Using Spectral Information to Precondition Saddle-Point Systems

Daniel Ruiz Annick Sartenaer Charlotte Tannier

IMA Department
INPT - ENSEEIHT - IRIT

Department of Mathematics
University of Namur

Recent Advances on Optimization, Toulouse, July 2013


Mrs Iserles
A.R. Conn
N. Gould
Ph. Toint
Mrs Powell

Outline

The target

Design efficient preconditioners to solve the system:

$$\mathcal{A}u = b \Leftrightarrow \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

with $A = A^T \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$ $(m \leq n)$

Overall goal

Find $\mathcal{P}$ such that $\mathcal{P}^{-1}\mathcal{A}$ has a better eigenvalue clustering and a reduced condition number

$\longrightarrow$ To reduce the number of MINRES iterations

"Ideal" block diagonal preconditioner

$$\mathcal{A}u = b \Leftrightarrow \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

Murphy, Golub and Wathen (2000):

$$\mathcal{P} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix}$$

where $S = B^T A^{-1} B$ is the Schur complement

Spectral properties

$\mathcal{P}^{-1}\mathcal{A}$ has at most four distinct eigenvalues $0, 1, \frac{1 \pm \sqrt{5}}{2}$

Color code*

As much as possible:

- ■ $\rightarrow$ when related to A
- ■ $\rightarrow$ when related to B
- ■ $\rightarrow$ when related to S (or any combination of A and B)

*Except for the pictures

"Approximate" block diagonal preconditioners

$$\mathcal{P} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \rightarrow \tilde{\mathcal{P}} = \begin{bmatrix} \tilde{A} & 0 \\ 0 & \tilde{S} \end{bmatrix}$$

- Golub, Greif and Varah (2006):

$$\tilde{A} = A + BWB^T \text{ and } \tilde{S} = B^T(A + BWB^T)^{-1}B$$

  with W symmetric positive semidefinite

- Rees, Dollar and Wathen (2010):

$$\tilde{A} \approx A \text{ and } \tilde{S} \approx S$$

  adapted to PDE-constrained optimization

- Olshanskii and Simoncini (2010) for a spectral-based analysis
- Benzi, Golub and Liesen (2005) for a survey

Assumptions and framework

$$\mathcal{A}u = b \Leftrightarrow \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

- A is symmetric positive definite
- B has full column rank

Additional assumptions

- A has few very small eigenvalues
- These eigenvalues and their associated eigenvectors are available (or good approximations)

In practice

How can we meet these additional assumptions?

- A has few very small eigenvalues

  $\rightarrow$ First-level preconditioner

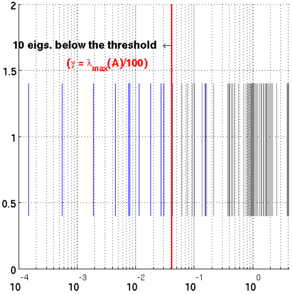
- These eigenvalues and their associated eigenvectors are available (or good approximations)

  $\rightarrow$ Krylov method with Chebyshev filters on $Ax = f$

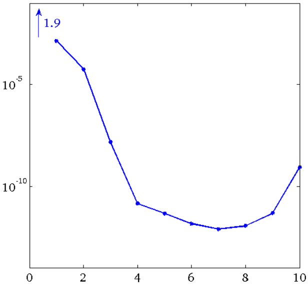
Chebyshev-based Krylov method

$\{\tilde{\lambda}_1, \cdots, \tilde{\lambda}_{10}\}$ and $[\tilde{u}_1 \ \cdots \ \tilde{u}_{10}]$:

obtained using a Conjugate Gradient method preconditioned by Chebyshev polynomials playing the role of spectral filter



$\tilde{\lambda}_i$ (in blue)



$$\frac{\|A\tilde{u}_i - \tilde{\lambda}_i \tilde{u}_i\|}{|\tilde{\lambda}_i|}$$

Golub, Ruiz and Touhami (2007)

SLRU-type spectral approximation of A^{-1}

$$\mathcal{P} = \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \rightarrow \tilde{\mathcal{P}} = \begin{bmatrix} \tilde{A} & 0 \\ 0 & \tilde{S} \end{bmatrix} \; ?$$

Consider the low-rank spectral approximation*:

$$A_\gamma^{-1} = \frac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

- $\lambda_{\min}(A) \leq \gamma \leq \lambda_{\max}(A)$
- $\Lambda_\gamma = \mathrm{diag}\{\lambda_i\}_{i=1}^p$ with $\lambda_i \leq \gamma$
- $U_\gamma \in \mathbb{R}^{n \times p}$ is the set of associated orthonormal eigenvectors
- $\alpha > 0$ is a scaling parameter

*Carpentieri, Duff and Giraud (2003)

Spectral properties of $A_\gamma^{-1}A$

$$A_\gamma^{-1} = \frac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

The eigenvalues $\{\mu_i\}_{i=1}^n$ of $A_\gamma^{-1}A$ satisfy:

$$\begin{cases} \mu_i = 1 + \frac{\lambda_i}{\alpha} & \text{if} \quad \lambda_i \leq \gamma \quad (p \text{ eigenvalues}) \\ \mu_i = \frac{\lambda_i}{\alpha} & \text{if} \quad \lambda_i > \gamma \quad (n - p \text{ eigenvalues}) \end{cases}$$

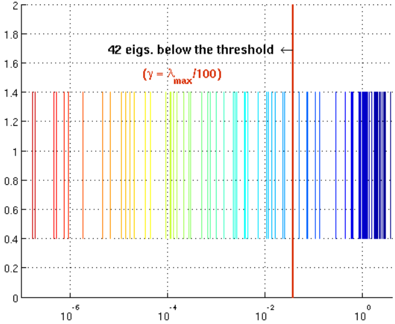
The eigenvalues $\{\mu_i\}_{i=1}^n$ of $A_\gamma^{-1}A$ are bounded within the interval

$$\left[\min\left(\frac{\alpha + \lambda_{\min}(A)}{\alpha}, \frac{\gamma}{\alpha}\right), \max\left(\frac{\alpha + \gamma}{\alpha}, \frac{\lambda_{\max}(A)}{\alpha}\right)\right]$$

$\longrightarrow$ In terms of $\lambda_{\min}(A), \lambda_{\max}(A), \gamma$ and α

Illustration

$$A \in \mathbb{R}^{300\times300},\ \lambda_{\min}(A) \approx 1.7\ 10^{-7},\ \lambda_{\max}(A) \approx 3.8$$



Spectrum of A

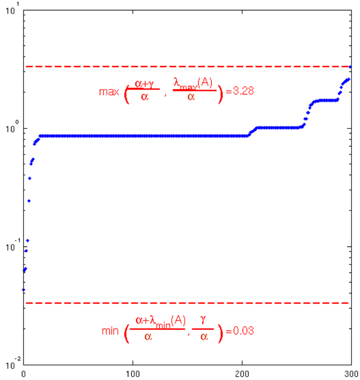
- $\gamma = \dfrac{\lambda_{\max}(A)}{100} \approx 3.8\ 10^{-2}$
- $\Lambda_\gamma = \mathrm{diag}\{\lambda_i\}_{i=1}^{42} \quad (\lambda_i \leq \gamma)$
- $U_\gamma \in \mathbb{R}^{300\times 42}$
- $\alpha = \dfrac{\mathrm{tr}(A)-\mathrm{tr}(\Lambda_\gamma)}{258} = 1.16$

