i^{(n)}$ which matches the first $2n$ moments for the distribution function $\omega(\lambda)$,

$$\int_0^\infty \lambda^k \, d\omega^{(n)}(\lambda) \equiv \sum_{i=1}^{n} \omega_i^{(n)} (\lambda_i^{(n)})^k = \xi_k, \quad k = 0, 1, \ldots, 2n-1 \,.$$

1 : Model reduction via matching moments I

Gauss-Christoffel quadrature formulation:

$$\int_0^\infty f(\lambda)\, d\omega(\lambda) \;\approx\; \sum_{i=1}^{n} \omega_i^{(n)} f(\lambda_i^{(n)})\,,$$

where the reduced model given by the distribution function with n points of increase $\omega^{(n)}$ matches the first $2n$ moments

$$\int_0^\infty \lambda^k\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} (\lambda_i^{(n)})^k\,, \quad k = 0, 1, \ldots, 2n-1\,.$$

1 : Stieltjes recurrence

Let $p_1(\lambda) \equiv 1, p_2(\lambda), \ldots, p_{n+1}(\lambda)$ be the first $n+1$ orthonormal polynomials corresponding to the distribution function $\omega(\lambda)$.

Then, writing $P_n(\lambda) = (p_1(\lambda), \ldots, p_n(\lambda))^T$,

$$\lambda \, P_n(\lambda) \; = \; T_n \, P_n(\lambda) \, + \, \delta_{n+1} \, p_{n+1}(\lambda) \, e_n$$

represents the Stieltjes recurrence (1883-4), with the Jacobi matrix

$$T_n \equiv \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_n \\ & & \delta_n & \gamma_n \end{pmatrix}, \qquad \delta_l > 0 \, .$$

1 : Stieltjes recurrence - earlier related works

Early continued fractions-related recurrences:

Brouncker (1655), Wallis (1656), Euler (1748).

1 : Matrix computation: Lanczos $\equiv$ Stieltjes

In matrix computations, T_n results from the Lanczos process (1951) applied to T_n starting with e_1. Therefore $p_1(\lambda) \equiv 1,\, p_2(\lambda),\, \ldots,\, p_n(\lambda)$ are orthonormal with respect to the inner product

$$(p_s, p_t) \;\equiv\; \sum_{i=1}^{n} |(z_i^{(n)}, e_1)|^2 \, p_s(\theta_i^{(n)}) \, p_t(\theta_i^{(n)}) \,,$$

where $z_i^{(n)}$ is the orthonormal eigenvector of T_n corresponding to the eigenvalue $\theta_i^{(n)}$, and $p_{n+1}(\lambda)$ has the roots $\theta_i^{(n)},\, i = 1, \ldots, n$.

Consequently,

$$\omega_i^{(n)} \;=\; |(z_i^{(n)}, e_1)|^2 \,, \quad \lambda_i^{(n)} \;=\; \theta_i^{(n)} \,,$$

1 : Conjugate gradients (CG) for $Ax = b$

Given A HPD, b, x_0, r_0 ,

$$\|x - x_n\|_A = \min_{u \,\in\, x_0 + \mathcal{K}_n(A, r_0)} \|x - u\|_A$$

with the formulation via the Lanczos process, $w_1 = r_0 / \|r_0\|$,

$$A\,W_n = W_n\,T_n + \delta_{n+1} w_{n+1} e_n^T\,, \quad T_n = W_n^*(A)\,A\,W_n(A)\,,$$

and the CG approximation given by

$$T_n\,y_n = \|r_0\|\,e_1\,, \quad x_n = x_0 + W_n\,y_n\,.$$

1 : CG $\equiv$ matrix formulation of the Gauss Q

$$Ax = b\,,\; x_0 \quad \longleftrightarrow \quad \int_{\zeta}^{\xi} \lambda^{-1}\, d\omega(\lambda)$$

$$\uparrow \qquad\qquad\qquad\qquad \uparrow$$

$$T_n\, y_n \;=\; \|r_0\|\, e_1 \quad \longleftrightarrow \quad \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1}$$

$$x_n \;=\; x_0 + W_n\, y_n$$

$$\boxed{\omega^{(n)}(\lambda) \longrightarrow \omega(\lambda)}$$

1 : Model reduction via matching moments II

CG (Lanczos) reduces for A HPD at the step n the original model

$$A\,x \;=\; b\,,\; r_0 = b - Ax_0$$

to

$$T_n\,y_n \;=\; \|r_0\|\,e_1\,,$$

such that the the $2n$ moments are matched,

$$w_1^*\,A^k\,w_1 \;=\; e_1^T\,T_n^k\,e_1\,, \quad k = 0, 1, \ldots, 2n-1\,.$$

Closely related to the works of
Gauss, Jacobi, Chebyshev, Stieltjes, Markov, ...

