Limited Memory Preconditioners for symmetric indefinite systems. Application to saddle point problems

Sylvain Mercier

EDF R&D, CERFACS

September 13th 2013

Serge Gratton, ENSEEIHT et CERFACS
Xavier Vasseur, CERFACS

Nicolas Tardieu, EDF R&D

Sommaire

Problem to solve

Saddle point systems (fluid mechanics, optimisation...)

$$Ax = b \Longleftrightarrow \begin{pmatrix} K & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}$$

$K \in \mathbb{R}^{n\times n}$ symmetric, $B \in \mathbb{R}^{m\times n}$, $m \leq n$ and
$C \in \mathbb{R}^{m\times m}$ symmetric positive semidefinite.

Problem to solve

Saddle point systems (fluid mechanics, optimisation...)

$$Ax = b \Longleftrightarrow \begin{pmatrix} K & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}$$

$K \in \mathbb{R}^{n \times n}$ symmetric, $B \in \mathbb{R}^{m \times n}$, $m \leq n$ and
$C \in \mathbb{R}^{m \times m}$ symmetric positive semidefinite.

Nonlinear problems in solid mechanics

- Dualizing Dirichlet boundary conditions with "double Lagrange" principle
- Newton-like method $\Longrightarrow$ sequence of linear systems to solve:

$$A_i x_i = b_i \text{ avec } A_i = \begin{pmatrix} K_i & B^T & B^T \\ B & -\alpha I & \alpha I \\ B & \alpha I & -\alpha I \end{pmatrix}, \quad \alpha \in \mathbb{R}^{+*}$$

Preconditioning [Benzi et al.,2005]

$$Ax = b \Longleftrightarrow \begin{pmatrix} K & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}$$

Krylov subspace methods $\Rightarrow$ Preconditioning

Examples:

- $P_{c,d} = \begin{pmatrix} I & 0 \\ cBK_0^{-1} & I \end{pmatrix} \begin{pmatrix} K_0 & 0 \\ 0 & S_0 \end{pmatrix} \begin{pmatrix} I & dK_0B^T \\ 0 & I \end{pmatrix}$,
  $K_0 \simeq K$ et $S_0 \simeq -C - BK^{-1}B^T$

  $\Rightarrow$ block diagonal, block triangular, constraint
- Inexact constraint preconditioners ($B_0 \simeq B$)
- Multigrid, domain decomposition
- ...

Objective: A class of Limited Memory Preconditioners (LMP) is available in a symmetric positive definite framework.

The aim is to generalize this class to symmetric **indefinite** systems.

Sommaire

Objective

- **Framework**: Solving a sequence of linear systems $A_i x_i = b_i$, with A_i SPD, using the Conjugate Gradient (CG) method.
- A first SPD preconditioner is available (*first-level*)
- It is possible to compute (cheaply) some directions slowing down the convergence of the CG method (e.g. Ritz, Harmonic Ritz...)
- **Goal**: Construct a *second-level* preconditioner with these directions to speed up the CG convergence.

Motivation

A^{-1} is the inverse Hessian of $q(x) = \frac{1}{2}x^T A x - b^T x$
The idea is to use the BFGS Hessian approximation:

$$H_i = (I_n - \frac{y_i s_i^T}{y_i^T s_i})^T H_{i-1} (I_n - \frac{y_i s_i^T}{y_i^T s_i}) + \frac{s_i s_i^T}{y_i^T s_i}$$

Motivation

A^{-1} is the inverse Hessian of $q(x) = \frac{1}{2}x^T Ax - b^T x$
The idea is to use the BFGS Hessian approximation:

$$H_i = (I_n - \frac{y_i s_i^T}{y_i^T s_i})^T H_{i-1}(I_n - \frac{y_i s_i^T}{y_i^T s_i}) + \frac{s_i s_i^T}{y_i^T s_i}$$

LMP definition:

$$H = (I_n - S(S^T AS)^{-1} S^T A)M(I_n - AS(S^T AS)^{-1} S^T) + S(S^T AS)^{-1} S^T$$

with k linearly independent column vectors in S, and M SPD.