Illustration (continued)

- $A \in \mathbb{R}^{300\times 300}$, $\lambda_{\min}(A) \approx 1.7\ 10^{-7}$, $\lambda_{\max}(A) \approx 3.8$
- $A_\gamma \in \mathbb{R}^{300\times 300}$, $\lambda_{\min}(A_\gamma) \approx 1.7\ 10^{-7}$, $\lambda_{\max}(A_\gamma) \approx 1.2$
- $A_\gamma^{-1} A \in \mathbb{R}^{300\times 300}$, $\mu_{\min} \approx 4.3\,10^{-2}$, $\mu_{\max} \approx 3.3$

The eigenvalues of $A_\gamma^{-1} A$ are guaranteed to be in the interval

$$[3.3\,10^{-2}, 3.3]$$



Eigenvalue dist. of $A_\gamma^{-1} A$

Approximation of the Schur complement and of its inverse

$$S_\gamma = B^T A_\gamma^{-1} B = B^T \left(\frac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T\right) B$$

whose inverse is given by (Sherman-Morrison-Woodbury formula):

$$S_\gamma^{-1} = \alpha (B^T B)^{-\frac{1}{2}} (I_m - K \underbrace{(\tfrac{1}{\alpha} \Lambda_\gamma + K^T K)^{-1}}_{p \times p} K^T)(B^T B)^{-\frac{1}{2}}$$

where $K = (B^T B)^{-\frac{1}{2}} B^T U_\gamma \in \mathbb{R}^{m \times p}$

Approximation of the Schur complement and of its inverse

$$S_\gamma = B^T A_\gamma^{-1} B = B^T \left(\tfrac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T\right) B$$

whose inverse is given by (Sherman-Morrison-Woodbury formula):

$$S_\gamma^{-1} = \alpha (B^T B)^{-\frac{1}{2}} (I_m - K \underbrace{(\tfrac{1}{\alpha}\Lambda_\gamma + K^T K)}_{p \times p}{}^{-1} K^T)(B^T B)^{-\frac{1}{2}}$$

where $K = (B^T B)^{-\frac{1}{2}} B^T U_\gamma \in \mathbb{R}^{m \times p}$

Remark: The singular values of K correspond to the cosines of the principal angles between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$

Spectral properties of $S_\gamma^{-1}S$

Main result

The eigenvalues $\{\nu_i\}_{i=1}^m$ of $S_\gamma^{-1}S$ are bounded within the interval

$$\left[\frac{\alpha}{\alpha+\lambda_{\max}(A)+\gamma}, \frac{\alpha+\gamma}{\gamma}\right]$$

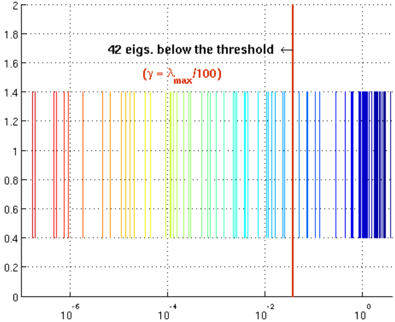
- In terms of $\lambda_{\max}(A), \gamma$ and α
- $\kappa(S_\gamma^{-1}S)$ is fully controlled by the choice of α and γ:

e.g., if $\gamma = \dfrac{\lambda_{\max}(A)}{100}$ and $\alpha = \dfrac{\lambda_{\max}(A)+\gamma}{2}$, then

$$\nu_i \in \left[\frac{1}{3}, \frac{103}{2}\right] \quad \text{and} \quad \kappa(S_\gamma^{-1}S) \leq 154.5$$

Illustration

$$A \in \mathbb{R}^{300\times 300},\ \lambda_{\min}(A) \approx 1.7\ 10^{-7},\ \lambda_{\max}(A) \approx 3.8$$



Spectrum of A

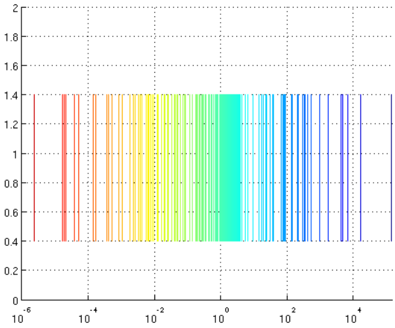
- $\gamma = \frac{\lambda_{\max}(A)}{100} \approx 3.8\ 10^{-2}$
- $\Lambda_\gamma = \mathrm{diag}\{\lambda_i\}_{i=1}^{42} \quad (\lambda_i \le \gamma)$
- $U_\gamma \in \mathbb{R}^{300\times 42}$
- $\alpha = \frac{\mathrm{tr}(A)-\mathrm{tr}(\Lambda_\gamma)}{258} = 1.16$

The eigenvalues of $S_\gamma^{-1} S$ are guaranteed in

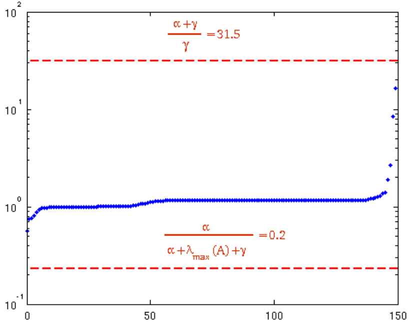
$[0.2, 31.5]$

Illustration (continued)

- $S \in \mathbb{R}^{150\times150}$, $\lambda_{\min}(S) \approx 2.4\ 10^{-6}$, $\lambda_{\max}(S) \approx 1.4\ 10^{5}$
- $S_\gamma \in \mathbb{R}^{150\times150}$, $\lambda_{\min}(S_\gamma) \approx 2.1\ 10^{-6}$, $\lambda_{\max}(S_\gamma) \approx 1.4\ 10^{5}$
- $S_\gamma^{-1} S \in \mathbb{R}^{150\times150}$, $\nu_{\min} \approx 0.56$, $\nu_{\max} \approx 16.28$



Spectrum of S



Eigenvalue dist. of $S_\gamma^{-1} S$

Two alternatives for an efficient spectral preconditioner

$$\mathcal{P}_1 = \begin{bmatrix} A & 0 \\ 0 & S_\gamma \end{bmatrix}$$

$$\mathcal{P}_2 = \begin{bmatrix} A_\gamma & 0 \\ 0 & S_\gamma \end{bmatrix}$$

with

$$A_\gamma^{-1} = \frac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

$$S_\gamma^{-1} = \alpha (B^T B)^{-\frac{1}{2}} (I_m - K(\frac{1}{\alpha}\Lambda_\gamma + K^T K)^{-1} K^T)(B^T B)^{-\frac{1}{2}}$$

where $K = (B^T B)^{-\frac{1}{2}} B^T U_\gamma$

Spectral properties of $\mathcal{A}$

$$\mathcal{A} = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}$$

- A is symmetric positive definite, with eigenvalues $\{\lambda_i\}_{i=1}^n$
- B has full column rank, with singular values $\{\sigma_i\}_{i=1}^m$