1 : Comments on literature

Proofs of results related to moments or model reduction are in the literature typically based on factorizations of the matrix of moments, Golub and Welsh (1969), Dahlquist, Golub and Nash (1978), ..., Kent(1989), ..., which is also true for Antoulas (2005).

It is interesting to recall the paper of Chebyshev from (1875), which uses a moment matching expansion of the continued fraction and which may be considered a starting point of the general theory of orthogonal polynomials, see Gragg (1974), Kuijlaars and Rakhmanov (1999).

$$w^*(\lambda I - A)^{-1} w \;=\; \sum_{j}^{N} \frac{\omega_j}{\lambda - \lambda_j} \;=\; \sum_{\ell}^{2n} \frac{\xi^{\ell-1}}{\lambda^{\ell}} \;+\; \mathcal{O}(\lambda^{-(2n+1)})$$

1 : Comments on literature

Moment matching techniques have been used for decades in computational physics and in computational chemistry, see the remarkable papers by Gordon (1968), Reinhard (1979) ...

Gauss quadrature formulation related to the nonsymmetric Lanczos process and to the Arnoldi process was given by Freund and Hochbruck (1993), motivated by Fischer and Freund (1992). Gauss quadrature was formally extended to the complex plane by Saylor and Smolarski (2001), with motivation from inverse scattering problems in electromagnetics by Warnick (1997), . . . , Golub, Stoll and Wathen (2008), S and Tichy (2009).

Matrix of moments as well as any formal generalization of the Gauss quadrature formulas to the complex plane can be avoided by using the Vorobyev method of moments.

1 : Vorobyev moment problem - 1958, 1965

The restricted og projected op. $A_n = W_n W_n^* \, A \, W_n W_n^* = W_n \, T_n \, W_n^*$

satisfies by definition

$$
\begin{aligned}
A_n \, w_1 &= A \, w_1 \,, \\
A_n \, (A \, w_1) \equiv A_n^2 \, w_1 &= A^2 w_1 \,, \\
&\vdots \\
A_n \, (A^{n-2} w_1) \equiv A_n^{n-1} \, w_1 &= A^{n-1} w_1 \,, \\
A_n \, (A^{n-1} w_1) \equiv A_n^{n} \, w_1 &= W_n W_n^* \, (A^n w_1) \,,
\end{aligned}
$$

which gives the result for CG, Lanczos

$$
w_1^* \, A^k \, w_1 = w_1^* \, A_n^k \, w_1 = e_1^T \, T_n^k \, e_1 \,, \quad k = 0, 1, \ldots, 2n-1 \,.
$$

1 : Non-Hermitian Lanczos (BiCG), [S-2009]

Given a nonsingular $A \in \mathbb{C}^{N\times N}$, $v_1 \in \mathbb{C}^N$, $w_1 \in \mathbb{C}^N$, $v_1^* w_1 = 1$,

$$
\begin{aligned}
A\,W_n &= W_n\,T_n + \delta_{n+1}\,w_{n+1}\,e_n^T\,, \\
A^*\,V_n &= V_n\,T_n^* + \beta_{n+1}^*\,v_{n+1}\,e_n^T\,, \\
V_n^*\,W_n &= I_n\,, \quad T_n = V_n^*(A, v_1, w_1)\,A\,W_n(A, v_1, w_1)\,.
\end{aligned}
$$

We assume that the algorithm does not break down in steps 1 through n (it can break down later). Then, using the Vorobyev method of moments,

$$v_1^*\,A^k\,w_1 \equiv e_1^T\,T_n^k\,e_1\,, \quad k = 0, 1, \ldots, 2n-1\,,$$

i.e., n steps of the non-Hermitian Lanczos (or BiCG) represent a model reduction which matches the first $2n$ moments.