Properties

First property

- H is symmetric

Properties

First property

- H is symmetric

Spectral properties:

- $HAS = S$, namely the k vectors of S are eigenvectors of HA associated to 1.
- H is SPD and the spectrum of HA is contracted relatively to the MA one. [Gratton et al., Th. 3.4, 2011]

Properties

First property

- H is symmetric

Spectral properties:

- $HAS = S$, namely the k vectors of S are eigenvectors of HA associated to 1.
- H is SPD and the spectrum of HA is contracted relatively to the MA one. [Gratton et al., Th. 3.4, 2011]

Invariance property:

- Let $Z = SX$ where X is a nonsingular matrix of order k. Then H with Z instead of S remains unchanged. [Gratton et al., th. 3.1, 2011]

  $\Rightarrow$ Construction of an A-conjugate basis of $range(S)$ such as:

$$H = (I_n - ZZ^T A)M(I_n - AZZ^T) + ZZ^T$$

Arnoldi/Lanczos, general case

The Arnoldi process constructs an orthonormal basis of the Krylov subspace $\mathcal{K}_m = \mathcal{K}_m(A, r_0)$ (Lanczos if A sym.)

$$AV_m = V_{m+1}H_{m+1,m} = V_m H_m + h_{m+1,m} v_{m+1} e_m^T$$

$H_{m+1,m}$ is an Hessenberg matrix
H_m is tridiagonal in the symmetric case

Remark: when $m = n$, $A = V_n H_n V_n^T$, A and H_n have same spectra.

Ritz/Harmonic Ritz

We search an approximation (w_i, β_i) of an eigenpair such as

$$\begin{cases} w_i \in \mathcal{K}_m \\ Aw_i - \beta_i w_i \perp \mathcal{L}_m \end{cases}$$

Ritz/Harmonic Ritz

We search an approximation (w_i, β_i) of an eigenpair such as

$$\begin{cases} w_i \in \mathcal{K}_m \\ Aw_i - \beta_i w_i \perp \mathcal{L}_m \end{cases}$$

- **Ritz**: $\mathcal{L}_m = \mathcal{K}_m$, $(w_i, \beta_i) = (V_m y_i, \beta_i)$

  with (y_i, β_i) eigenpair of H_m

Ritz/Harmonic Ritz

We search an approximation (w_i, β_i) of an eigenpair such as

$$\begin{cases} w_i \in \mathcal{K}_m \\ Aw_i - \beta_i w_i \perp \mathcal{L}_m \end{cases}$$

- **Ritz**: $\mathcal{L}_m = \mathcal{K}_m$, $(w_i, \beta_i) = (V_m y_i, \beta_i)$

  with (y_i, β_i) eigenpair of H_m

- **Harmonic Ritz**: $\mathcal{L}_m = A\mathcal{K}_m$, $(w_i, \beta_i) = (V_m y_i, \beta_i)$

  with (y_i, β_i) eigenpair of $(H_m + (H_m^T)^{-1} e_m h_{m+1,m}^2 e_m^T)$

Sommaire

Notation

Formula:

$$H = (I_n - S(S^T AS)^{-1} S^T A) M (I_n - AS(S^T AS)^{-1} S^T) + S(S^T AS)^{-1} S^T$$

Notation:
It is possible to use $M^{\frac{1}{2}} A M^{\frac{1}{2}}$ as matrix and I_n as *first-level* preconditioner. Subsequently, we will refer to the LMP H as follows:

$$H = (I_n - S(S^T KS)^{-1} S^T K) H_0 (I_n - KS(S^T KS)^{-1} S^T) + S(S^T KS)^{-1} S^T$$

Sign matrix

Problem: When K is symmetric **indefinite**, there is no spectrum contraction.

Sign matrix

Problem: When K is symmetric **indefinite**, there is no spectrum contraction.

Idea: Decouple H to work separately in the positive (I^+) and negative (I^-) invariant subspaces of K.

Sign matrix

Problem: When K is symmetric **indefinite**, there is no spectrum contraction.

Idea: Decouple H to work separately in the positive (I^+) and negative (I^-) invariant subspaces of K.