Rusten and Winther (1992)

The eigenvalues of $\mathcal{A}$ are bounded within $I^- \cup I^+$ with

$$I^- = \left[\frac{\lambda_{\min} - \sqrt{\lambda_{\min}^2 + 4\sigma_{\max}^2}}{2}, \frac{\lambda_{\max} - \sqrt{\lambda_{\max}^2 + 4\sigma_{\min}^2}}{2} \right]$$

$$I^+ = \left[\lambda_{\min}, \frac{\lambda_{\max} + \sqrt{\lambda_{\max}^2 + 4\sigma_{\max}^2}}{2} \right]$$

For our illustration: $\mathrm{Spec}(\mathcal{A})$ in $[-2, -2.4\,10^{-7}] \cup [1.7\,10^{-7}, 4.7]$ and $\kappa(\mathcal{A}) \leq 2.\,10^7$

Spectral properties of $\mathcal{P}_1^{-1}\mathcal{A}$

$$\mathcal{A} = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \text{ preconditioned by } \boxed{\mathcal{P}_1 = \begin{bmatrix} A & 0 \\ 0 & S_\gamma \end{bmatrix}}$$

With Rusten and Winther, applied on

$$\mathcal{P}_1^{-1/2}\mathcal{A}\mathcal{P}_1^{-1/2} = \begin{bmatrix} I_n & Q_1 \\ Q_1^T & 0 \end{bmatrix},$$

where $Q_1 = A^{-1/2}BS_\gamma^{-1/2}$ satisfies $Q_1^TQ_1 = S_\gamma^{-1/2}SS_\gamma^{-1/2} \sim S_\gamma^{-1}S$

$$\boxed{\Longrightarrow \ \lambda_i(I_n) = 1 \quad \text{and} \quad \sigma_i^2(Q_1) = \lambda_i(S_\gamma^{-1}S) = \nu_i}$$

Spectral properties of $\mathcal{P}_1^{-1}\mathcal{A}$ (Rusten and Winther)

The eigenvalues of $\mathcal{P}_1^{-1}\mathcal{A}$ are bounded within the intervals

$$\left[\frac{1-\sqrt{1+4\nu_{\max}}}{2}, \frac{1-\sqrt{1+4\nu_{\min}}}{2}\right] \bigcup \left[1, \frac{1+\sqrt{1+4\nu_{\max}}}{2}\right]$$

where

$$\nu_{\min} \text{ and } \nu_{\max} \in \left[\frac{\alpha}{\alpha+\lambda_{\max}(A)+\gamma}, \frac{\alpha+\gamma}{\gamma}\right]$$

(eigenvalues of $S_\gamma^{-1}S$)

$\longrightarrow$ In terms of $\lambda_{\max}(A), \gamma$ and α

Spectral properties of $\mathcal{P}_1^{-1}\mathcal{A}$ (direct proof)

The eigenvalues of $\mathcal{P}_1^{-1}\mathcal{A}$ are bounded within the (refined) intervals

$$\left[\frac{1-\sqrt{1+4\nu_{\text{max}}}}{2}, \frac{1-\sqrt{1+4\nu_{\text{min}}}}{2}\right] \bigcup \{1\} \bigcup \left[\frac{1+\sqrt{1+4\nu_{\text{min}}}}{2}, \frac{1+\sqrt{1+4\nu_{\text{max}}}}{2}\right]$$

where

$$\nu_{\text{min}} \text{ and } \nu_{\text{max}} \in \left[\frac{\alpha}{\alpha+\lambda_{\text{max}}(A)+\gamma}, \frac{\alpha+\gamma}{\gamma}\right]$$

(eigenvalues of $S_\gamma^{-1}S$)

$\longrightarrow$ In terms of $\lambda_{\text{max}}(A), \gamma$ and α

Spectral properties of $\mathcal{P}_2^{-1}\mathcal{A}$

$$\mathcal{A} = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix} \text{ preconditioned by } \boxed{\mathcal{P}_2 = \begin{bmatrix} A_\gamma & 0 \\ 0 & S_\gamma \end{bmatrix}}$$

With Rusten and Winther, applied on

$$\mathcal{P}_2^{-1/2}\mathcal{A}\mathcal{P}_2^{-1/2} = \begin{bmatrix} A_\gamma^{-1/2} A A_\gamma^{-1/2} & Q_2 \\ Q_2^T & 0 \end{bmatrix},$$

where $Q_2 = A_\gamma^{-1/2} B S_\gamma^{-1/2}$ satisfies $Q_2^T Q_2 = S_\gamma^{-1/2} S_\gamma S_\gamma^{-1/2} = I_m$

$$\Longrightarrow \ \lambda_i(A_\gamma^{-1/2} A A_\gamma^{-1/2}) = \lambda_i(A_\gamma^{-1} A) = \mu_i \quad \text{and} \quad \sigma_i^2(Q_2) = 1$$

Spectral properties of $\mathcal{P}_2^{-1}\mathcal{A}$ (Rusten and Winther)

The eigenvalues of $\mathcal{P}_2^{-1}\mathcal{A}$ are bounded within the intervals

$$\left[\frac{\mu_{\mathsf{min}} - \sqrt{\mu_{\mathsf{min}}^2 + 4}}{2}, \frac{\mu_{\mathsf{max}} - \sqrt{\mu_{\mathsf{max}}^2 + 4}}{2}\right] \bigcup \left[\mu_{\mathsf{min}}, \frac{\mu_{\mathsf{max}} + \sqrt{\mu_{\mathsf{max}}^2 + 4}}{2}\right]$$

where

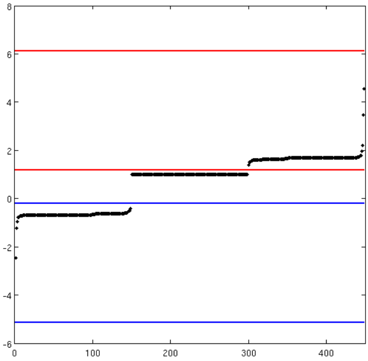
$$\mu_{\mathsf{min}} = \min\left(\frac{\alpha + \lambda_{\mathsf{min}}(A)}{\alpha}, \frac{\gamma}{\alpha}\right) \quad \text{and} \quad \mu_{\mathsf{max}} = \max\left(\frac{\alpha + \gamma}{\alpha}, \frac{\lambda_{\mathsf{max}}(A)}{\alpha}\right)$$

(eigenvalues of $A_\gamma^{-1}A$)

$\longrightarrow$ In terms of $\lambda_{\mathsf{min}}(A), \lambda_{\mathsf{max}}(A), \gamma$ and α

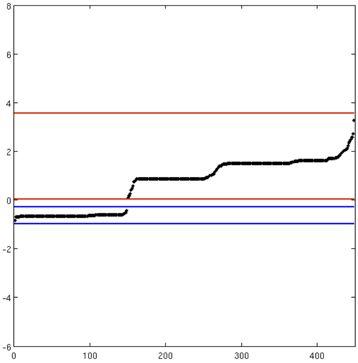
$\mathcal{P}_1^{-1}\mathcal{A}$ versus $\mathcal{P}_2^{-1}\mathcal{A}$ (Intervals in terms of $\lambda_{\max}(A), \lambda_{\min}(A), \gamma$ and α)

