Iterative solvers for saddle point algebraic linear systems: tools of the trade

V. Simoncini

Dipartimento di Matematica, Università di Bologna
and CIRSA, Ravenna, Italy
valeria@dm.unibo.it

The problem

$$\begin{bmatrix} A & B^T \\ B & -C \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \qquad A = A^T, \; C = C^T \geq 0$$

$$\mathcal{M}x = b$$

with $\mathcal{M}$ large, real indefinite symmetric matrix

- Computational Fluid Dynamics (Elman, Silvester, Wathen 2005)
- Elasticity problems
- Mixed (FE) formulations of II and IV order elliptic PDEs
- Linearly Constrained Programs
- Linear Regression in Statistics
- Image restoration and registration
- ... Survey: Benzi, Golub and Liesen, Acta Num 2005

Iterative solver. Convergence considerations.

$$\mathcal{M}x = b$$

$\mathcal{M}$ is symmetric and indefinite $\quad\rightarrow\quad$ MINRES

(Paige & Saunders, '75)

$$x_k \in x_0 + K_k(\mathcal{M}, r_0), \quad \text{s.t.} \qquad \min \|b - \mathcal{M}x_k\|$$

$r_k = b - \mathcal{M}x_k$, $k = 0, 1, \ldots$, x_0 starting guess

Iterative solver. Convergence considerations.

$$\mathcal{M}x = b$$

$\mathcal{M}$ is symmetric and indefinite $\quad\to\quad$ MINRES

(Paige & Saunders, '75)

$$x_k \in x_0 + K_k(\mathcal{M}, r_0), \quad \text{s.t.} \qquad \min \|b - \mathcal{M}x_k\|$$

$r_k = b - \mathcal{M}x_k$, $k = 0, 1, \ldots$, x_0 starting guess

If $\mathrm{spec}(\mathcal{M}) \subset [-a, -b] \cup [c, d]$, with $|b - a| = |d - c|$, then

$$\|b - \mathcal{M}x_{2k}\| \le 2 \left(\frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \right)^k \|b - \mathcal{M}x_0\|$$

Note: more general but less tractable bounds available

Features of MINRES

- Residual minimizing solver for *indefinite* linear systems
- Short-term recurrence (possibly with Lanczos recurrence)
- Positive definite preconditioner, but not necessarily factorized (not as LL^T)

Features of MINRES

- Residual minimizing solver for *indefinite* linear systems
- Short-term recurrence (possibly with Lanczos recurrence)
- Positive definite preconditioner, but not necessarily factorized (not as LL^T)

Outline

- Harmonic Ritz values and superlinear convergence
- Enhancing MINRES convergence
- Estimating the Saddle-point problem inf-sup constant and stopping criteria

Harmonic Ritz values

$$K_k(\mathcal{M}, r_0) = \mathrm{span}\{r_0, \mathcal{M}r_0, \ldots, \mathcal{M}^{k-1}r_0\} \qquad (x_0 = 0)$$

$$x_k = \phi_{k-1}(\mathcal{M})r_0 \in K_k(\mathcal{M}, r_0), \quad \phi_{k-1} \text{ polyn. of deg} \leq k-1$$

Therefore

$$r_k = r_0 - \mathcal{M}x_k = \varphi_k(\mathcal{M})r_0, \quad \varphi_k \text{ polyn. of deg} \leq k, \varphi_k(0) = 1$$

Harmonic Ritz values

$K_k(\mathcal{M}, r_0) = \text{span}\{r_0, \mathcal{M}r_0, \ldots, \mathcal{M}^{k-1}r_0\}$ $\qquad (x_0 = 0)$

$x_k = \phi_{k-1}(\mathcal{M})r_0 \in K_k(\mathcal{M}, r_0)$, $\quad \phi_{k-1}$ polyn. of deg $\leq k-1$

Therefore

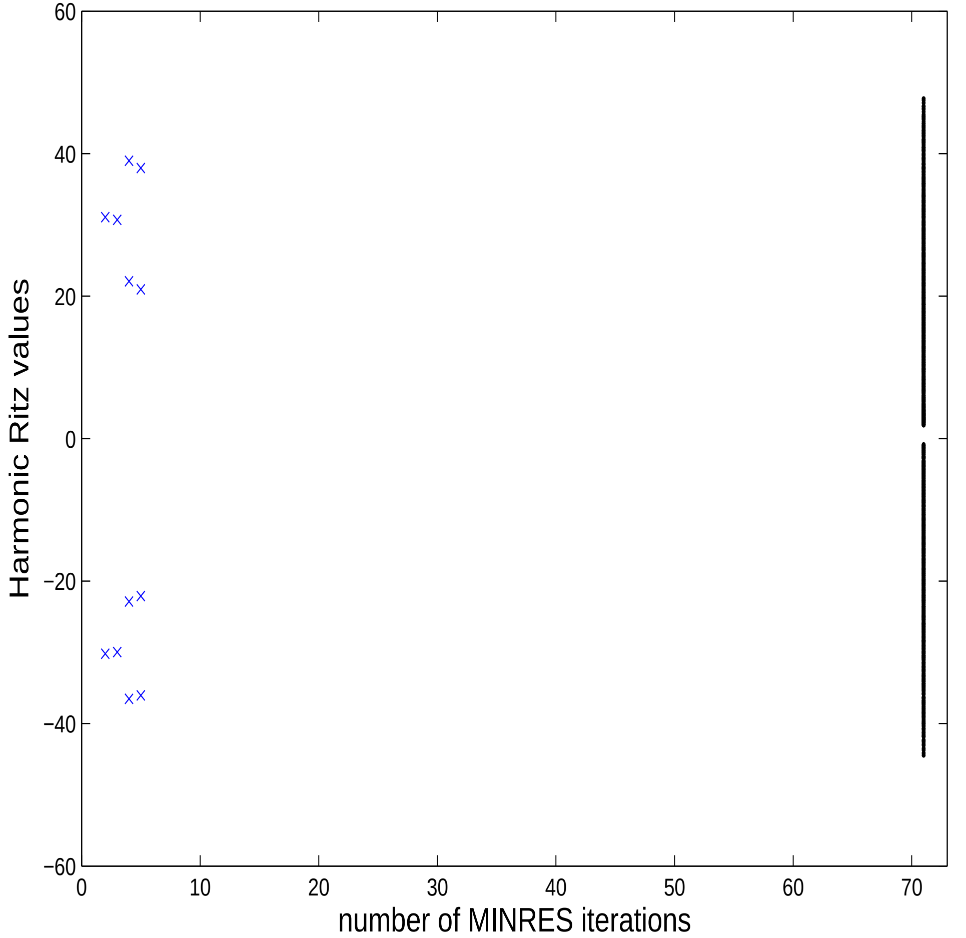
$r_k = r_0 - \mathcal{M}x_k = \varphi_k(\mathcal{M})r_0$, $\quad \varphi_k$ polyn. of deg $\leq k$,$\varphi_k(0) = 1$

Harmonic Ritz values: roots of φ_k (residual polynomial)

Remark: Harmonic Ritz values approximate eigenvalues of $\mathcal{M}$

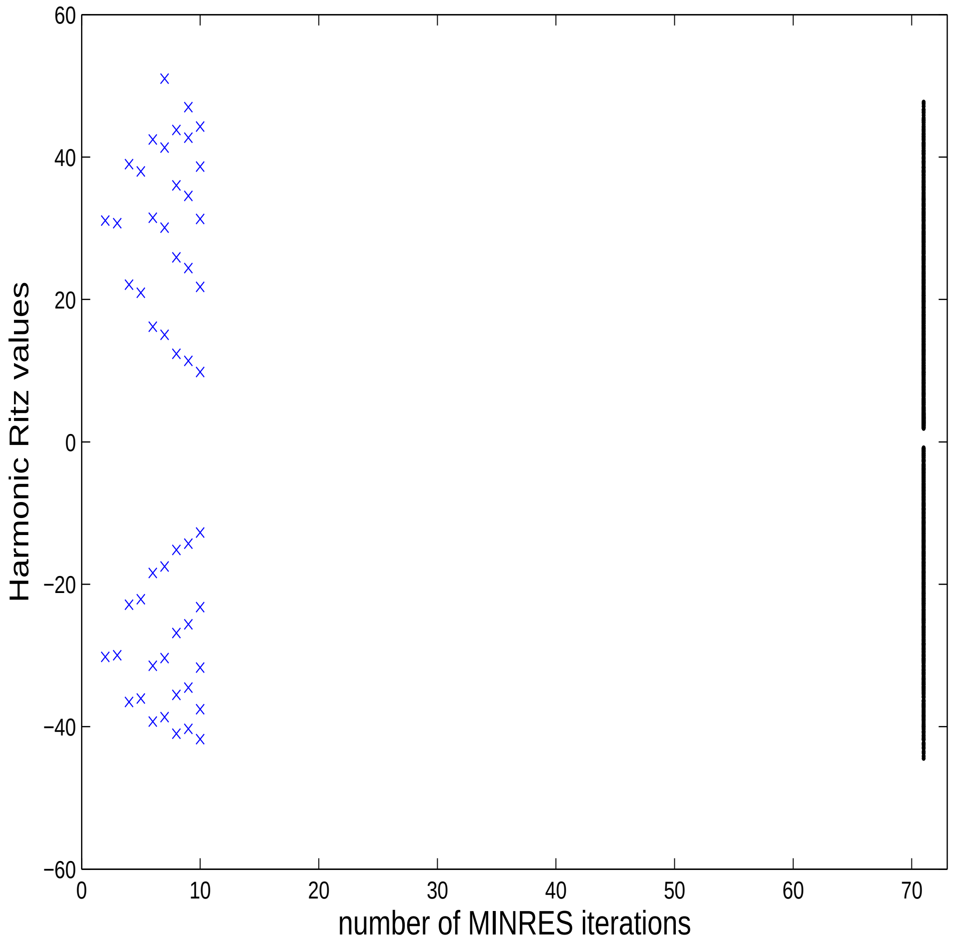
(Paige, Parlett & van der Vorst, '95)