1 : Arnoldi (FOM (GMRES)) [S -2009]

The construction looks similar to the Hermitian case.
Due to A non-Hermitian the last line of the Vorobyev moment problem can not be used as above,

$$v_1^T A^k w_1 \equiv v_1^T A_n^k w_1 , \quad k = 0, 1, \ldots, n-1 ,$$

and we get

$$v_1^T A^k w_1 = (W^T v_1)^T H_n^k e_1 , \quad k = 0, 1, \ldots, n-1 ,$$

$$w_1^T A^k w_1 = w_1^T A_n^k w_1 = e_1^T H_n^k e_1 , \quad k = 0, 1, \ldots, n ,$$

i.e., n steps of the Arnoldi (or FOM) represent
a model reduction which matches the first n or $(n+1)$ moments.

1 : Question:

When does the spectral information determine the solution of the problem of moments?

Outline

1. Krylov subspace methods as matching moments model reduction

2. Spectral information and GMRES

From Hermitian through normal to non-normal

Normal matrices have a full set of eigenvectors forming a basis of $\mathbb{C}^N$ which can be chosen orthonormal. Therefore the change to (orthonormal) eigenvector coordinates does not involve any distortion of geometry.

Substantial difference from the Hermitian case which causes enormous technical difficulties in proofs and in deriving bounds - the eigenvalues are not real. However, principal difficulties come with non-normality.

We restrict ourselves to the GMRES method.

Minimal residual methods

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

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

(Hermitian) MINRES [Paige, Saunders - 75]
and 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.

Bound by Elman for A normal

$$
\begin{aligned}
\|r_n\| &= \|p_n(A) r_0\| = \min_{p \in \Pi_n} \|p(A) r_0\| = \min_{p \in \Pi_n} \| S \, [p(\Lambda) \, S^* r_0] \| \\
&= \min_{p \in \Pi_n} \|p(\Lambda) \, S^* r_0\| = \min_{p \in \Pi_n} \left\{ \sum_i | \, (s_i^* r_0) \, p(\lambda_i) \, |^2 \right\}^{\frac{1}{2}} \\
&\leq \|r_0\| \min_{p \in \Pi_n} \max_i |p(\lambda_i)| \, .
\end{aligned}
$$

$|p_n(\lambda_i)|$ represents a multiplicative correction to the absolute values $|s_i^* r_0|$ 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 for a general A

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

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

reflects that, while the term in the bound

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

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

Problem :

The rate of convergence is often automatically linked to the distribution of eigenvalues of the matrix A.

There are, however, examples showing that any (nonincreasing) convergence curve is possible for GMRES with matrix A having any given (nonzero) eigenvalues. [Greenbaum, S - 94], [Greenbaum, Pták, S - 96] , [Arioli , Pták, S - 98]

Assume convergence exactly in N steps (generalization to $m < N$ possible). For simplicity of notation $r_0 = b$ $(x_0 = 0)$.

Problem:

Given convergence curve, given N nonzero eigenvalues (not necessarily distinct), describe the set of all $\{A, b\}$ such that GMRES (A, b) generates the curve while the spectrum of A is prescribed.

Convergence curve

$$\|r_0\| \geq \|r_1\| \geq \cdots \geq \|r_{N-1}\| > \|r_N\| = 0,$$

$$h \equiv (\eta_1, \ldots, \eta_N)^T, \quad \eta_j \equiv ((\|r_{j-1}\|)^2 - \|r_j\|^2)^{1/2}.$$

$$d \equiv (\nu_1, \ldots, \nu_N), \quad \nu_1 = \frac{1}{\eta_N},\ \nu_2 = -\frac{\eta_1}{\eta_N},\ \ldots,\ \nu_N = -\frac{\eta_{N-1}}{\eta_N}\,.$$

Meaning? Let $W = (w_1, \ldots, w_j)$ be the orthonormal basis of $AK_j(A, r_0)$. Then

$$r_n = r_0 - \sum_{j=1}^{n} w_j \eta_j, \quad r_0 = \sum_{j=1}^{n} w_j \eta_j + r_n, \quad \|r_0\|^2 = \sum_{j=1}^{n} \eta_j^2 + \|r_n\|^2$$