How to: project a part of S in I^+ and the other in I^-. To do this, we will use the sign matrix of K, denoted X_K, and the projectors:

$$P_+ = \frac{I_n + X_K}{2} \quad \text{and} \quad P_- = \frac{I_n - X_K}{2}$$

Definition of the preconditioner

Notation

$$S = [S_+, S_-] \text{ with } S_+ \in I^+ \text{ et } S_- \in I^-$$

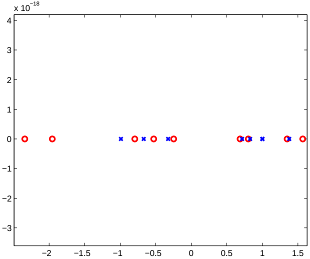
Decoupling

$$\begin{aligned} H = \; & (I_n - [S_+, S_-] \begin{pmatrix} S_+^T K S_+ & 0 \\ 0 & S_-^T K S_- \end{pmatrix}^{-1} [S_+, S_-]^T K) \times H_0 \times \\ & (I_n - K[S_+, S_-] \begin{pmatrix} S_+^T K S_+ & 0 \\ 0 & S_-^T K S_- \end{pmatrix}^{-1} [S_+, S_-]^T) + \\ & [S_+, S_-] \begin{pmatrix} S_+^T K S_+ & 0 \\ 0 & S_-^T K S_- \end{pmatrix}^{-1} [S_+, S_-]^T \end{aligned}$$

Theoretical results

Contraction on each side of the real axis

$\mathcal{P}_I$ property



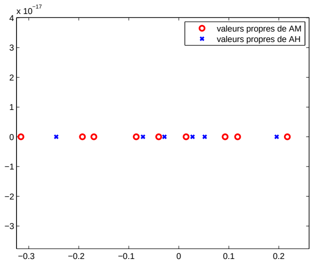
legend: $\sigma(KH_0)$ in red and $\sigma(KH)$ in blue

Assumptions: H_0 **symmetric positive definite** and $KH_0 = H_0K$

Theoretical results

Global contraction

$\mathcal{P}_g$ property



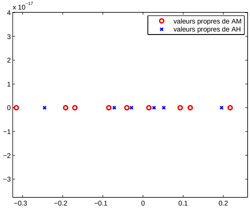
legend: $\sigma(KH_0)$ in red and $\sigma(KH)$ in blue

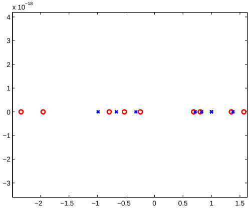
Assumptions: H_0 **symmetric indefinite** and $KH_0 = H_0 K$

Theoretical results

Spectral contraction

$\mathcal{P}_g$ property (orange ●) | $\mathcal{P}_l$ property (green ●)





	$K = A, H_0 = M$	$K = AM, H_0 = I$	$K = M^{\frac{1}{2}} A M^{\frac{1}{2}}, H_0 = I$
M SPD, $MA = AM$	green ●	green ●	green ●
M NPD, $MA = AM$	orange ●	green ●	×
M SPD, $MA \neq AM$	red ●	red ●	green ●

SQMR

QMR method [Freund, 1992]:

Let the the system $Ax = b$, with A nonsymmetric.

Nonsymmetric Lanczos:
$AV_m = V_{m+1}T_m$ with V_m **non-orthonormal basis** of $\mathcal{K}_m(A, r_0)$

At each iteration, we search y such as $x_m = x_0 + V_m y$

$$b - Ax_m = V_{m+1}(\|r_0\|e_1 - T_m y)$$

$$\text{QMR} \;\Rightarrow y = \text{argmin}\|\|r_0\|e_1 - T_m y\|$$

SQMR

SQMR method [Freund, 1995]:
If we know J nonsingular such as $A^T J = JA$, the Lanczos algorithm is simplified.