Typical convergence pattern



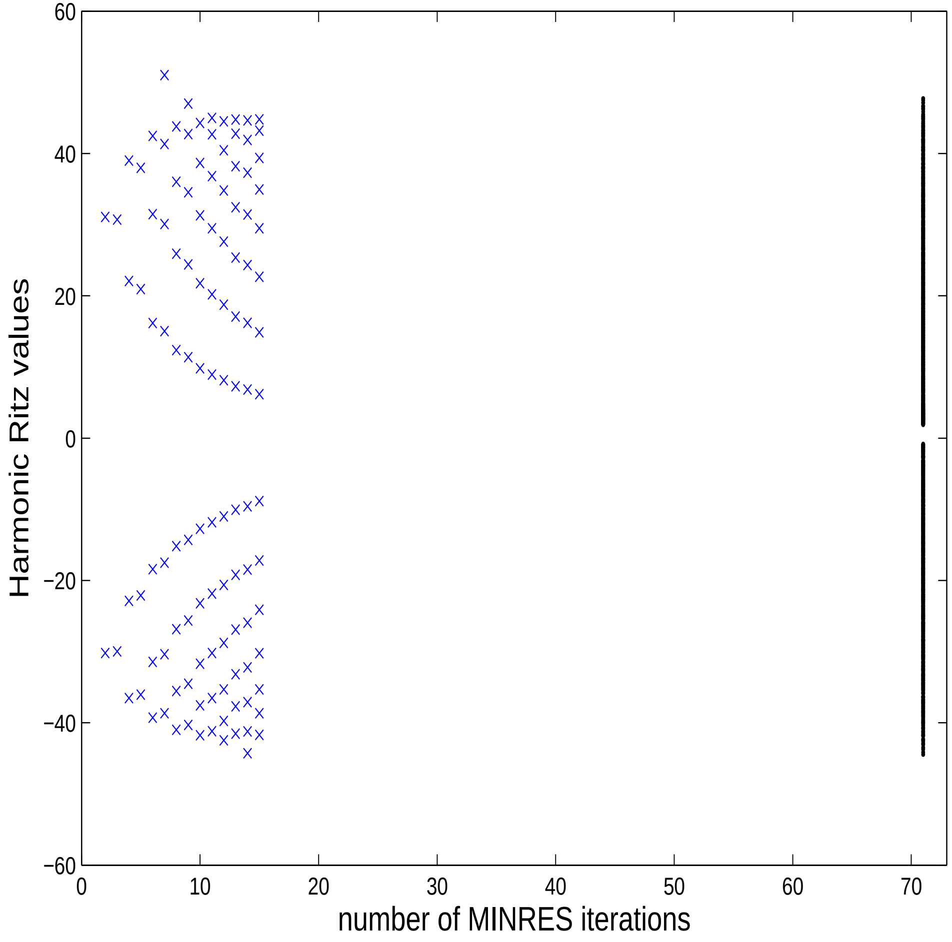
Harmonic Ritz values as iterations proceed

Typical convergence pattern



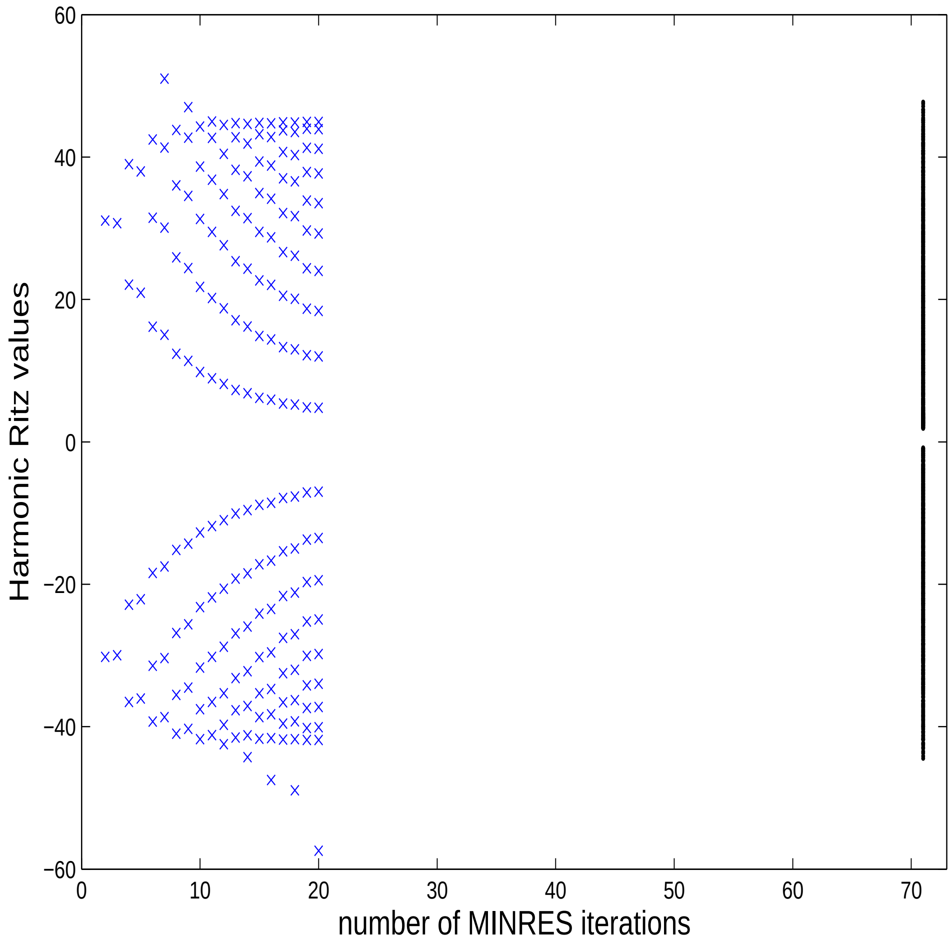
Harmonic Ritz values as iterations proceed

Typical convergence pattern



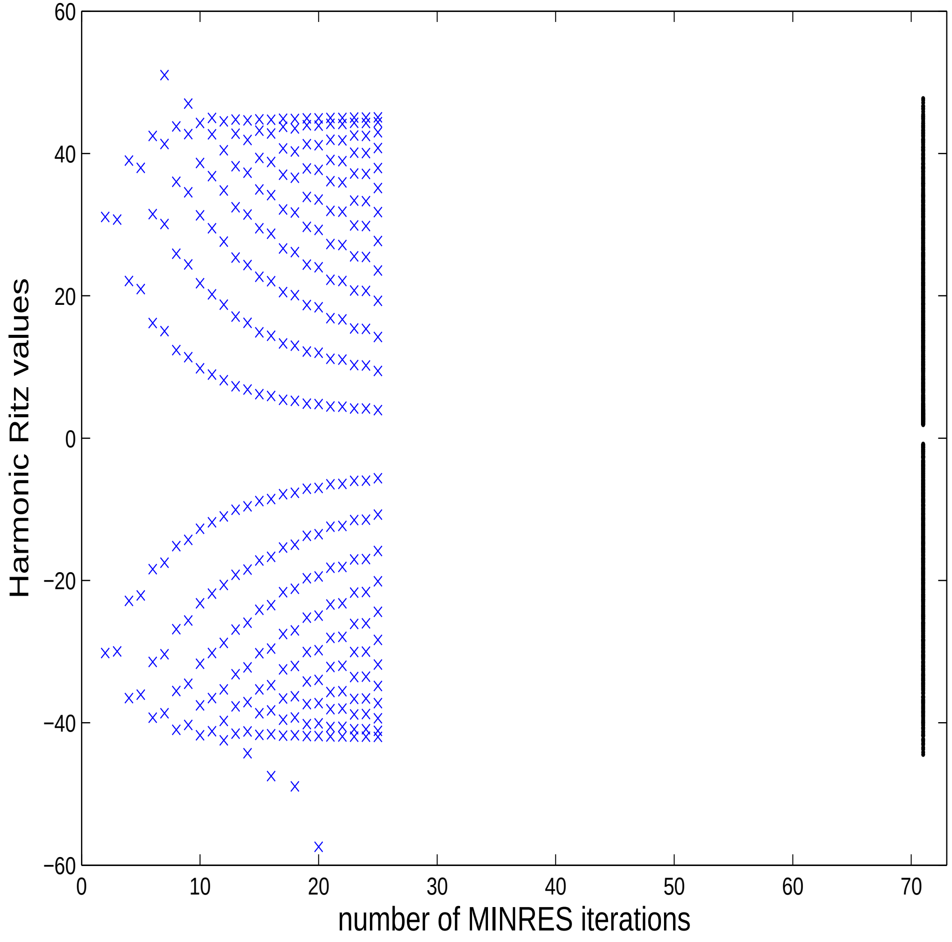
Harmonic Ritz values as iterations proceed

Typical convergence pattern



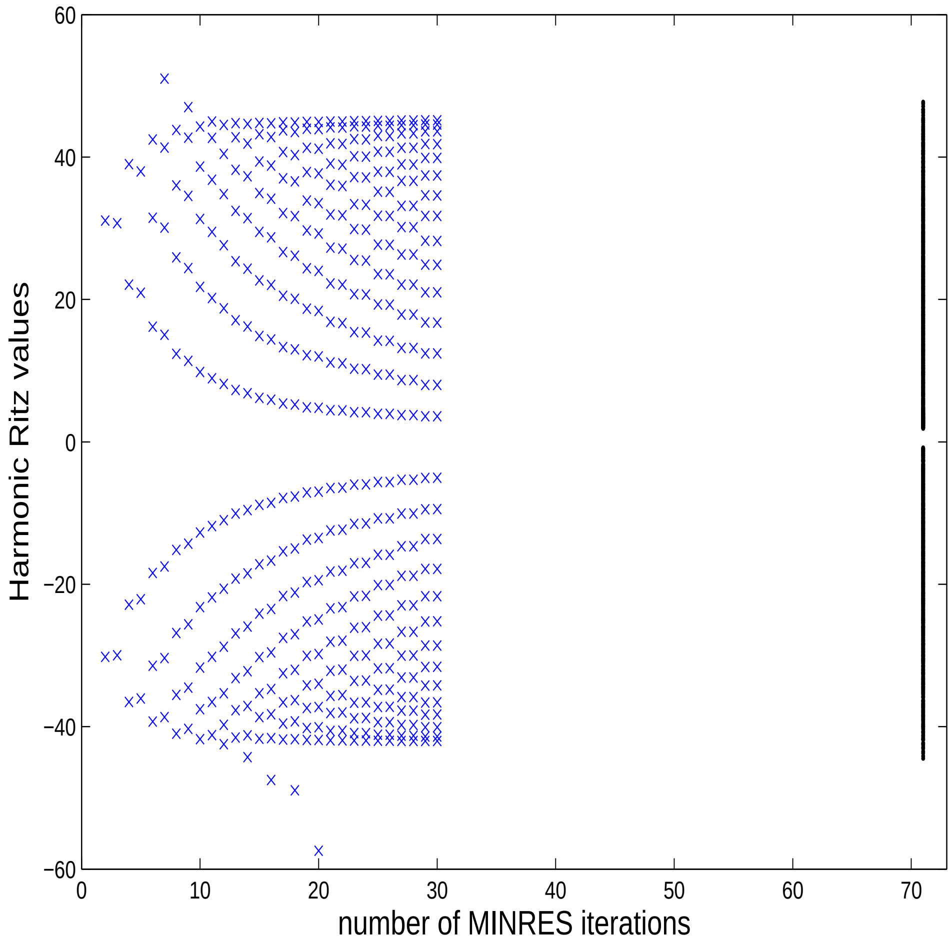
Harmonic Ritz values as iterations proceed