Eigenvalues

$$\{\lambda_1, \lambda_2, \ldots, \lambda_N\}, \qquad \lambda_j \neq 0, \quad j = 1, \ldots, n\,.$$

$$q_N(z) \;\equiv\; z^N - \sum_{j=0}^{N-1} \alpha_j z^j = (z-\lambda_1)(z-\lambda_2)\ldots(z-\lambda_N),$$

$$p_N(z) \;\equiv\; 1 - \sum_{j=1}^{N} \xi_j z^j = -\,\frac{1}{\alpha_0}\, q_N(z), \quad \xi_N = \frac{1}{\alpha_0}, \qquad \xi_j = -\,\frac{\alpha_j}{\alpha_0}\,,$$

$$s \;\equiv\; (\xi_1, \ldots, \xi_N)^T\,, \quad a = (\alpha_0, \ldots, \alpha_{N-1})^T$$

Spectral companion matrix

$$C = \begin{pmatrix} 0 & & & \alpha_0 \\ 1 & \ddots & & \alpha_1 \\ & \ddots & 0 & \vdots \\ & & 1 & \alpha_{N-1} \end{pmatrix} = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \boxed{a} \\ & \ddots & 0 & \\ & & 1 & \end{pmatrix}$$

$$C^{-1} = \begin{pmatrix} & 1 & & \\ \boxed{s} & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

Problem solution

Theorem 2

The following two assertions are equivalent:

1° The spectrum of A is $\{\lambda_1, \dots, \lambda_N\}$ and GMRES(A, b) yields residuals with the prescribed nonincreasing sequence

$$\|r_0\| \geq \|r_1\| \geq \cdots \geq \|r_{N-1}\| > \|r_N\| = 0\,.$$

2° Matrix A is of the form $A = W R C R^{-1} W^*$ and $b = W h$ where C is the companion matrix corresponding to the polynomial $q_N(z)$, W is unitary and R a nonsingular upper triangular matrix such that $Rs = h$.

Corollary: Any noninreasing convergence curve can be generated by GMRES for a matrix having any prescribed eigenvalues.

Problem solution

Denoting

$$Y \equiv RC^{-1} = R \begin{pmatrix} \boxed{s} & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} = \begin{pmatrix} \boxed{h} & \boxed{R_{N-1}} \\ & 0 \end{pmatrix},$$

Then

$$\begin{aligned} A &= WRCR^{-1}W^* = W(RC^{-1})C(CR^{-1})W^* = \\ &= W(RC^{-1})C(RC^{-1})^{-1}W^* = WYCY^{-1}W^* \end{aligned}$$

Assertions 1° and 2° of Theorem 2 are equivalent to

Problem solution

Theorem 2 (continuation)