$[-5.13, -0.19] \cup \{1\} \cup [1.19, 6.13]$



$\boxed{\kappa(\mathcal{P}_1^{-1}\mathcal{A}) \leq 32.3}$

$[-0.98, -0.28] \cup [0.03, 3.56]$



$\boxed{\kappa(\mathcal{P}_2^{-1}\mathcal{A}) \leq 119}$

$\longrightarrow$ Reduced condition numbers in both cases

Varying γ ("True" intervals)

γ	$\lvert\{\lambda_i \le \gamma\}\rvert$	$\mathrm{Spec}(\mathcal{P}_1^{-1}\mathcal{A})$	$\kappa(\mathcal{P}_1^{-1}\mathcal{A})$
$\frac{\lambda_{\max}(A)}{100}$	42	$[-3.57, -0.40] \cup [1, 4.57]$	11.43
$\frac{\lambda_{\max}(A)}{1000}$	33	$[-12.12, -0.39] \cup [1, 13.12]$	33.64
$\frac{\lambda_{\max}(A)}{10000}$	23	$[-27.20, -0.38] \cup [1, 28.20]$	74.21

γ	$\lvert\{\lambda_i \le \gamma\}\rvert$	$\mathrm{Spec}(\mathcal{P}_2^{-1}\mathcal{A})$	$\kappa(\mathcal{P}_2^{-1}\mathcal{A})$
$\frac{\lambda_{\max}(A)}{100}$	42	$[-0.90, -0.45] \cup [9.7\,10^{-2}, 3.28]$	33.81
$\frac{\lambda_{\max}(A)}{1000}$	33	$[-0.98, -0.43] \cup [1.1\,10^{-2}, 3.40]$	309.09
$\frac{\lambda_{\max}(A)}{10000}$	23	$[-0.99, -0.42] \cup [2.5\,10^{-3}, 3.52]$	1408.00

Convergence bound of preconditioned MINRES

Assume that $\mathrm{Spec}(\mathcal{P}^{-1}\mathcal{A}) \subset [-a, -b] \cup [c, d]$, with $a, b, c, d > 0$

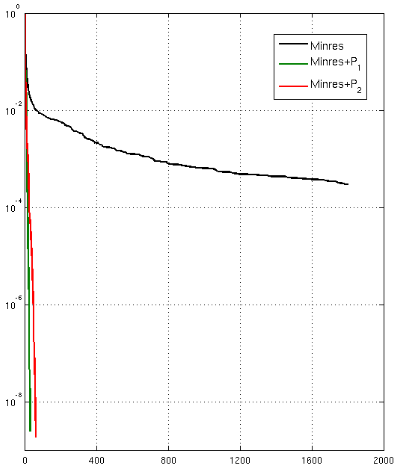
The iteration residual $r^{2k} = b - \mathcal{A}u^{2k}$ satisfies the bound

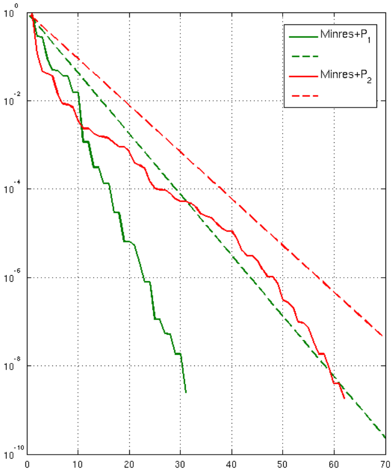
$$\frac{\|r^{2k}\|_{\mathcal{P}^{-1}}}{\|r^0\|_{\mathcal{P}^{-1}}} \leq 2 \left(\underbrace{\frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}}_{\to \text{ slope}} \right)^k$$

$\to$ Illustration on our example for various γ's

$$\left(\text{Stopping criterion in MINRES: } \frac{\|r^k\|_2}{\|r^0\|_2} \leq 10^{-8} \right)$$

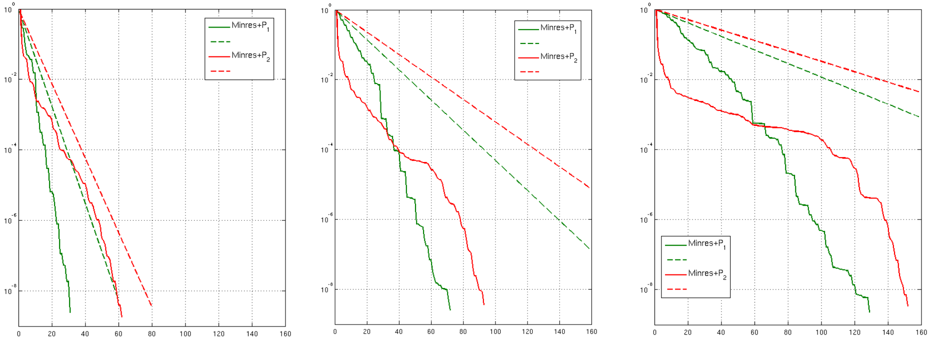
MINRES for $\gamma = \frac{\lambda_{max}(A)}{100}$





Convergence curves ($\mathcal{P}_1^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

MINRES for $\gamma = \frac{\lambda_{max}(A)}{100}$, $\gamma = \frac{\lambda_{max}(A)}{1000}$ and $\gamma = \frac{\lambda_{max}(A)}{10000}$



Convergence curves ($\mathcal{P}_1^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

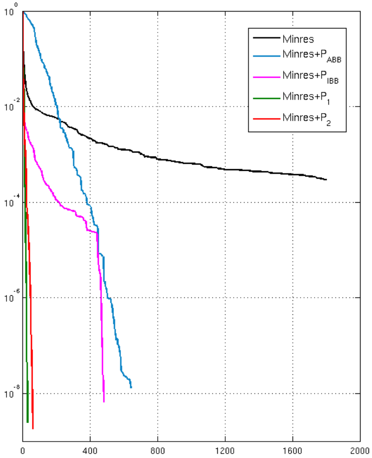
Further comparisons (“True” intervals, $\gamma = \frac{\lambda_{max}(A)}{100}$)

$$\mathcal{P}_1 = \begin{bmatrix} A & 0 \\ 0 & S_\gamma \end{bmatrix} \qquad \mathcal{P}_{IBB} = \begin{bmatrix} I_n & 0 \\ 0 & B^T B \end{bmatrix}$$

$$\mathcal{P}_2 = \begin{bmatrix} A_\gamma & 0 \\ 0 & S_\gamma \end{bmatrix} \qquad \mathcal{P}_{ABB} = \begin{bmatrix} A & 0 \\ 0 & B^T B \end{bmatrix}$$

$\mathcal{P}$	Spec($\mathcal{P}^{-1}\mathcal{A}$)			$\kappa(\mathcal{P}^{-1}\mathcal{A})$
	$[-1.6, -2.4\,10^{-6}]$	$\cup$	$[2.3\,10^{-5}, 3.8]$	$1.6\,10^{6}$
$\mathcal{P}_1$	$[-3.6, -0.4]$	$\cup$	$[1.0, 4.6]$	11.4
$\mathcal{P}_2$	$[-0.9, -0.5]$	$\cup$	$[9.7\,10^{-2}, 3.3]$	33.8
$\mathcal{P}_{IBB}$	$[-1.0, -0.4]$	$\cup$	$[3.0\,10^{-5}, 3.8]$	$1.3\,10^{5}$
$\mathcal{P}_{ABB}$	$[-8.9\,10^{2}, -0.4]$	$\cup$	$[1.0, 8.9\,10^{2}]$	$2.5\,10^{3}$