Typical convergence pattern



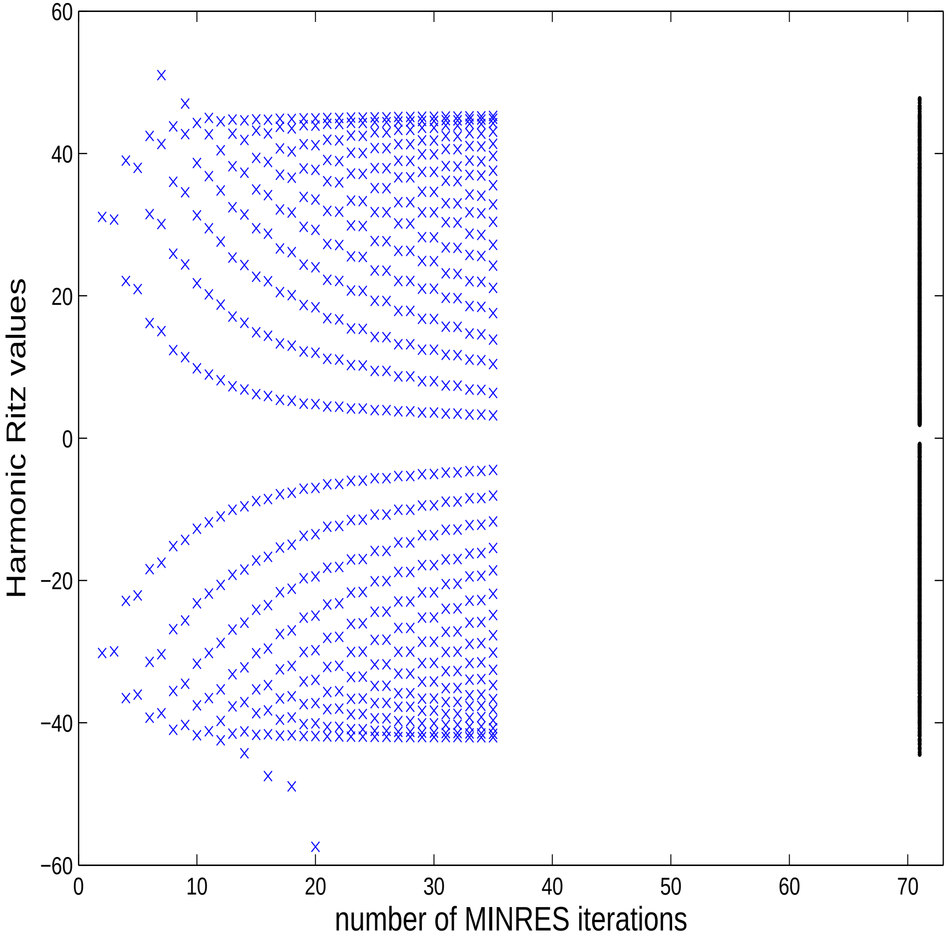
Harmonic Ritz values as iterations proceed

Typical convergence pattern



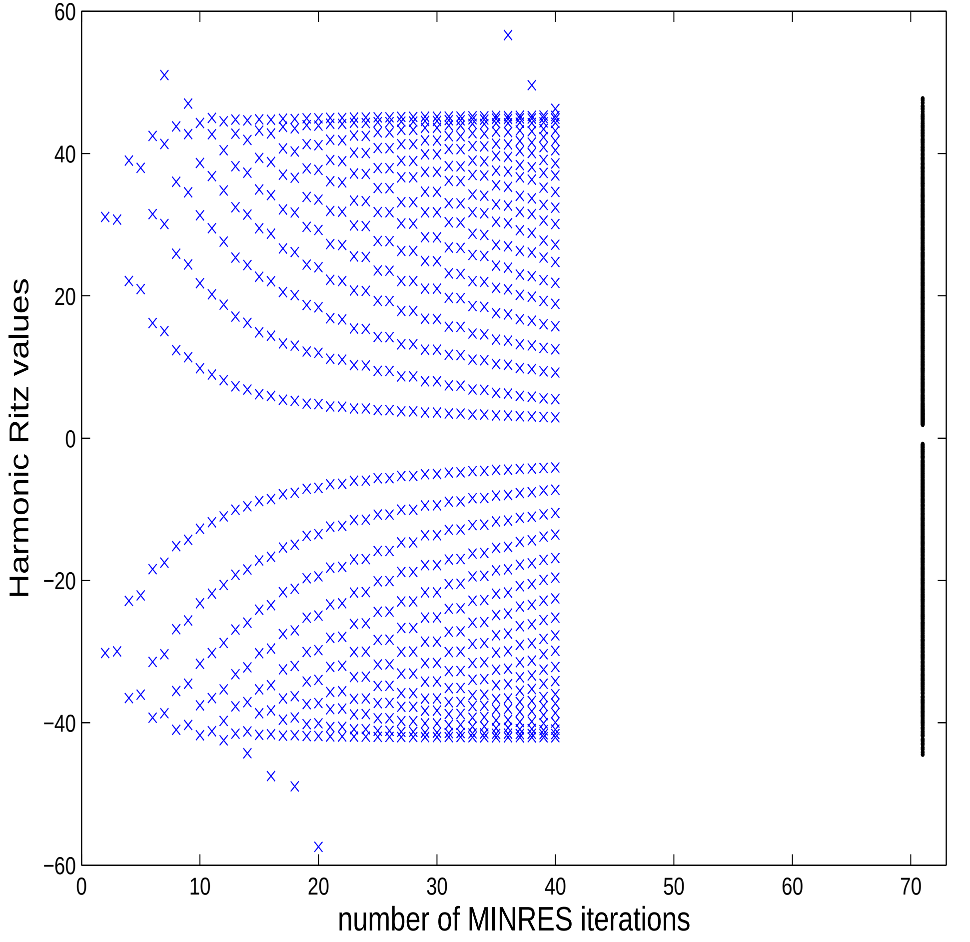
Harmonic Ritz values as iterations proceed

Typical convergence pattern



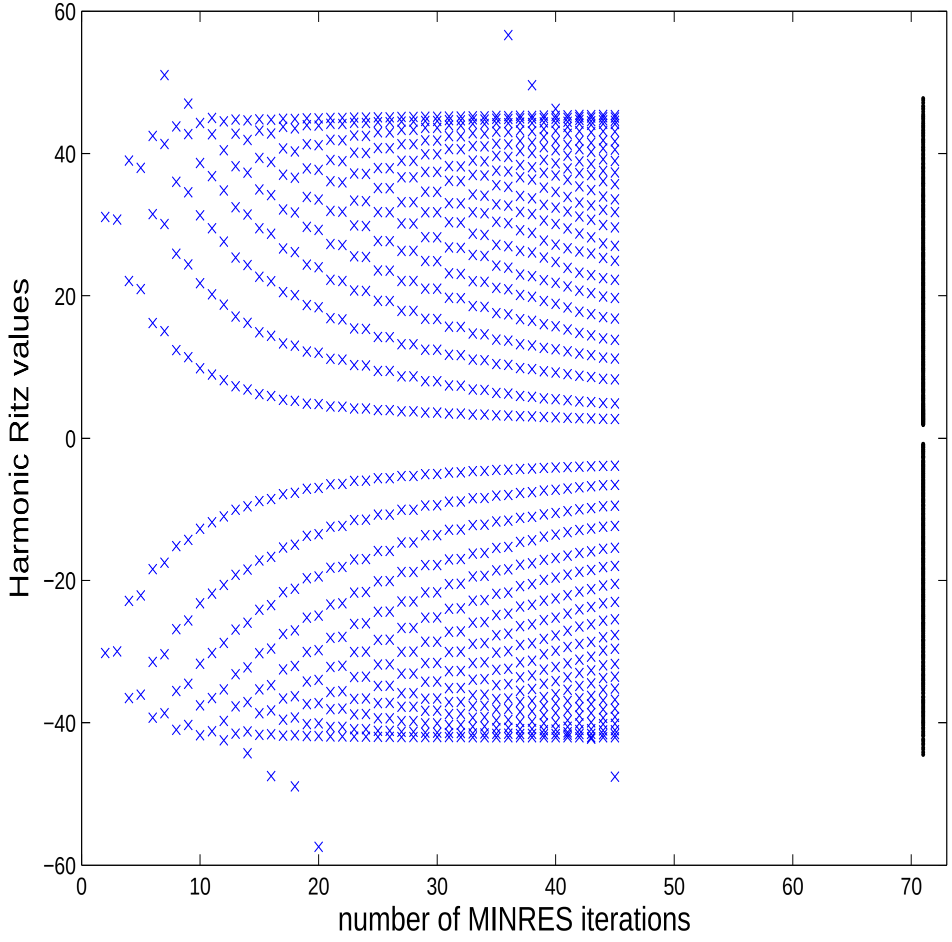
Harmonic Ritz values as iterations proceed

Typical convergence pattern



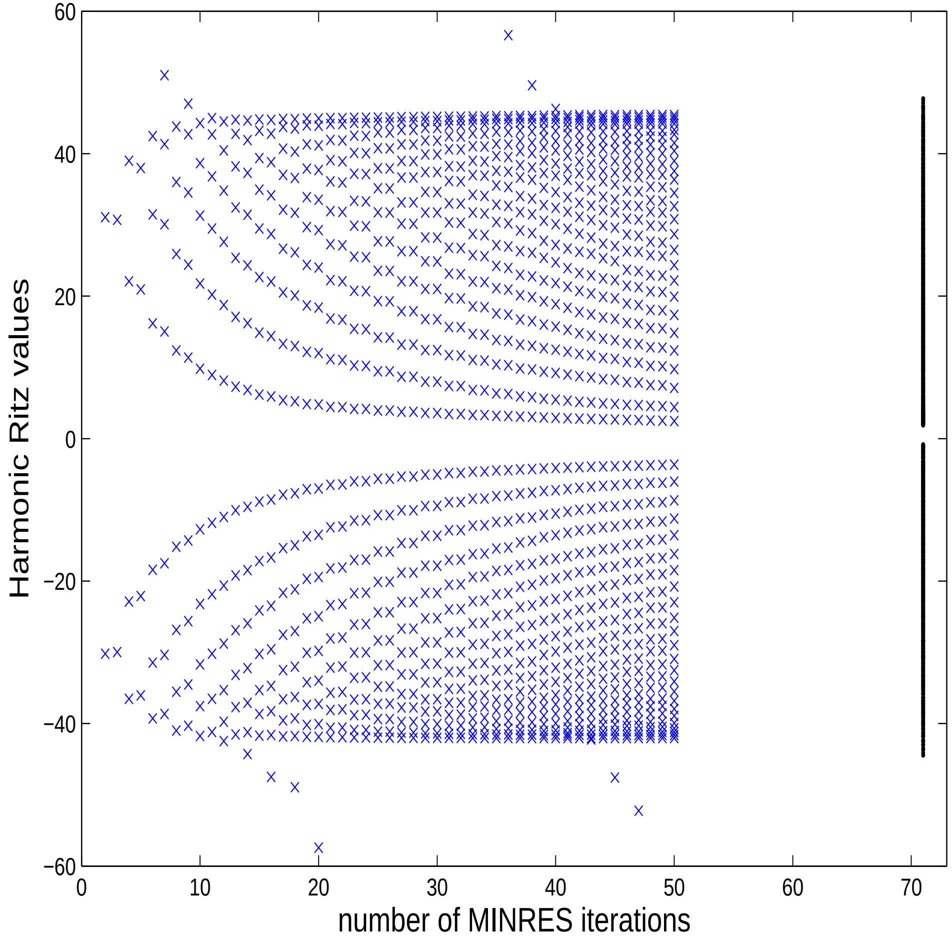
Harmonic Ritz values as iterations proceed