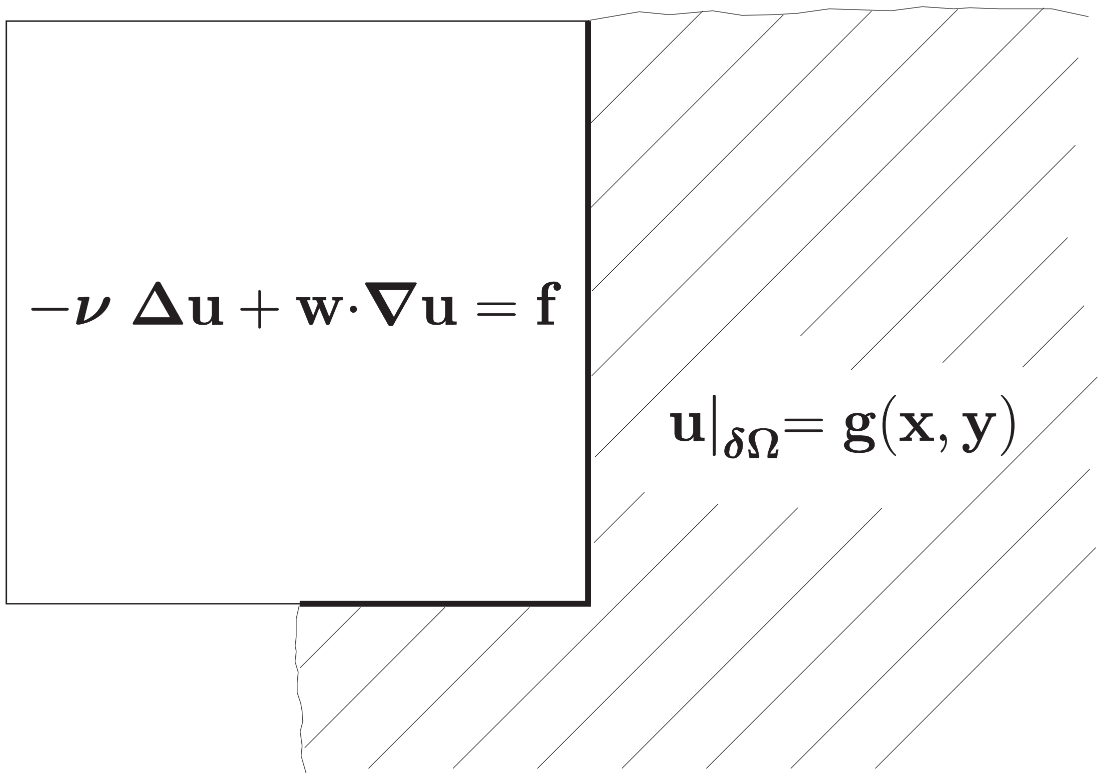
3° Matrix A is of the form

$$A = WYCY^{-1}W^*$$

and $b = Wh$ where C is the companion matrix corresponding to the polynomial $q(\lambda)$, W is unitary and R_{N-1} part of Y is **any** $(N-1)$ by $(N-1)$ nonsingular upper triangular matrix.