MINRES for $\gamma = \frac{\lambda_{max}(A)}{100}$



Convergence curves ($\mathcal{P}^{-1}$-norm of relative residuals)

First-level preconditioner

Preconditioner on A:
$M = R^T R$

Preconditioner on B:
$N = W^T W$

$$\Downarrow$$

$$\mathcal{P}_{Init} = \begin{bmatrix} M & 0 \\ 0 & N \end{bmatrix}$$

A $\mathcal{P}_2$-type preconditioner applied to $\mathcal{P}_{Init}^{-1}\mathcal{A}$ amounts to

$\mathcal{A} = \begin{bmatrix} A & B \\ B^T & 0 \end{bmatrix}$ preconditioned by $\tilde{\mathcal{P}}_2 = \begin{bmatrix} \tilde{A}_\gamma & 0 \\ 0 & \tilde{S}_\gamma \end{bmatrix}$

Two levels of preconditioner in one

where

$$\tilde{A}_\gamma^{-1} = \frac{1}{\alpha} M^{-1} + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

$$\tilde{S}_\gamma^{-1} = \alpha (B^T M^{-1} B)^{-\frac{1}{2}} (I_m - K(\tfrac{1}{\alpha}\Lambda_\gamma + K^T K)^{-1} K^T)(B^T M^{-1} B)^{-\frac{1}{2}}$$

$$K = (B^T M^{-1} B)^{-\frac{1}{2}} B^T U_\gamma$$

and

- $\lambda_{\min}(M^{-1}A) \leq \gamma \leq \lambda_{\max}(M^{-1}A)$
- $\Lambda_\gamma = \mathrm{diag}\{\lambda_i(M^{-1}A)\}_{i=1}^p$ with $\lambda_i \leq \gamma$
- $U_\gamma \in \mathbb{R}^{n \times p}$ is the set of associated orthonormal eigenvectors
- $\alpha > 0$ is a scaling parameter

Construction of $\tilde{A}_\gamma^{-1}$ and $\tilde{S}_\gamma^{-1}$

$$\tilde{A}_\gamma^{-1} = \frac{1}{\alpha} M^{-1} + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

$$\tilde{S}_\gamma^{-1} = \alpha (B^T M^{-1} B)^{-\frac{1}{2}} (I_m - K(\tfrac{1}{\alpha}\Lambda_\gamma + K^T K)^{-1} K^T)(B^T M^{-1} B)^{-\frac{1}{2}}$$

$$K = (B^T M^{-1} B)^{-\frac{1}{2}} B^T U_\gamma$$

1. Extract U_γ and Λ_γ:
   - Chebyshev-based Krylov method on $AU_\gamma = MU_\gamma\Lambda_\gamma$
2. $(B^T M^{-1} B)^{-1} v$

$$\hookrightarrow \begin{bmatrix} M & B \\ B^T & 0 \end{bmatrix} \quad \cdots$$

Interaction between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$

Recombination issue

$$S_\gamma^{-1} = \alpha(B^TB)^{-\frac{1}{2}}(I_m - K\underbrace{(\tfrac{1}{\alpha}\Lambda_\gamma + K^TK)^{-1}}_{\text{Key part}}K^T)(B^TB)^{-\frac{1}{2}}$$

where $K = (B^TB)^{-\frac{1}{2}}B^TU_\gamma = Q^TU_\gamma$, with

- Q: orthonormal basis of $\mathcal{I}m(B)$
- U_γ: orthonormal basis of $\mathcal{I}m(U_\gamma)$

$\longrightarrow$ The singular values of K correspond to the cosines of the principal angles between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$

Interaction between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$ (continued)

Consider the Singular Value Decomposition of $K = Q^T U_\gamma$:

$$Y^T K Z = \underbrace{Y^T Q^T}_{\substack{\text{Princ. vec.} \\ \text{in } \mathcal{I}m(B) \\ (p \text{ first lines})}} \quad \underbrace{U_\gamma Z}_{\substack{\text{Princ. vec.} \\ \text{in } \mathcal{I}m(U_\gamma)}} = \begin{array}{|c|} \hline C_\gamma \\ \hline 0 \\ \hline \end{array}$$

where

- $Y \in \mathbb{R}^{m \times m}$ and $Z \in \mathbb{R}^{p \times p}$ are orthogonal matrices
- $C_\gamma = \mathrm{diag}\{\cos\theta_i\}_{i=1}^p$ (that we assume nonsingular)
- $\{\theta_i\}_{i=1}^p \in \; [0, \pi/2[$ are the principal angles

Interaction between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$ (continued)

$$S_\gamma^{-1} = \alpha(B^T B)^{-\frac{1}{2}}(I_m - K\underbrace{(\tfrac{1}{\alpha}\Lambda_\gamma + K^T K)}_{\text{Key part}}{}^{-1}K^T)(B^T B)^{-\frac{1}{2}}$$

becomes, using $Y_\gamma = Y(:,1{:}p)$ and $V_\gamma = U_\gamma Z$ (princ. vec.),

$$S_\gamma^{-1} = \alpha(B^T B)^{-\frac{1}{2}}(I_m - Y_\gamma\underbrace{(\tfrac{1}{\alpha}C_\gamma^{-1}(V_\gamma^T A V_\gamma)C_\gamma^{-1} + I_p)}_{\text{Key part}}{}^{-1}Y_\gamma^T)(B^T B)^{-\frac{1}{2}}$$

where $V_\gamma^T A V_\gamma = Z^T U_\gamma^T A U_\gamma Z = Z^T \Lambda_\gamma Z$

Interaction between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$ (continued)

$$S_\gamma^{-1} = \alpha(B^TB)^{-\frac{1}{2}}(I_m - Y_\gamma\underbrace{(\tfrac{1}{\alpha}C_\gamma^{-1}(V_\gamma^TAV_\gamma)C_\gamma^{-1} + I_p)}_{\text{Key part}}{}^{-1}Y_\gamma^T)(B^TB)^{-\frac{1}{2}}$$

where $V_\gamma^TAV_\gamma = Z^TU_\gamma^TAU_\gamma Z = Z^T\Lambda_\gamma Z$

Observing that:

$$\frac{1}{\alpha}\|C_\gamma^{-1}(V_\gamma^TAV_\gamma)C_\gamma^{-1}\|_2 \leq \frac{1}{\alpha}\,\frac{\max\{\lambda_i\}_{i=1}^p}{\min\{\cos^2\theta_i\}_{i=1}^p} \leq \frac{1}{\alpha}\,\frac{\gamma}{\min\{\cos^2\theta_i\}_{i=1}^p},$$

roughly speaking, small eigenvalues "have an impact" on S_γ^{-1} when:

$$\boxed{\min\{\cos^2\theta_i\}_{i=1}^p \lesssim \mathcal{O}(\gamma/\alpha)}$$

Interaction between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$ (continued)