Typical convergence pattern



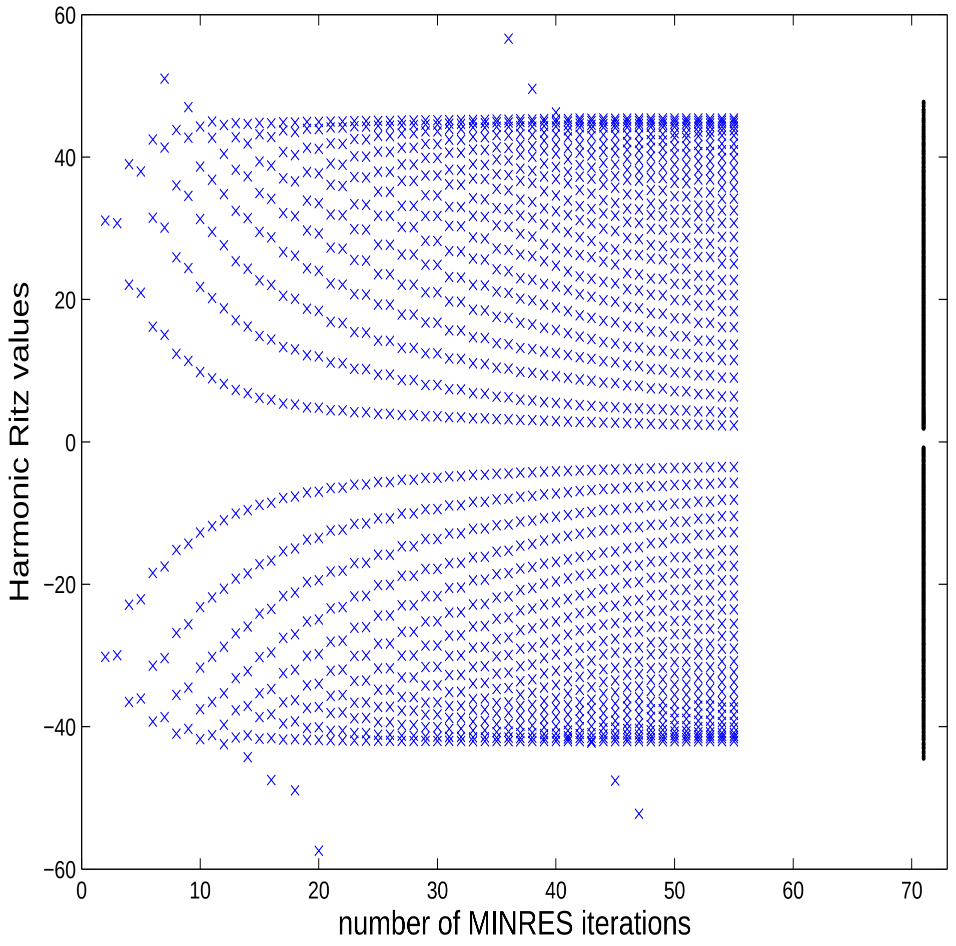
Harmonic Ritz values as iterations proceed

Typical convergence pattern



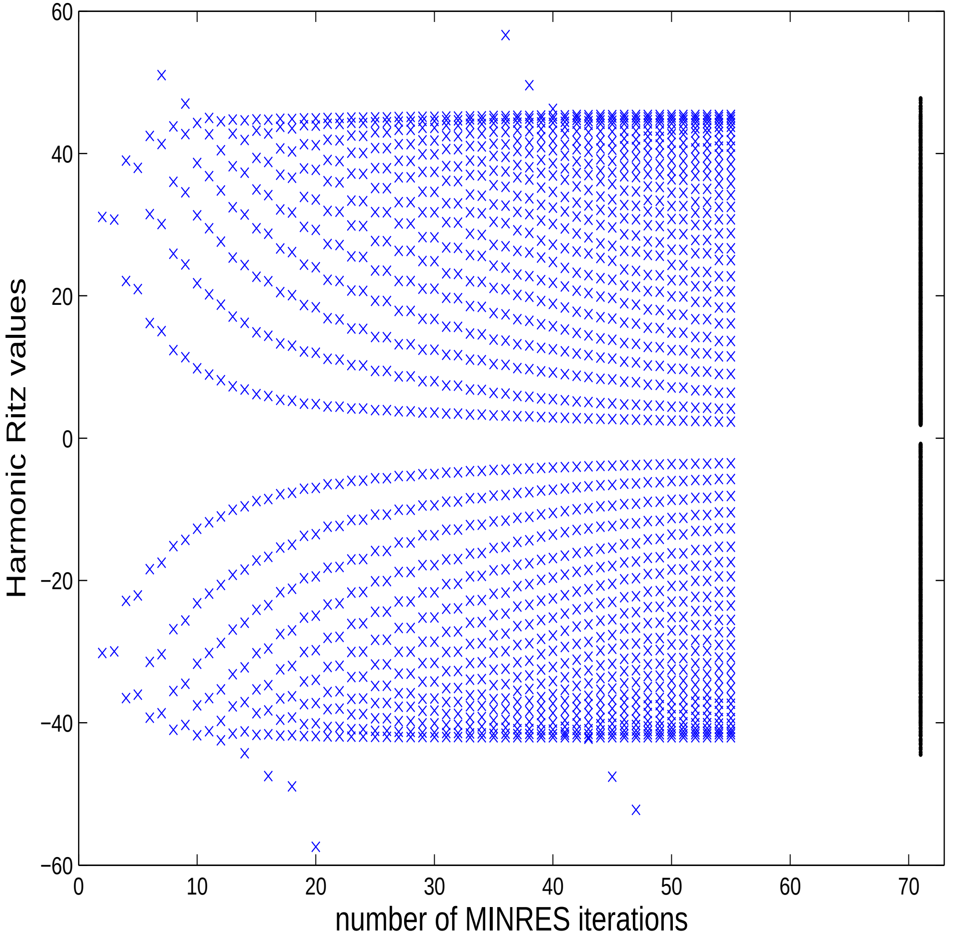
Harmonic Ritz values as iterations proceed

Typical convergence pattern



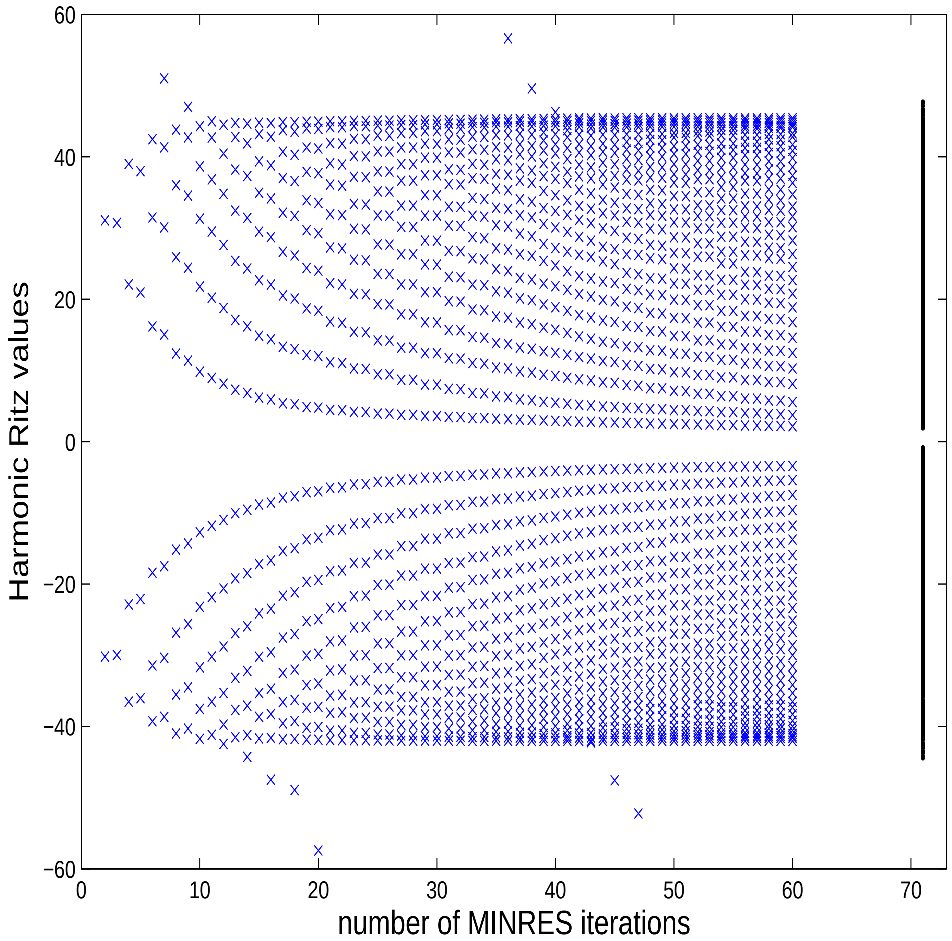
Harmonic Ritz values as iterations proceed

Typical convergence pattern



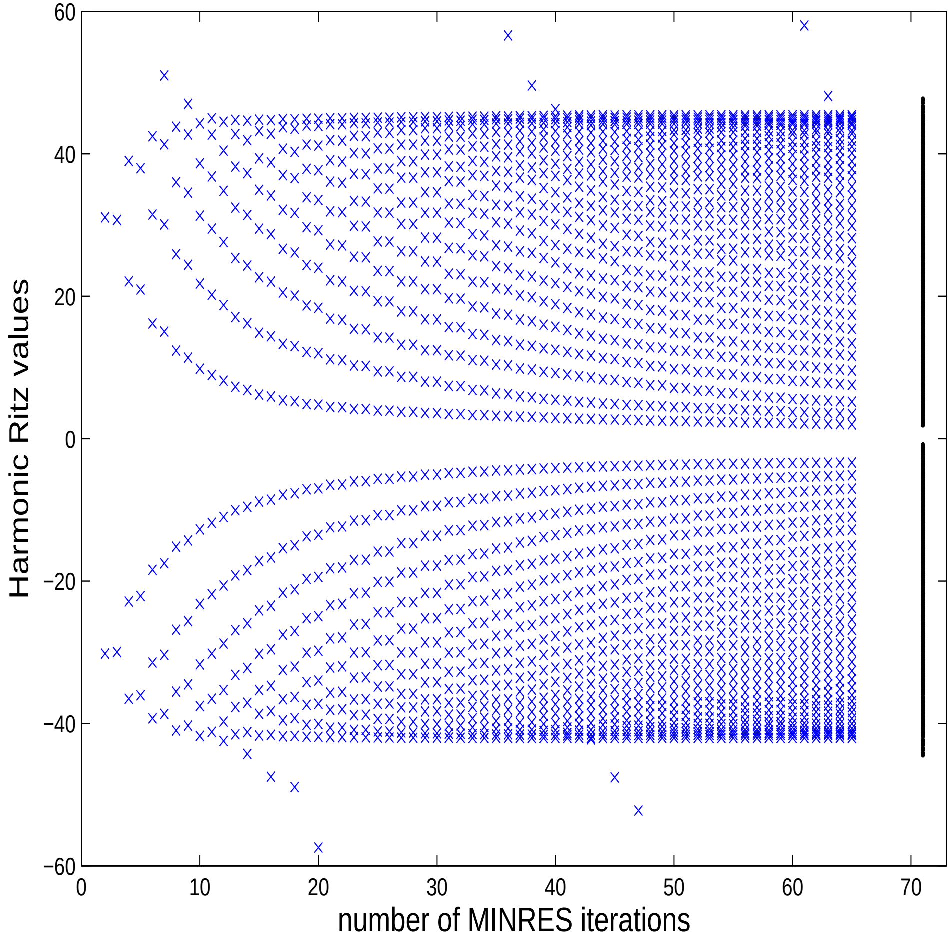
Harmonic Ritz values as iterations proceed