Application: Right preconditioning with H

$$Ax = b \Leftrightarrow AH\tilde{x} = b, \;\; x = H\tilde{x}$$

If A is symmetric, $H(AH) = (AH)^T H$ and we take $J = H$

Interest: **symmetric preconditioner** , contrary to MINRES or SYMMLQ method (requiring a SPD preconditioner)

Application of the preconditioner, computational cost

Construction of H

- Computational cost for the projection of S thanks to the sign matrix
- Construction of Z_+ (resp. Z_-) K-conjugate basis (resp. $-K$) of $range(S_+)$ (resp. $range(S_-)$)
  computational cost: $2n(k_+^2 + k_-^2)$ where $k = k_+ + k_-$, and k matrix-vector products by K

Application of the preconditioner, computational cost

Construction of H

- Computational cost for the projection of S thanks to the sign matrix
- Construction of Z_+ (resp. Z_-) K-conjugate basis (resp. $-K$) of $range(S_+)$ (resp. $range(S_-)$)
  computational cost: $2n(k_+^2 + k_-^2)$ where $k = k_+ + k_-$, and k matrix-vector products by K

Application of H

Let $Z = [Z_+, Z_-]$, $\overline{Z} = [Z_+, -Z_-]$ and $Y = [KZ_+, -KZ_-]$:

$$Hq = (I_n - ZY^T)H_0(I_n - YZ^T)q + Z\overline{Z}q$$

Computational cost: $10kn$ and one matrix-vector product by H_0
Storage: $2k$ vectors of length n

Main steps

Framework: M SPD, $K = M^{\frac{1}{2}} A M^{\frac{1}{2}}$ and $H_0 = I$.

Algorithm:

Main steps

Framework: M SPD, $K = M^{\frac{1}{2}} A M^{\frac{1}{2}}$ and $H_0 = I$.

Algorithm:

- Several iterations of GMRES with the matrix K
  $\Rightarrow$ information collected to construct S

Main steps

Framework: M SPD, $K = M^{\frac{1}{2}} A M^{\frac{1}{2}}$ and $H_0 = I$.

Algorithm:

- Several iterations of GMRES with the matrix K
  $\Rightarrow$ information collected to construct S
- Projection of S in I^+ and I^- (exact sign of K)

Main steps

Framework: M SPD, $K = M^{\frac{1}{2}} A M^{\frac{1}{2}}$ and $H_0 = I$.

Algorithm:

- Several iterations of GMRES with the matrix K
  $\Rightarrow$ information collected to construct S
- Projection of S in I^+ and I^- (exact sign of K)
- Construction of $Z = [Z_+, Z_-]$ and $Y = [KZ_+, -KZ_-]$

Main steps

Framework: M SPD, $K = M^{\frac{1}{2}} A M^{\frac{1}{2}}$ and $H_0 = I$.

Algorithm:

- Several iterations of GMRES with the matrix K
  $\Rightarrow$ information collected to construct S
- Projection of S in I^+ and I^- (exact sign of K)
- Construction of $Z = [Z_+, Z_-]$ and $Y = [KZ_+, -KZ_-]$
- SQMR or GMRES for K right preconditioned by H

Main steps

Framework: M SPD, $K = M^{\frac{1}{2}} A M^{\frac{1}{2}}$ and $H_0 = I$.

Algorithm:

- Several iterations of GMRES with the matrix K
  $\Rightarrow$ information collected to construct S
- Projection of S in I^+ and I^- (exact sign of K)
- Construction of $Z = [Z_+, Z_-]$ and $Y = [KZ_+, -KZ_-]$
- SQMR or GMRES for K right preconditioned by H

Remark: It is possible to reuse directly H in a sequence framework (reduction of the computational cost)

Sommaire

Academic test problem

- Stationary Navier-Stokes problem: "Lid driven cavity" in a square domain (from IFISS library)
- $Q_2 - Q_1$ finite element
- Picard method to linearize the problem

Sequence: $$\begin{pmatrix} K + N_i & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u_i \\ p_i \end{pmatrix} = \begin{pmatrix} f_i \\ g_i \end{pmatrix}$$