$$S_\gamma^{-1} = \alpha(B^TB)^{-\frac{1}{2}}(I_m - Y_\gamma\underbrace{(\tfrac{1}{\alpha}C_\gamma^{-1}(V_\gamma^T A V_\gamma)C_\gamma^{-1} + I_p)^{-1}}_{\text{Key part}}Y_\gamma^T)(B^TB)^{-\frac{1}{2}}$$

with $\displaystyle \frac{1}{\alpha}\|C_\gamma^{-1}(V_\gamma^T A V_\gamma)C_\gamma^{-1}\|_2 \leq \frac{1}{\alpha}\frac{\gamma}{\min\{\cos^2\theta_i\}_{i=1}^p}$

that is, if

$$\frac{1}{\alpha}\frac{\gamma}{\min\{\cos^2\theta_i\}_{i=1}^p} \ll 1, \quad \text{or equivalently,} \quad \min\{\cos^2\theta_i\}_{i=1}^p \gg \frac{\gamma}{\alpha},$$

the spectral information contained in $V_\gamma^T A V_\gamma$ is inhibited
($\rightarrow$ problem dependent)

Illustration on a toy example

- $A \in \mathbb{R}^{500\times 500}$ is a diagonal matrix with entries in $]0,1]$ and s.t.:
  - $p = 5 \quad (\gamma = 10^{-1}$ and $\alpha = 1)$
  - $\Lambda_\gamma = A(1{:}5,1{:}5)$
  - $U_\gamma = I_{500}(:,1{:}5)$
- $B \in \mathbb{R}^{500\times 200}$ is set to $B = \begin{bmatrix} C_\gamma & 0 \\ B_1 S_\gamma & B_2 \end{bmatrix}$, where
  - $C_\gamma = \mathrm{diag}\{\cos\theta_i\}_{i=1}^5$ and $S_\gamma = \mathrm{diag}\{\sin\theta_i\}_{i=1}^5$
  - $Q = \begin{bmatrix} B_1 & B_2 \end{bmatrix} \in \mathbb{R}^{495\times 200}$ with $Q^T Q = I_{200}$

Ensuring that

- B has othonormal columns
- $(B^T B)^{-\frac{1}{2}} B^T U_\gamma = B^T U_\gamma = \boxed{\begin{array}{c} C_\gamma \\ \hline 0 \end{array}}$

Illustration on a toy example (continued)

Implies a one to one match

- Eigenvectors versus principal vectors:

$$V_\gamma = U_\gamma$$

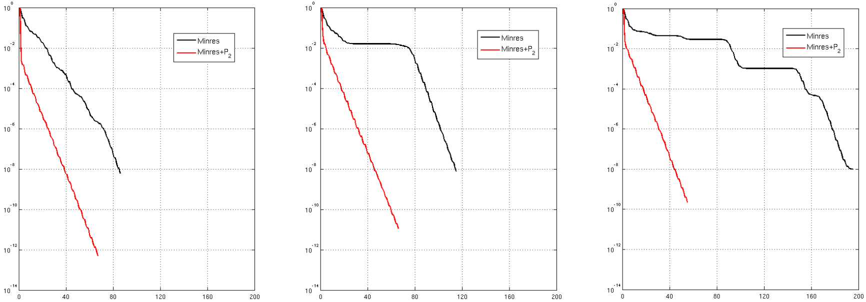
- Eigenvalues versus cosines of principal angles:

$$\text{Key part} = C_\gamma^{-1} \Lambda_\gamma C_\gamma^{-1} + I_5$$

Illustration on a toy example (continued)

$\Lambda_\gamma = \mathrm{diag}(10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1})$ and $C_\gamma = \mathrm{diag}\{\cos\theta_i\}_{i=1}^{5}$, where:

$$\begin{cases} \cos\theta_1 = 0.3 \\ \cos\theta_2 = 0.3 \\ \cos\theta_3 = 0.3 \\ \cos\theta_4 = 0.3 \\ \cos\theta_5 = 0.3 \end{cases} \qquad \begin{cases} \cos\theta_1 = 0.3 \\ \cos\theta_2 = 10^{-3} \\ \cos\theta_3 = 0.3 \\ \cos\theta_4 = 0.3 \\ \cos\theta_5 = 0.3 \end{cases} \qquad \begin{cases} \cos\theta_1 = 10^{-6} \\ \cos\theta_2 = 10^{-5} \\ \cos\theta_3 = 10^{-4} \\ \cos\theta_4 = 10^{-3} \\ \cos\theta_5 = 0.3 \end{cases}$$



Convergence curves (2-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

Interaction between $\mathcal{I}m(B)$ and $\mathcal{I}m(U_\gamma)$ (continued)

From these observations, we can extrapolate (bet?) that:

The bad conditioning contained in the smallest eigenvalues of A (if any)
will impact and spoil the convergence of MINRES
if (some of) the principal angles between $\mathcal{I}m(B)$
and the associated invariant subspace $\mathcal{I}m(U_\gamma)$, are close to $\pi/2$

because

the key part: $\frac{1}{\alpha} C_\gamma^{-1} (V_\gamma^T A V_\gamma) C_\gamma^{-1} + I_p$ reveals that
the square of the inverse of the cosines of these principal angles
push the corresponding bad conditioning of A
to "show up" in the Schur complement inverse

Short recap

$\Lambda_\gamma = \mathrm{diag}\{\lambda_i \leq \gamma\}_{i=1}^{p}$ and U_γ contains the orthonormal eigenvectors

$$\text{(SLRU)} \quad A_\gamma^{-1} = \frac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

$$S_\gamma^{-1} = B^T A_\gamma^{-1} B = \alpha (B^T B)^{-\frac{1}{2}} (I_m - \underbrace{K(\tfrac{1}{\alpha}\Lambda_\gamma + K^T K)^{-1} K^T}_{\text{Key part}}) (B^T B)^{-\frac{1}{2}}$$

where $K = (B^T B)^{-\frac{1}{2}} B^T U_\gamma = Q^T U_\gamma$ such that

$$\text{(SVD)} \quad Y^T K Z = \underbrace{Y^T Q^T}_{\substack{\text{Princ. vec.} \\ \text{in } \mathcal{I}m(B)}} \quad \underbrace{U_\gamma Z}_{\substack{\text{Princ. vec.} \\ \text{in } \mathcal{I}m(U_\gamma)}} = \begin{array}{|c|} \hline C_\gamma \\ \hline 0 \\ \hline \end{array}$$

$C_\gamma = \mathrm{diag}\{\cos\theta_i\}_{i=1}^{p}$ $(\{\theta_i\}_{i=1}^{p} \in [0, \pi/2[$ the principal angles$)$

Short recap (continued) and a question

$$S_\gamma^{-1} = \alpha(B^TB)^{-\frac{1}{2}}(I_m - Y_\gamma\underbrace{(\tfrac{1}{\alpha}C_\gamma^{-1}(V_\gamma^T A V_\gamma)C_\gamma^{-1} + I_p)^{-1}}_{\text{Key part}}Y_\gamma^T)(B^TB)^{-\frac{1}{2}}$$

where $V_\gamma = U_\gamma Z$ are the principal vectors in $\mathcal{I}m(U_\gamma)$ and where the smaller the associated cosines are, the "greater" the impact may be

A question:

Among the "available" information (Λ_γ, U_γ, V_γ and C_γ),
which one is the most relevant to accelerate the convergence of MINRES
between
a subset of smallest eigenvalues and their associated eigenvectors and
a subset of smallest cosines and their associated principal vectors in $\mathcal{I}m(U_\gamma)$?