Typical convergence pattern



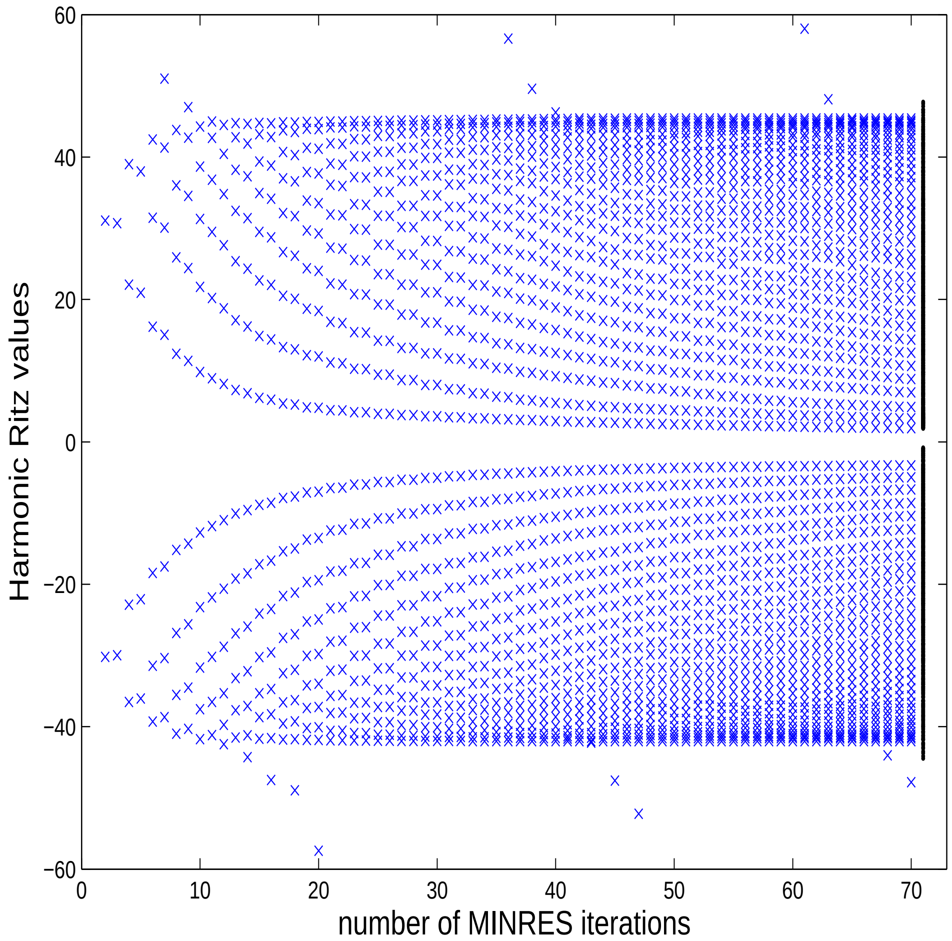
Harmonic Ritz values as iterations proceed

Typical convergence pattern



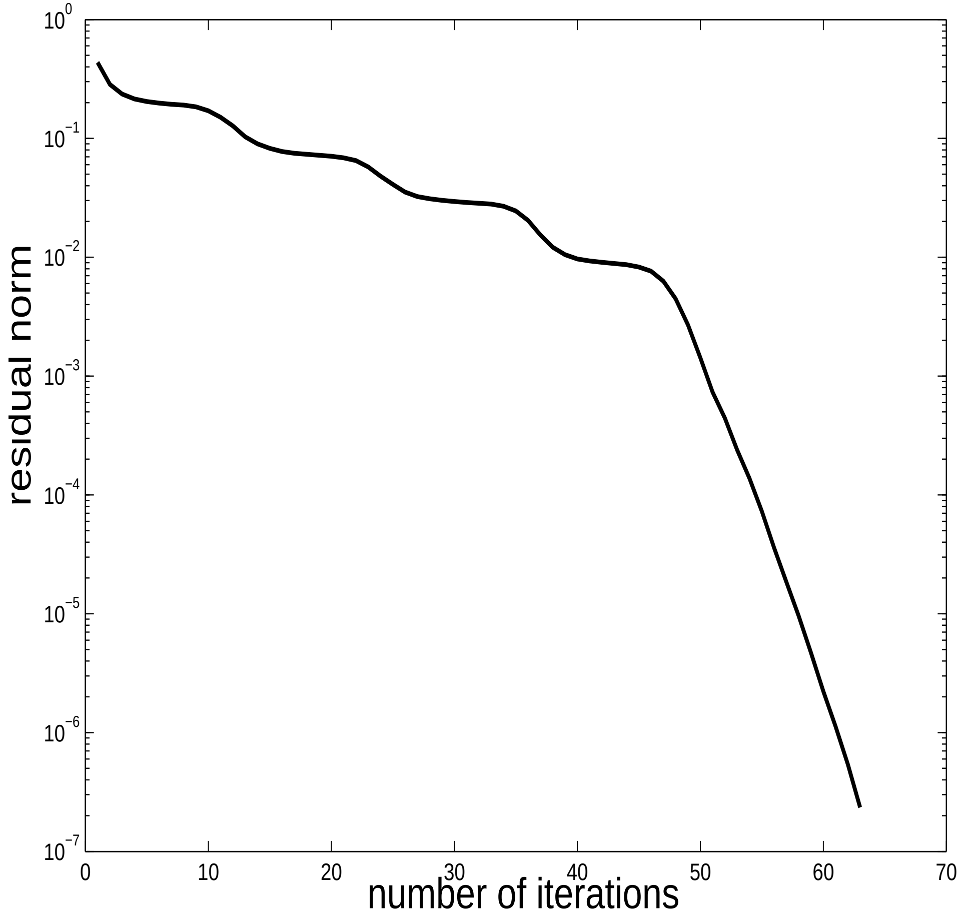
Harmonic Ritz values as iterations proceed

Typical convergence pattern



Harmonic Ritz values as iterations proceed

Superlinear convergence



Superlinear convergence

Generalization of CG well-known result (van der Sluis & van der Vorst '86)

MINRES: (van der Vorst and Vuik '93, van der Vorst '03)

(λ_i, z_i) eigenpairs of $\mathcal{M}$

TH: Assume $r_k = \bar{r}_0 + s$ with $\bar{r}_0 \perp z_1$

Let $\bar{r}_j$ be the MINRES residual after j iterations with $\bar{r}_0$. Then

$$\|r_{k+j}\| \le F_k \|\bar{r}_j\|, \quad \text{where} \quad F_k = \max_{k \ge 2} \frac{|\theta_1^{(k)}|}{|\lambda_1|} \frac{|\theta_1^{(k)} - \lambda_k|}{|\lambda_1 - \lambda_k|}$$

and $\theta_1^{(k)}$ is the harmonic Ritz value closest to λ_1 in $K_k(A, r_0)$.

(for a proof, Simoncini & Szyld '11, unpublished)

Superlinear convergence. A simple example.

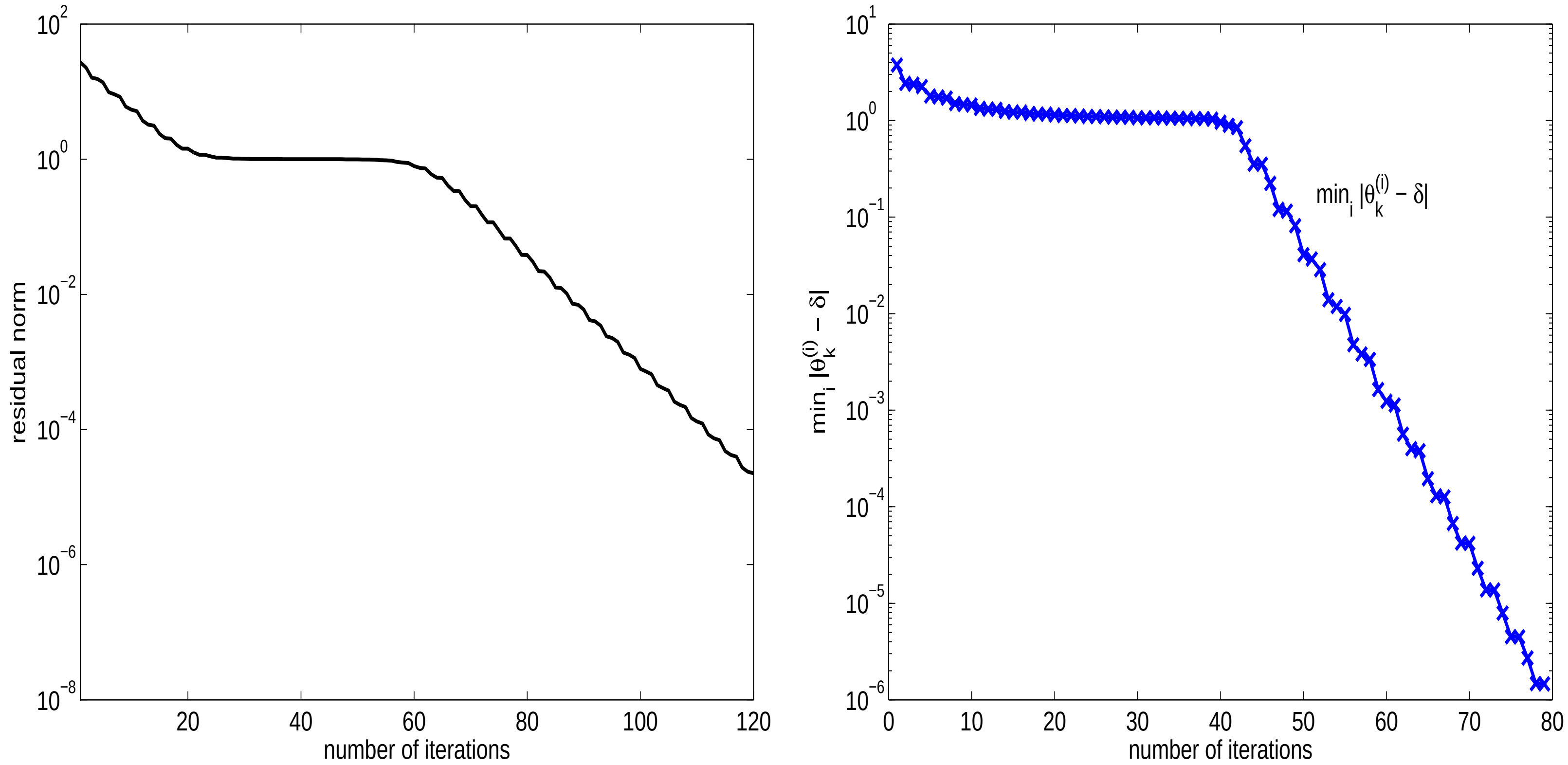
$$\mathcal{M} = \begin{bmatrix} M_- & & \\ & -\delta & \\ & & M_+ \end{bmatrix}$$