Stokes with multiple right-hand sides, $n = 2467$

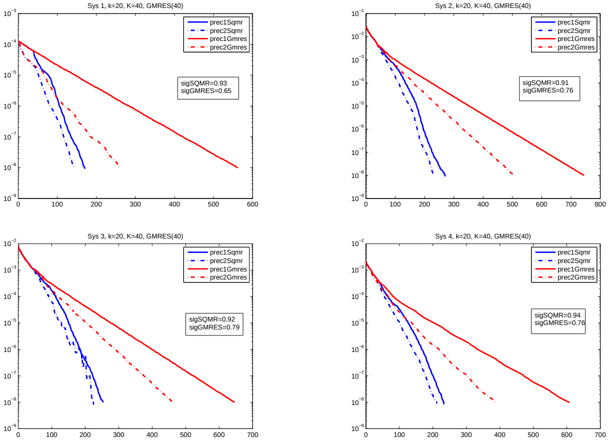
Sequence: $Ax_i = b_i \Leftrightarrow \begin{pmatrix} K & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u_i \\ p_i \end{pmatrix} = \begin{pmatrix} f_i - N_i u_{i-1} \\ g_i \end{pmatrix}$

Characteristics:

- $n = 2467$, 4 systems to solve.
- $M = LL^T$ Incomplete Cholesky factorization of $\begin{pmatrix} K & 0 \\ 0 & Q \end{pmatrix}$ with Q the pressure mass matrix (drop tolerance 10^{-2}).
- H is "built" during the solution of the 1st system. It is also applied to the 3 other systems.
- S contains k Ritz vectors (projected) corresponding to the k Ritz values with smallest modulus.
- The sign matrix of $L^{-1}AL^{-T}$ is exact (with signm.m from the matrix function toolbox by N.Higham)

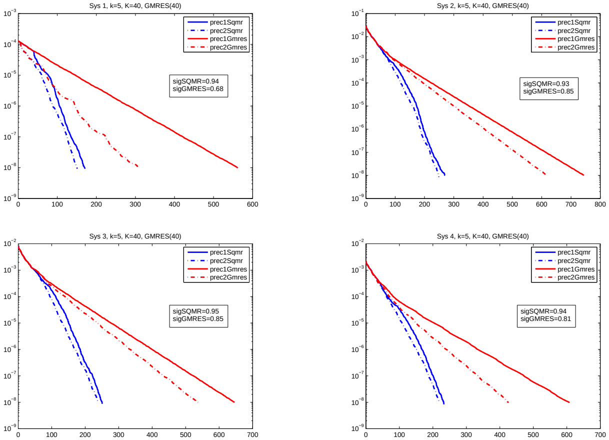
Stokes with multiple right-hand sides, $n = 2467$

Convergence curves of relative residual $\frac{\|b - Ax_i\|_2}{\|b\|_2}$ with $k = 20$ ($K = 40$):



Stokes with multiple right-hand sides, $n = 2467$

Convergence curves of relative residual $\frac{\|b - Ax_i\|_2}{\|b\|_2}$ with $k = 5$ ($K = 40$):



Sommaire

Prospectives

- Application of the sign matrix on a block of vectors (approximate sign matrix or approximate projection)

Prospectives

- Application of the sign matrix on a block of vectors (approximate sign matrix or approximate projection)
- Application in solid mechanics
  - Find a *first-level* symmetric indefinite preconditioner verifying the assumptions of spectral contraction
  - Update the preconditioner in a sequence with changing matrices

Prospectives

- Application of the sign matrix on a block of vectors (approximate sign matrix or approximate projection)
- Application in solid mechanics
  - Find a *first-level* symmetric indefinite preconditioner verifying the assumptions of spectral contraction
  - Update the preconditioner in a sequence with changing matrices
- Generalization to the nonsymmetric case

Prospectives

- Application of the sign matrix on a block of vectors (approximate sign matrix or approximate projection)
- Application in solid mechanics
  - Find a *first-level* symmetric indefinite preconditioner verifying the assumptions of spectral contraction
  - Update the preconditioner in a sequence with changing matrices
- Generalization to the nonsymmetric case
- Non-standard inner product

Thank you for your attention!