"Smart" selection of principal vectors in $\mathcal{I}m(U_\gamma)$

$$S_\gamma^{-1} = \alpha(B^TB)^{-\frac{1}{2}}(I_m - Y_\gamma(\tfrac{1}{\alpha}C_\gamma^{-1}(V_\gamma^T A V_\gamma)C_\gamma^{-1} + I_p)^{-1}Y_\gamma^T)(B^TB)^{-\frac{1}{2}}$$

Select the principal vectors $V_\gamma = U_\gamma Z$ whose associated principal angles satisfy (let ℓ be the number of such vectors):

$$\cos^2\theta_i \le c\,\frac{\gamma}{\alpha}, \qquad c \in \left[\frac{1}{4}, 4\right]$$

i.e., select the appropriate columns in Z (yielding $Z_\theta \in \mathbb{R}^{p\times\ell}$), and "restrict" S_γ^{-1} accordingly:

$$S_\theta^{-1} = \alpha(B^TB)^{-\frac{1}{2}}(I_m - Y_\theta(\tfrac{1}{\alpha}C_\theta^{-1}(V_\theta^T A V_\theta)C_\theta^{-1} + I_\ell)^{-1}Y_\theta^T)(B^TB)^{-\frac{1}{2}}$$

where $V_\theta = U_\gamma Z_\theta \in \mathbb{R}^{n\times\ell}$, $C_\theta \in \mathbb{R}^{\ell\times\ell}$ and $Y_\theta \in \mathbb{R}^{m\times\ell}$

A "smarter" $\mathcal{P}_2$

In practice, this amounts to use the preconditioner:

$$\mathcal{P}_\theta = \begin{bmatrix} A_\theta & 0 \\ 0 & S_\theta \end{bmatrix}$$

where

$$A_\theta^{-1} = \frac{1}{\alpha} I_n + V_\theta (V_\theta^T A V_\theta)^{-1} V_\theta^T$$

$$S_\theta^{-1} = \alpha (B^T B)^{-\frac{1}{2}} \left(I_m - K_\theta \left(\frac{1}{\alpha} V_\theta^T A V_\theta + K_\theta^T K_\theta\right)^{-1} K_\theta^T \right) (B^T B)^{-\frac{1}{2}}$$

with $K_\theta = (B^T B)^{-\frac{1}{2}} B^T V_\theta$

$\mathcal{P}_2$ versus $\mathcal{P}_\theta$

$$\mathcal{P}_2 = \begin{bmatrix} A_\gamma & 0 \\ 0 & S_\gamma \end{bmatrix} \qquad \mathcal{P}_\theta = \begin{bmatrix} A_\theta & 0 \\ 0 & S_\theta \end{bmatrix}$$

with

$$A_\gamma^{-1} = \frac{1}{\alpha} I_n + U_\gamma \Lambda_\gamma^{-1} U_\gamma^T$$

$$S_\gamma^{-1} = \alpha (B^T B)^{-\frac{1}{2}} \left(I_m - K(\tfrac{1}{\alpha}\Lambda_\gamma + K^T K)^{-1} K^T \right) (B^T B)^{-\frac{1}{2}}$$

$$K = (B^T B)^{-\frac{1}{2}} B^T U_\gamma$$

$$A_\theta^{-1} = \frac{1}{\alpha} I_n + V_\theta (V_\theta^T A V_\theta)^{-1} V_\theta^T$$

$$S_\theta^{-1} = \alpha (B^T B)^{-\frac{1}{2}} \left(I_m - K_\theta(\tfrac{1}{\alpha} V_\theta^T A V_\theta + K_\theta^T K_\theta)^{-1} K_\theta^T \right) (B^T B)^{-\frac{1}{2}}$$

$$K_\theta = (B^T B)^{-\frac{1}{2}} B^T V_\theta$$

$\mathcal{P}_2$ versus $\mathcal{P}_\theta$ on our example

- $\gamma = \frac{\lambda_{max}(A)}{100} \approx 3.8\ 10^{-2}$ $\longrightarrow$ $p = 42$
- $\cos^2\theta_i \leq 4\,\frac{\gamma}{\alpha}$ $(c = 4)$ $\longrightarrow$ $\ell = 22$

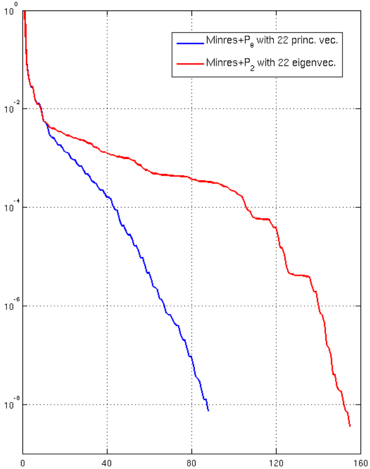
Comparison on Minres preconditioned by:

$\mathcal{P}_2$ built with $\ell = 22$ eigenvectors
corresponding to the ℓ smallest eigenvalues λ_i's
(among the p ones below γ)

versus

$\mathcal{P}_\theta$ built with $\ell = 22$ principal vectors
corresponding to the ℓ smallest $\cos\theta_i$'s
(among the p ones)

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ ($\ell = 22$)



Convergence curves ($\mathcal{P}_\theta^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ (for various ℓ)

Comparison on Minres preconditioned by:

$\mathcal{P}_2$ built with ℓ eigenvectors
corresponding to the ℓ smallest eigenvalues λ_i's
(among the p ones below γ)

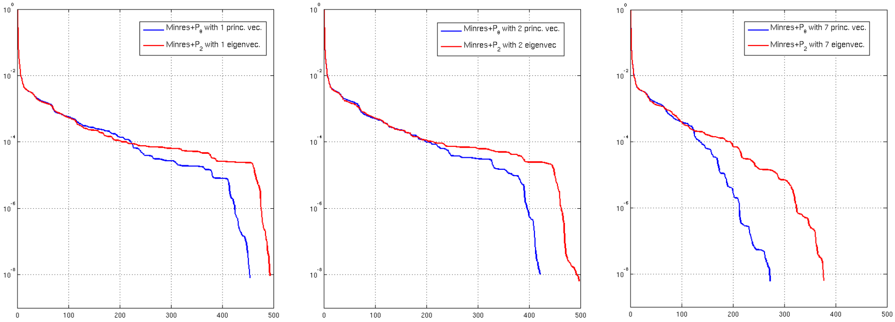
versus

$\mathcal{P}_\theta$ built with ℓ principal vectors
corresponding to the ℓ smallest $\cos\theta_i$'s
(among the p ones)

where

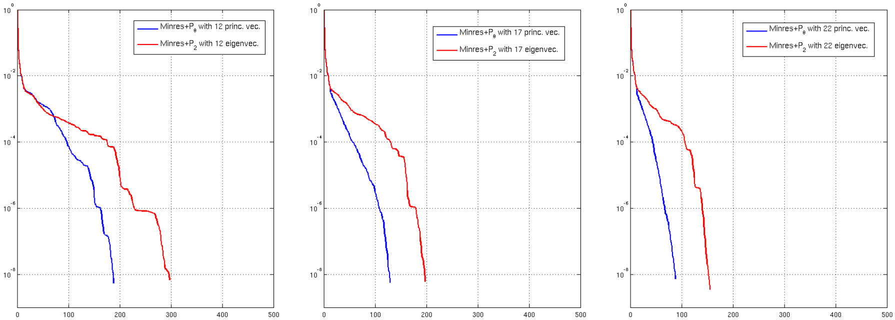
$$\ell = 1, 2, 7, 12, 17, 22, 27, 32, 37, 42$$