$\delta = 10^{-3}$, $\quad \mathrm{spec}(M_-) \in \mathbb{R}^-, \mathrm{spec}(M_+) \in \mathbb{R}^+$ both well clustered

$b = \mathbf{1}$

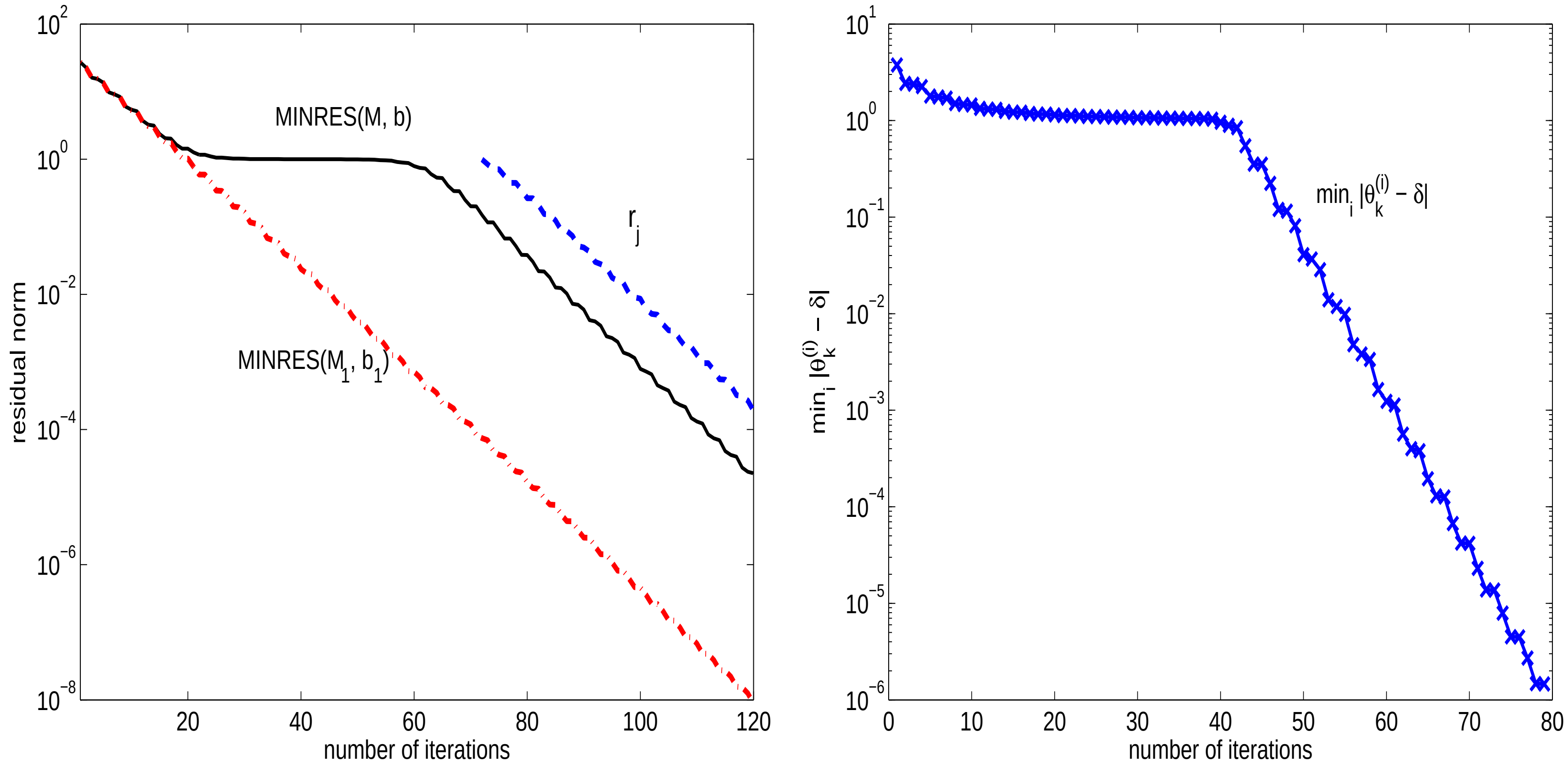
Superlinear convergence. An experiment.

$\mathcal{M}, b$ data with one tiny negative eigenvalue



Superlinear convergence. An experiment.

$\mathcal{M}, b$ data with one tiny negative eigenvalue



$\mathcal{M}_1, b_1$ data with negative eigenvalue closest to zero removed

Enhancing MINRES convergence for Saddle Point Problems

Spectral properties

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & O \end{bmatrix} \qquad \begin{array}{ll} 0 < \lambda_n \le \cdots \le \lambda_1 & \text{eigs of } A \\ 0 < \sigma_m \le \cdots \le \sigma_1 & \text{sing. vals of } B \end{array}$$

$\mathrm{spec}(\mathcal{M})$ subset of (Rusten & Winther 1992)

$$\left[\frac{1}{2}(\lambda_n - \sqrt{\lambda_n^2 + 4\sigma_1^2}), \frac{1}{2}(\lambda_1 - \sqrt{\lambda_1^2 + 4\sigma_m^2})\right] \quad \cup \quad \left[\lambda_n, \frac{1}{2}(\lambda_1 + \sqrt{\lambda_1^2 + 4\sigma_1^2})\right]$$

More results under different hypotheses

Block diagonal Preconditioner

⋆ $\mathcal{P}_{\text{ideal}} = \begin{bmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{bmatrix}$ MINRES converges in at most 3 its.

Block diagonal Preconditioner

⋆ $\mathcal{P}_{\text{ideal}} = \begin{bmatrix} A & 0 \\ 0 & BA^{-1}B^T \end{bmatrix}$ MINRES converges in at most 3 its.

A more practical choice:

$$\mathcal{P} = \begin{bmatrix} \widetilde{A} & 0 \\ 0 & \widetilde{S} \end{bmatrix} \qquad \text{spd.} \quad \widetilde{A} \sim A \qquad \widetilde{S} \sim BA^{-1}B^T$$

spectrum in $\quad [-a, -b] \cup [c, d], \qquad a, b, c, d > 0$

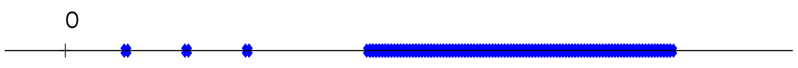
"Robust" preconditioners for PDE systems (Survey, Mardal & Winther, '11)

A quasi-optimal approximate Schur complement
(joint with M. Olshanskii)

$$\widetilde{S} \approx BA^{-1}B^T$$

For certain operators, $\widetilde{S}$ is quasi-optimal:

spec$(BA^{-1}B^T\widetilde{S}^{-1})$ well clustered except for few eigenvalues



A quasi-optimal approximate Schur complement (joint with M. Olshanskii)

$$\widetilde{S} \approx BA^{-1}B^T$$

For certain operators, $\widetilde{S}$ is quasi-optimal:

spec$(BA^{-1}B^T\widetilde{S}^{-1})$ well clustered except for few eigenvalues

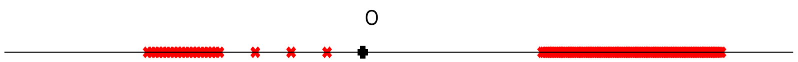
Often: well clustered eigs also mesh/parameter-independent

The role of $\widetilde{S}$

Claim:

The presence of outliers in $BA^{-1}B^T\widetilde{S}^{-1}$ is accurately inherited by the preconditioned matrix $\mathcal{M}\mathcal{P}^{-1}$

spec($BA^{-1}B^T\widetilde{S}^{-1}$)



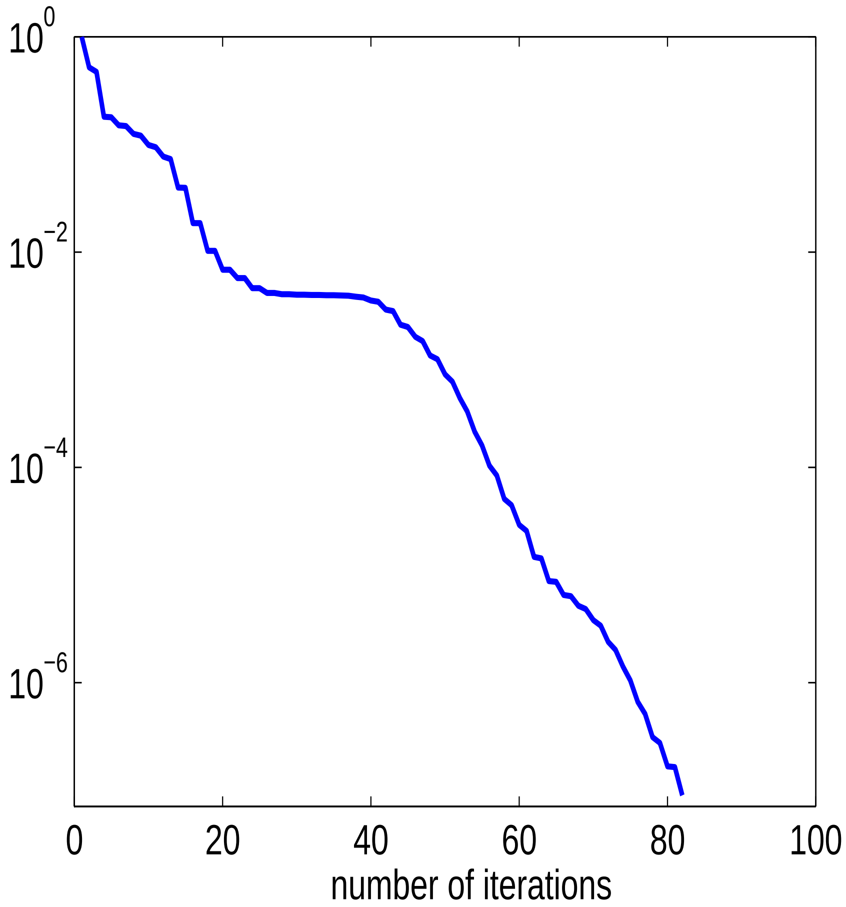
spec($\mathcal{M}\mathcal{P}^{-1}$)

(for a proof, see Olshanskii & Simoncini, SIMAX '10)

Side effects

- These spectral peculiarities have an effect on the preconditioned problem $\mathcal{M}\mathcal{P}^{-1}$
- ... and influence the convergence of MINRES on $\mathcal{M}\mathcal{P}^{-1}$
- Can we eliminate this influence?