Bibliography I

M. Benzi, G.H. Golub, J.Liesen.
Numerical Solution of saddle point problems.
Acta Numerica, 1-137, 2005.

P. Krzyzanowski.
On block preconditioners for saddle point problems with singular or indefinite (1,1) block.
Numerical Linear Algebra with Applications 18: 123-140, 2011.

M.F. Murphy, G.H. Golub, A.J. Wathen.
A note on preconditioning for indefinite linear systems.
SIAM J. Sci. Comput., 21-6: 1969-1972, 2000.

A. Klawonn.
An optimal preconditioner for a class of saddle point problems with a penalty term.
SIAM J. Sci. Comput. 19-2: 540-552, 1998.

H.S. Dollar, N.I.M. Gould, M. Stoll, A.J. Wathen.
Preconditioning saddle point systems with applications in optimization.
SIAM J. Sci. Comput. 32-1: 249-270, 2012.

C. Siefert, E. de Sturler.
Preconditioners for generalized saddle point problems.
SIAM J. Numer. Anal. 44-3: 1275-1296, 2006.

Bibliography II

M. Benzi, V. Simoncini.
On the eigenvalues of a class of saddle point matrices.
Numerische Mathematik 103: 173-196, 2006.

L. Bergamaschi, J. Gondzio, M. Venturin, G. Zilli.
Inexact constraint preconditioners for linear systems arising in interior point methods.
Computational Optimization and Applications, 36: 137-147, 2007.

J. Pestana, A.J. Wathen.
Combination preconditioning of saddle point systems for positive definiteness.
Numerical Linear Algebra with Applications, 2012.

M. Stoll, A.J. Wathen.
Combination preconditioning and the Bramble-Pasciak^{+} preconditioner.
SIAM J. Matrix Anal. Appl. 30-2: 582-608, 2008.

S. Bellavia, D. Bertaccini, B. Morini.
Nonsymmetric preconditioner updates in Newton-Krylov methods for nonlinear systems.
SIAM J. Sci. Comput. 33-5: 2595-2619, 2011.

Bibliography III

S. Gratton, A. Sartenaer, J. Tshimanga.
On a class of limited memory preconditioners for large scale linear systems with multiple right-hand sides.
SIAM J. Optim. 21-3: 912-935, 2011.

J. Tshimanga.
On a class of limited memory preconditioners for large-scale nonlinear least-squares problems.
PhD Thesis, 2007.

G. Fasano, M. Roma.
AINVK: a class of approximate inverse preconditioners based on Krylov subspace methods, for large indefinite linear systems.
Preprint, 2012.

Y. Saad.
Iterative methods for sparse linear systems - second edition.
SIAM, Philadelphia, PA, 2003.

N.J. Higham.
Functions of matrices: theory and computation.
SIAM, chap. 5, 2008.

Bibliography IV

A. Frommer, K. Kahl, T. Lippert, H. Rittich.
Error bounds for the sign function.
IMACM, 2011.

A. Frommer, K. Kahl, T. Lippert, H. Rittich.
2-norm error bounds and estimates for Lanczos approximations to linear systems and rational matrix functions.
SIAM J. Matrix Anal. Appl. 34-3: 1046-1065, 2013.

J.C.R. Bloch, S.Heybrock.
A nested Krylov subspace method to compute the sign function of lare complex matrices.
Computer Physics Communications 182: 878-889, 2011.

R.W. Freund, N.M. Nachtigal.
Software for simplified Lanczos and QMR algorithms.
Applied Numerical Mathematics 19: 319-341, 1995.

R.W. Freund, T. Szeto.
A quasi minimal residual squared algorithm for non hermitian linear systems.
.

H. Elman, D. Silvester, A. Wathen.
Finite elements and fast iterative solvers with applications in incompressible fluid dynamics.
Numerical Mathematics and Scientific Computation, 2005.

H. Elman, A. Ramage, D. Silvester.
IFISS, a matlab toolbox for modelling incompressible flow.
ACM Transactions on Mathematical Software 33-2, 2007.

E. de Sturler,J. Liesen.
Block-diagonal and constraint preconditioners for nonsymmetric indefinite linear systems. Part I: theory.
SIAM J. Sci. Comput. 26-5: 1598-1619, 2005.