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ (for $\ell = 1, 2, 7$)



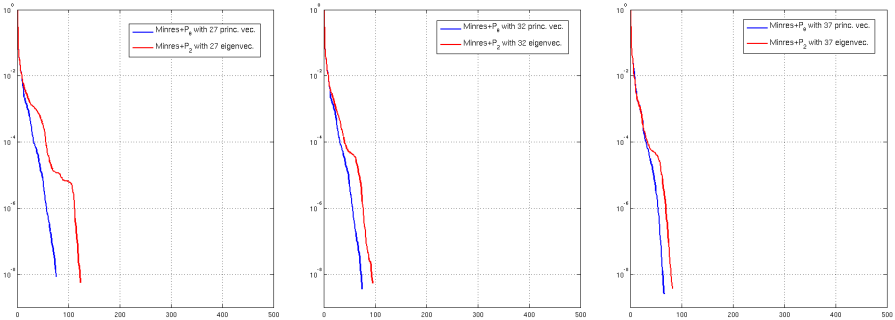
Convergence curves ($\mathcal{P}_\theta^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ (for $\ell = 12, 17, 22$)



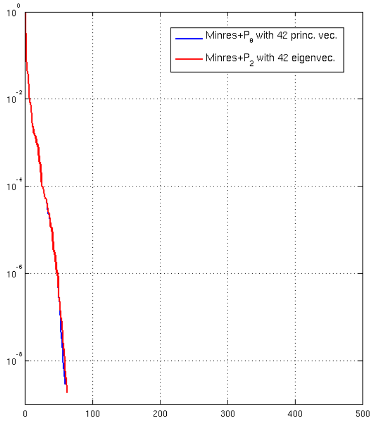
Convergence curves ($\mathcal{P}_\theta^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ (for $\ell = 27, 32, 37$)



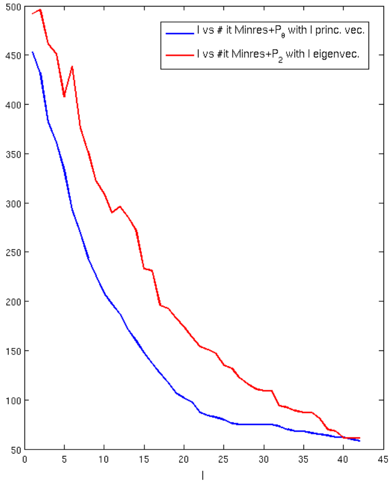
Convergence curves ($\mathcal{P}_\theta^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ (for $\ell = 42$)

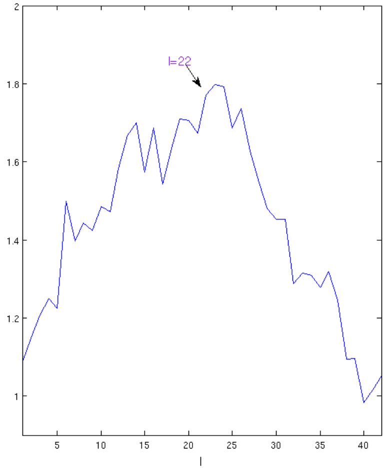


Convergence curves ($\mathcal{P}_\theta^{-1}$-norm and $\mathcal{P}_2^{-1}$-norm of relative residuals)

MINRES preconditioned using $\mathcal{P}_2$ versus $\mathcal{P}_\theta$ (overall picture)



ℓ versus # MINRES it.



ℓ versus $\dfrac{\#\text{MINRES } +\mathcal{P}_2 \text{ it.}}{\#\text{MINRES } +\mathcal{P}_\theta \text{ it.}}$

Conclusion

- Address the (bad) conditioning of A and that of B separately

  $\rightarrow$ Using a low-rank spectral approximation

- Recombine this spectral information appropriately to build an efficient block diagonal preconditioner

  $\rightarrow$ Through the Schur complement approximation

- Get some insight on the recombination issue between A and B

  $\rightarrow$ Through some analytic dissection ☺ and illustrations

Perspectives

- Look deeper into the recombination issue
- Use a similar spectral approach to address the bad conditioning of B
- Analyse the cost and the amortization (multiple r.h.s.)
- Derive practical and efficient implementations
- Perform numerical experiments on academic problems and applications

Thank you for your attention!

Construction of $\tilde{S}_\gamma^{-1}$

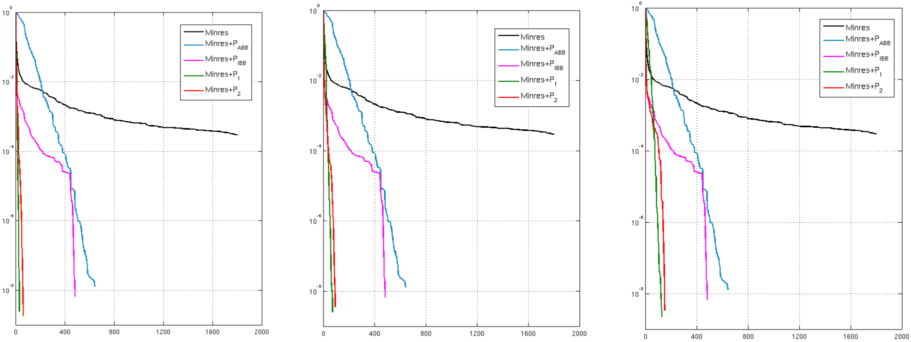
How to deal with $\begin{bmatrix} M & B \\ B^T & 0 \end{bmatrix}$?

1. Using Schilders' factorization, see Dollar and Wathen (2006)

2. Using a preconditioner of the form $\begin{bmatrix} D^{-1} & 0 \\ 0 & (B^T D^{-1} B)^{-1} \end{bmatrix}$ where D is diagonal, see Golub, Greif and Varah (2006)

3. Using a similar spectral approach on $B^T M^{-1} B$. How can we include efficiently spectral information in the Schur complement? $\Rightarrow$ We work on it !

MINRES for $\gamma = \frac{\lambda_{max}(A)}{100}$, $\gamma = \frac{\lambda_{max}(A)}{1000}$ and $\gamma = \frac{\lambda_{max}(A)}{10000}$



Convergence curves ($\mathcal{P}^{-1}$-norm of relative residuals)

Spectrum of $\mathcal{P}_{IBB}$ and $\mathcal{P}_{ABB}$