Effect on MINRES convergence



Stokes-type problem

Eliminating the stagnation phase: “Deflated” MINRES

$Y = [y_1, \ldots, y_s]$: *approximate* eigenbasis of $\mathcal{M}$

* Approximation space: Augmented Lanczos sequence

$$v_{j+1} \perp \mathrm{span}\{Y, v_1, v_2, \ldots, v_j\}, \qquad \|v_{j+1}\| = 1$$

obtained by standard Lanczos method with coeff.matrix

$$\mathcal{G} := \mathcal{M} - \mathcal{M}Y(Y^T\mathcal{M}Y)^{-1}Y^T\mathcal{M}$$

* MINRES method:

$$r_j = \hat{b} - \mathcal{M}\hat{u}_j \ \perp \ \mathcal{G}K_j(\mathcal{G}, v_1)$$

$\Rightarrow$ $\hat{u}_j$ obtained with a short-term recurrence http://www.stanford.edu/group/SOL/software/minres.html

Augmented MINRES: Given $\mathcal{M}, b$, maxit, tol, $\mathcal{P}$, and Y with orthonormal columns

$\mathbf{u} = \mathbf{Y}(\mathbf{Y^T}\mathcal{M}\mathbf{Y})^{-1}\mathbf{Y^T b}$ starting approx, $r = b - \mathcal{M}u$, $r_1 = r$, $y = \mathcal{P}^{-1}r$, etc

while ($i <$ maxit)

$i = i + 1 \qquad v = y/\beta;$

$y = \mathcal{M}v - \mathcal{M}\mathbf{Y}(\mathbf{Y^T}\mathcal{M}\mathbf{Y})^{-1}\mathbf{Y^T}\mathcal{M}\mathbf{v}$

if $i \geq 2$, $y = y - (\beta/\beta_0)r_1$

$\alpha = v^T y, \qquad y = y - r_2\alpha/\beta$

$r_1 = r_2, \qquad r_2 = y$

$y = \mathcal{P}^{-1}r_2$

$\beta_0 = \beta, \qquad \beta = \sqrt{r_2^T y}$

$e_0 = e, \qquad \delta = c\bar{d} + s\alpha \qquad \bar{g} = s\bar{d} - c\alpha \qquad e = s\beta \qquad \bar{d} = -c\beta$

$\gamma = \|[\bar{g}, \beta]\| \qquad c = \bar{g}/\gamma, \qquad s = \beta/\gamma, \qquad \phi = c\bar{\phi}, \qquad \bar{\phi} = s\bar{\phi}$

$w_1 = w_2, \qquad w_2 = w$

$w = (v - e_0 w_1 - \delta w_2)\gamma^{-1}$

$\mathbf{g} = \mathbf{Y}(\mathbf{Y^T}\mathcal{M}\mathbf{Y})^{-1}\mathbf{Y^T}\mathcal{M}\mathbf{w}\boldsymbol{\phi}$

$u = u - \mathbf{g} + \phi w$

$\zeta = \chi_1/\gamma, \qquad \chi_1 = \chi_2 - \delta z, \qquad \chi_2 = -e\zeta$

Check preconditioned residual norm ($\bar{\phi}$) for convergence

end

Stokes-type problem with variable viscosity in $\Omega \subset \mathbb{R}^d$

$$\begin{aligned} -\mathbf{div}\nu(\mathbf{x})\mathbf{Du} + \nabla p = \mathbf{f} \quad &\text{in} \quad \Omega, \\ -\operatorname{div}\, \mathbf{u} = 0 \quad &\text{in} \quad \Omega, \\ \mathbf{u} = 0 \quad &\text{on} \quad \partial\Omega, \end{aligned}$$

with $0 < \nu_{\min} \leq \nu(\mathbf{x}) \leq \nu_{\max} < \infty$. (Here, $\nu(\mathbf{x}) = 2\mu + \frac{\tau_s}{\sqrt{\varepsilon^2 + |\mathbf{Du}(\mathbf{x})|^2}}$)

$\mathbf{u}$: velocity vector field $\qquad$ p : pressure

$\mathbf{Du} = \frac{1}{2}(\nabla\mathbf{u} + \nabla^T\mathbf{u})$ rate of deformation tensor

Stokes-type problem with variable viscosity in $\Omega \subset \mathbb{R}^d$

$$
\begin{aligned}
-\mathbf{div}\nu(\mathbf{x})\mathbf{Du} + \nabla p = \mathbf{f} &\quad \text{in} \quad \Omega, \\
-\operatorname{div}\,\mathbf{u} = 0 &\quad \text{in} \quad \Omega, \\
\mathbf{u} = 0 &\quad \text{on} \quad \partial\Omega,
\end{aligned}
$$

with $0 < \nu_{\min} \leq \nu(\mathbf{x}) \leq \nu_{\max} < \infty$. (Here, $\nu(\mathbf{x}) = 2\mu + \frac{\tau_s}{\sqrt{\varepsilon^2 + |\mathbf{Du}(\mathbf{x})|^2}}$)

$\mathbf{u}$: velocity vector field $\qquad p$: pressure

$\mathbf{Du} = \frac{1}{2}(\nabla\mathbf{u} + \nabla^T\mathbf{u})$ rate of deformation tensor

Prec. S: pressure mass matrix wrto weighted product $(\nu^{-1}\cdot, \cdot)_{L^2(\Omega)}$

$\Rightarrow (BA^{-1}B^T)S^{-1}$ well clustered spectrum,

except for few small positive eigenvalues (Grinevich & Olshanskii, '09)

Bercovier-Engelman regularized model of Bingham viscoplastic fluid

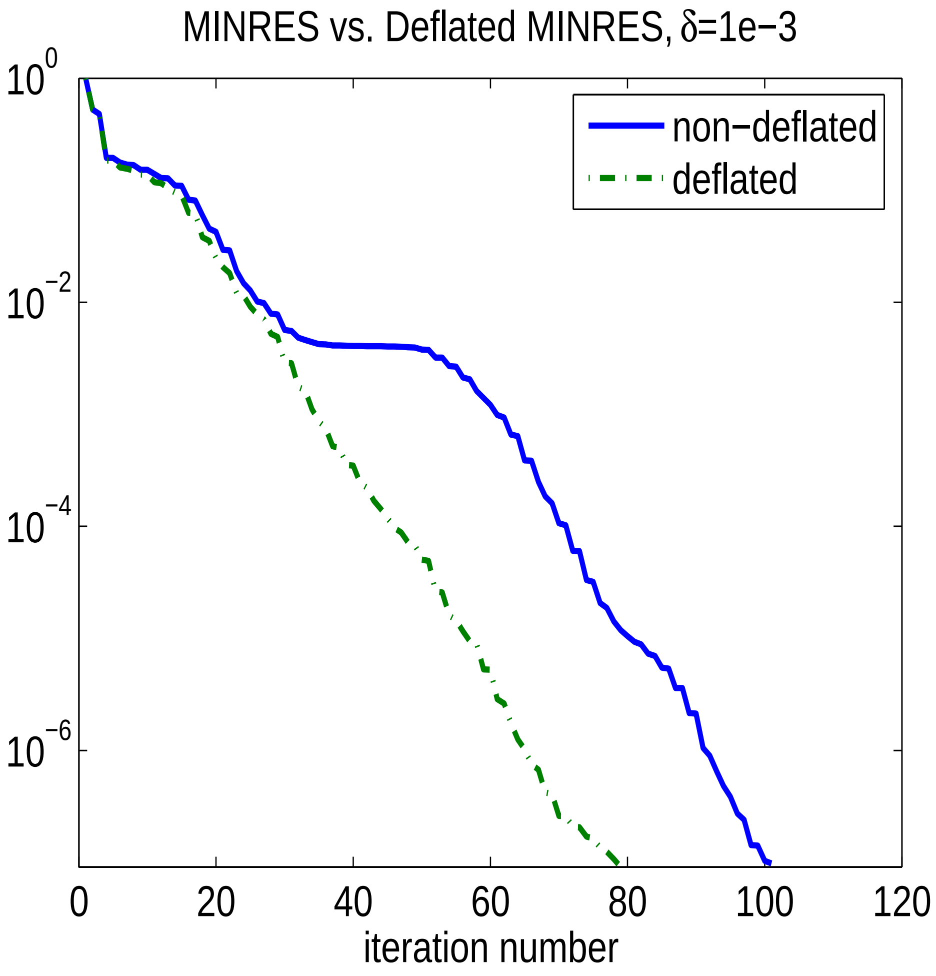
* One zero pressure mode (eigvec easy to approx)

* One small eigenvalue of precon'd Schur Complement

$\Rightarrow$ rough eigvec approximation : $\{\widetilde{u}_2, \widetilde{p}_2\}^T \approx \{u_2, p_2\}^T$

$$\widetilde{p}_2 = \begin{cases} 0 & \text{if } \frac{1}{2} - \tau_s \leq y \leq \frac{1}{2} + \tau_s, \\ 1 & \text{if } 0 \leq y < \frac{1}{2} - \tau_s, \\ -1 & \text{if } 1 \geq y > \frac{1}{2} + \tau_s, \end{cases} \qquad \text{and} \quad \widetilde{u}_2 = -\widetilde{A}^{-1} B^T \widetilde{p}_2$$

Original and "Deflated" convergence histories



$\widetilde{A} = \mathrm{IC}(A, \delta), \delta = 10^{-3}$ (same with $\delta = 0$)

A stopping criterion for Stokes mixed approximation (joint with D. Silvester)

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} \widetilde{A} & 0 \\ 0 & \widetilde{S} \end{bmatrix} \qquad \widetilde{A} \sim A, \quad \widetilde{S} \sim BA^{-1}B^T$$

"Natural" norm for the problem:

energy norm in the u-space and p-space, $\|x - x_k\|_{\mathcal{P}_{\mathrm{ideal}}}$

A stopping criterion for Stokes mixed approximation
(joint with D. Silvester)

$$\mathcal{M} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} \widetilde{A} & 0 \\ 0 & \widetilde{S} \end{bmatrix} \qquad \widetilde{A} \sim A, \quad \widetilde{S} \sim BA^{-1}B^T$$

"Natural" norm for the problem:

energy norm in the u-space and p-space, $\|x - x_k\|_{\mathcal{P}_{\text{ideal}}}$