Convection-diffusion model problem

Convection-diffusion model problem

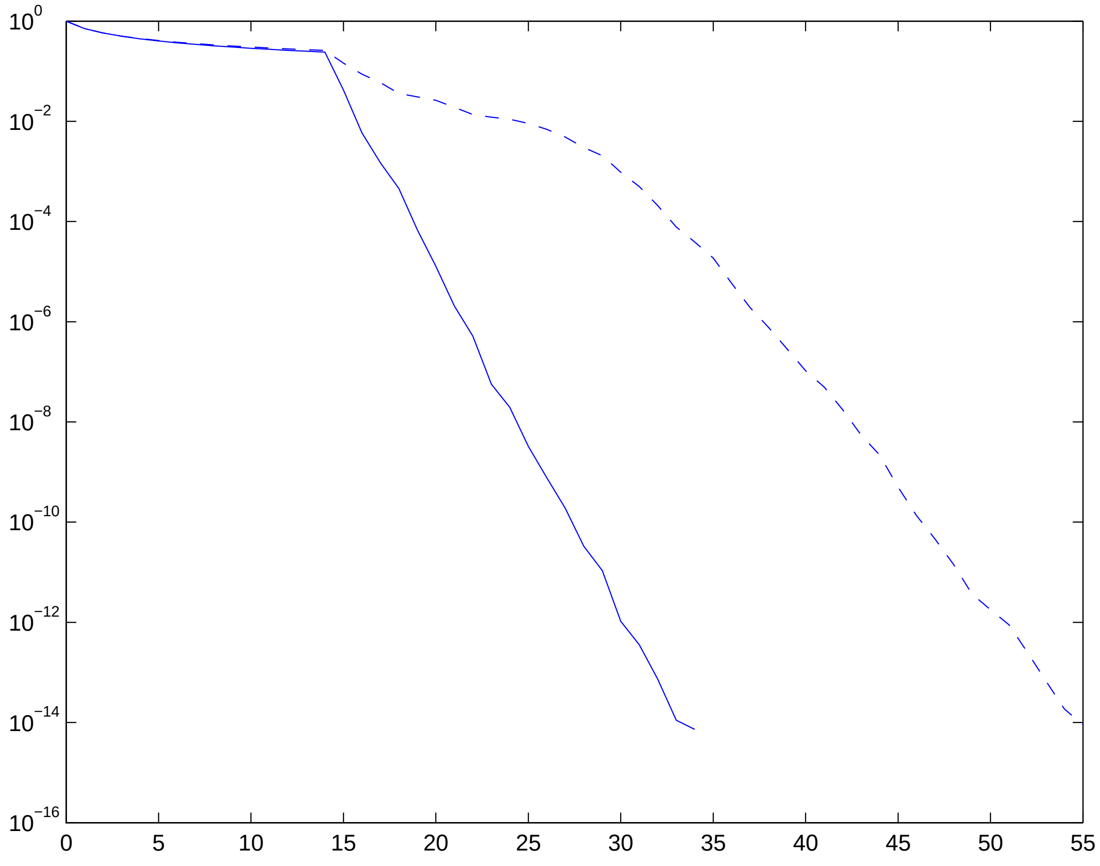


$\uparrow \quad \mathbf{w} = [\mathbf{0}, \mathbf{1}]^{\mathbf{T}}$

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

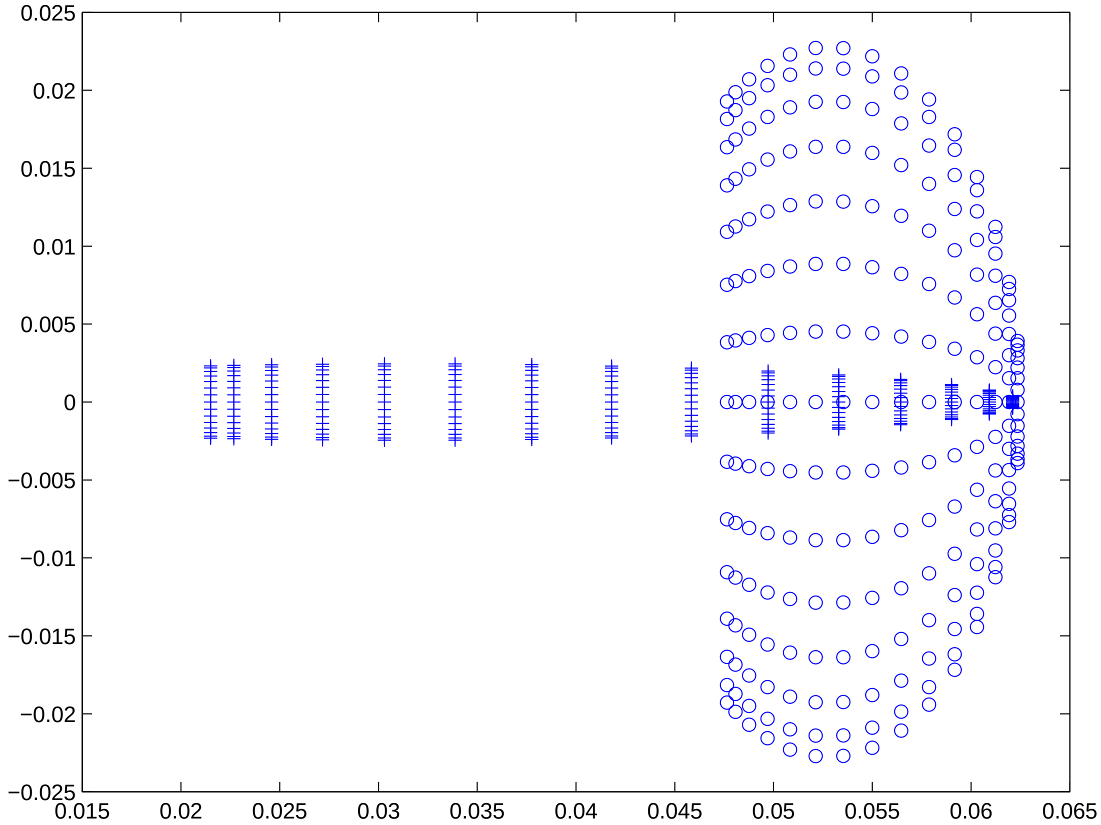
$\nu = 0.01$ and $\nu = 0.0001$

[Liesen, S - 2004, 2005]



$\nu = 0.01$ and $\nu = 0.0001$

Which spectrum corresponds to which convergence curve?



Concluding remarks

- initial phase is important, it depends on the right hand side!
- technique: orthonormal transformation to Jordan-like-structure (for the convection-diffusion model problem
  the matrix is diagonalizable!
- Given parametrization - a tool ?
- Eigenvalue solvers ?