For stable discretization, heuristic relation between error and residual:

$$\|x - x_k\|_{\mathcal{P}_{\text{ideal}}} \leq \frac{\sqrt{2}}{\gamma^2} \|b - \mathcal{M}x_k\|_{\mathcal{P}_{\text{ideal}}^{-1}} \sim \frac{\sqrt{2}}{\gamma^2} \|b - \mathcal{M}x_k\|_{\mathcal{P}^{-1}} < tol$$

γ inf-sup constant: $\qquad \gamma^2 \leq \frac{q^T BA^{-1}B^T q}{q^T Q q}, \quad \forall q \neq 0$

Estimating the inf-sup constant

For the preconditioned problem (Elman etal, '05):

$$\lambda_- \le \frac{1}{2}(\delta - \sqrt{\delta^2 + 4\delta\gamma^2}) \qquad \delta \le \lambda_+$$

with $\delta = \lambda_{\min}(A\widetilde{A}^{-1})$. If these bounds are tight (equalities), then

$$\gamma^2 = \frac{\lambda_-^2 - \lambda_-\lambda_+}{\lambda_+}$$

Estimating the inf-sup constant

For the preconditioned problem (Elman etal, '05):

$$\lambda_- \le \frac{1}{2}(\delta - \sqrt{\delta^2 + 4\delta\gamma^2}) \qquad \delta \le \lambda_+$$

with $\delta = \lambda_{\min}(A\widetilde{A}^{-1})$. If these bounds are tight (equalities), then

$$\gamma^2 = \frac{\lambda_-^2 - \lambda_-\lambda_+}{\lambda_+}$$

In practice, adaptive estimate with Harmonic Ritz values:

$$\gamma^2 \approx \gamma_k^2 = \frac{(\theta_-^{(k)})^2 - \theta_-^{(k)}\theta_+^{(k)}}{\theta_+^{(k)}}, \qquad kth \text{ MINRES iteration}$$

Corresponding analogous results for Potential flow problem

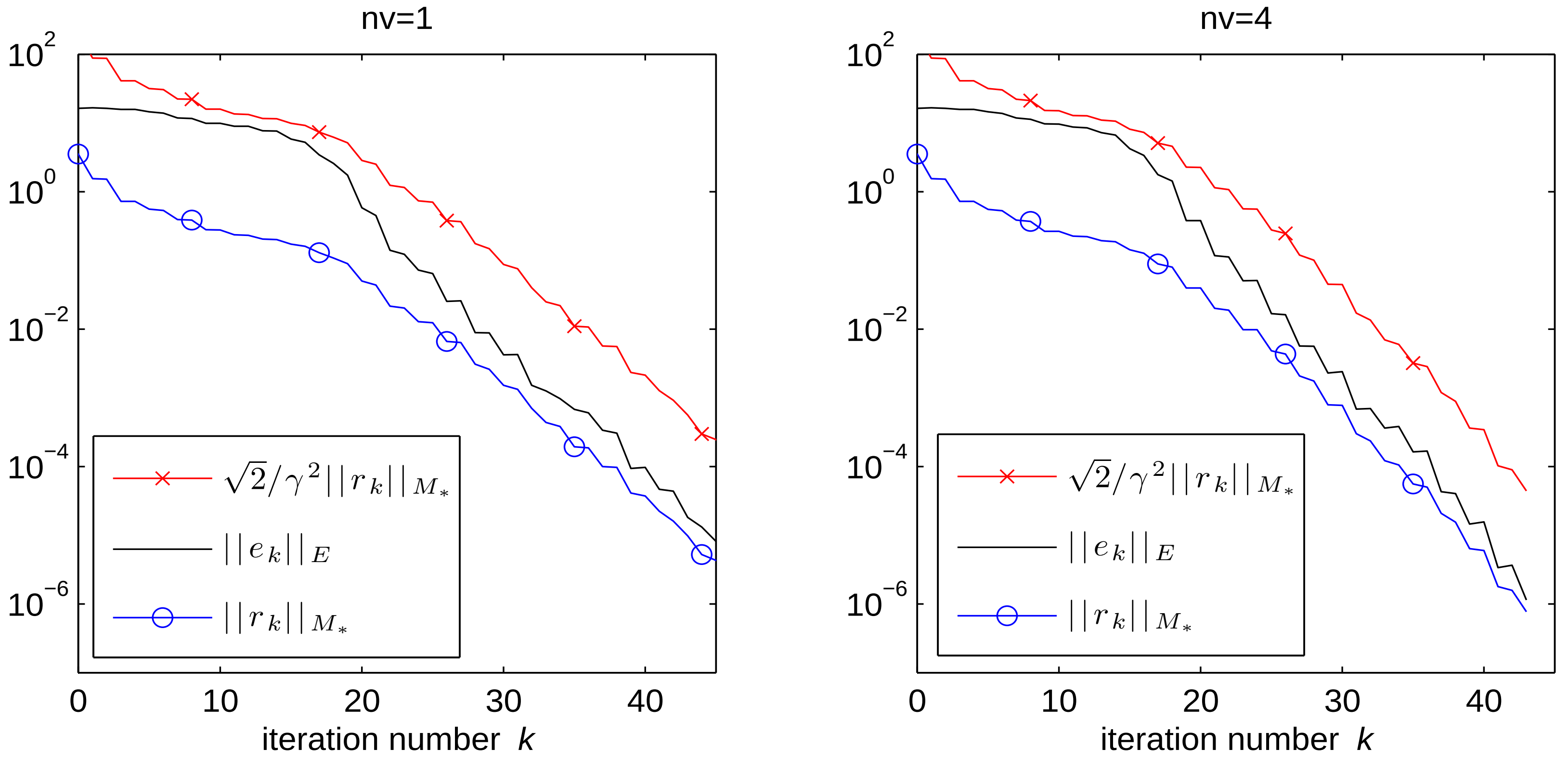
IFISS Problem: smooth colliding flow with quartic polyn velocity soln

k	γ_k	$\lambda_k \searrow \lambda_+$
4	0.9807	1.0212
5	0.9340	1.0065
6	0.9147	1.0024
7	0.6689	0.9896
8	0.6618	0.9887
9	0.3581	0.9705
10	0.3502	0.9671
11	0.2992	0.9305
12	0.2917	0.9218
13	0.2684	0.9071
14	0.2576	0.9031
15	0.2468	0.9000
16	0.2328	0.8973
17	0.2320	0.8971
18	0.2239	0.8945
19	0.2238	0.8945
20	0.2177	0.8904
21	0.2174	0.8897

`Optimal gamma=0.2162`

Another IFISS example. Flow over a step. Optimal γ

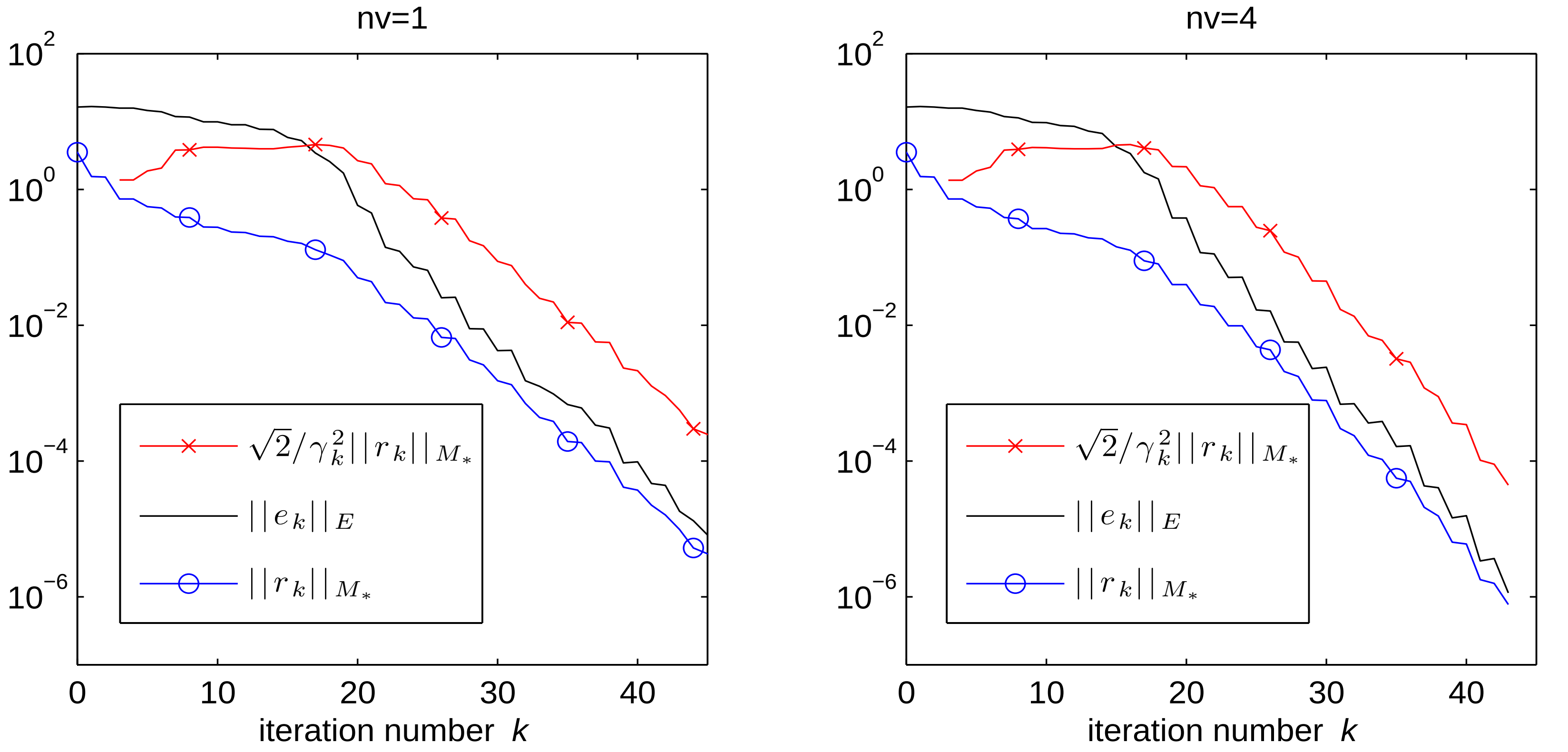
e_k: error at iteration k r_k: residual at iteration k



$$E = \mathcal{P}_{\text{ideal}}, \qquad M_* = \mathcal{P}^{-1}$$

Another IFISS example. Flow over a step. Adaptive γ_k

e_k: error at iteration k $\qquad$ r_k: residual at iteration k



$$E = \mathcal{P}_{\text{ideal}}, \qquad M_* = \mathcal{P}^{-1}$$

Conclusions

- MINRES is rich in information to be exploited
- Adaptive problem-related stopping criteria available

References:

* Maxim A. Olshanskii and V. Simoncini *Acquired clustering properties and solution of certain saddle point systems*. SIMAX, 2010.

* David J. Silvester and V. Simoncini *An Optimal Iterative Solver for Symmetric Indefinite Systems stemming from Mixed Approximation*. ACM TOMS, 2011.

* V. Simoncini and Daniel B. Szyld , unpublished, 2011.