Iterative solvers for large sparse linear systems

Jocelyne Erhel - ALADIN group -

INRIA-RENNES

ERCIM WG meeting - May 2001

Séminaire du 3eme Cycle Romand d'Informatique

Krylov iterative methods

- Symmetric positive definite
- Symmetric indefinite
- Non symmetric

A symmetric positive definite

One method of choice : Conjugate Gradient (CG)

- algorithm
- properties
- convergence
- preconditioning

Conjugate Gradient

Algorithm

Initialisation

choose x_0

$p_0 = r_0 = b - Ax_0$

For $k = 0, 1 \ldots$

$\alpha_k = \frac{\|r_k\|^2}{(Ap_k, p_k)}$

$x_{k+1} = x_k + \alpha_k p_k$

$r_{k+1} = r_k - \alpha_k A p_k$

$\beta_{k+1} = \frac{\|r_{k+1}\|^2}{\|r_k\|^2}$

$p_{k+1} = r_{k+1} + \beta_{k+1} p_k$

End For

Properties

$(r_{k+1}, p_k) = 0$

$(r_{k+1}, r_k) = 0$

$(p_{k+1}, Ap_k) = 0$

$\|r_{k+1}\|_{A^{-1}} = \min_\alpha \|r_k - \alpha A p_{k-1}\|_{A^{-1}}$

Conjugate Gradient - properties

Orthogonality and minimisation

$$(r_k,p_i) = (r_k,r_i) = 0,\ \ i \leq k-1$$
$$(p_k,Ap_i) = 0,\ \ i \leq k-1$$
$$\|r_{k+1}\|_{A^{-1}} \leq \|r_k\|_{A^{-1}}$$
$$\|r_k\|_{A^{-1}} = \min_{x \in x_0 + Span(p_0,\ldots,p_{k-1})} \|b - Ax\|_{A^{-1}}$$

Krylov method

$\mathcal{K}_k(A,r_0) = Span(r_0,Ar_0,\ldots,A^{k-1}r_0)$ Krylov space

$\mathcal{K}_k(A,r_0) = Span(r_0,r_1,\ldots,r_{k-1}) = Span(p_0,p_1,\ldots,p_{k-1})$

Projection method

$x_k \in x_0 + \mathcal{K}_k(A,r_0)$ Space condition

$r_k \perp \mathcal{K}_k(A,r_0)$ Galerkin condition

Conjugate Gradient - convergence

Polynomial method

$$x_k = x_0 + Q_{k-1}(A) r_0$$
$$r_k = (I - AQ_{k-1}(A)) r_0 = P_k(A) r_0 \text{ with } P_k(0) = 1$$

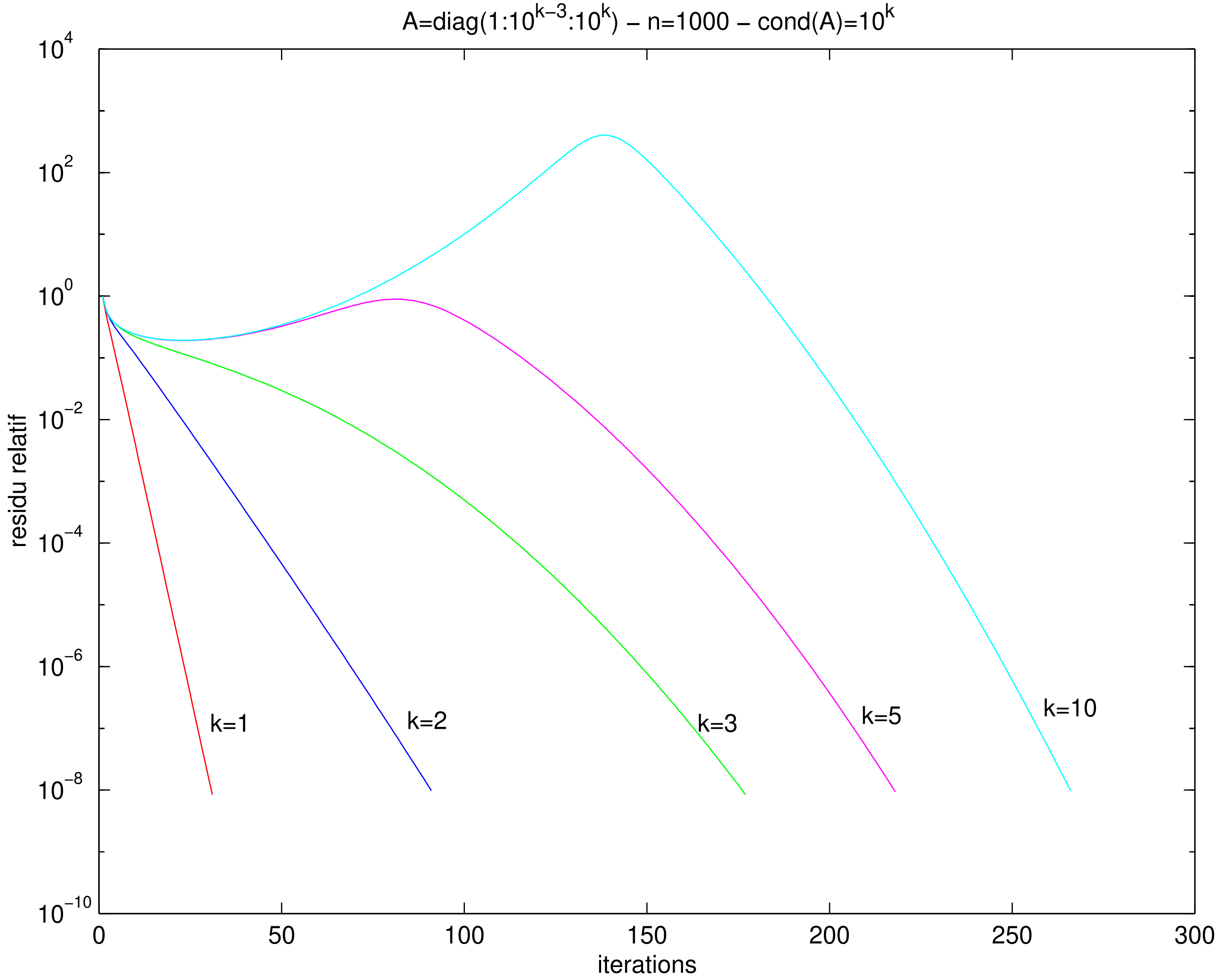
Minmax property - asymptotic convergence

$$A = V \Delta V^{-1} \text{ avec } \Delta = diag(\lambda_1, \ldots, \lambda_n)$$
$$0 < \lambda_1 \leq \ldots \leq \lambda_n \text{ and } \kappa(A) = \lambda_n / \lambda_1$$

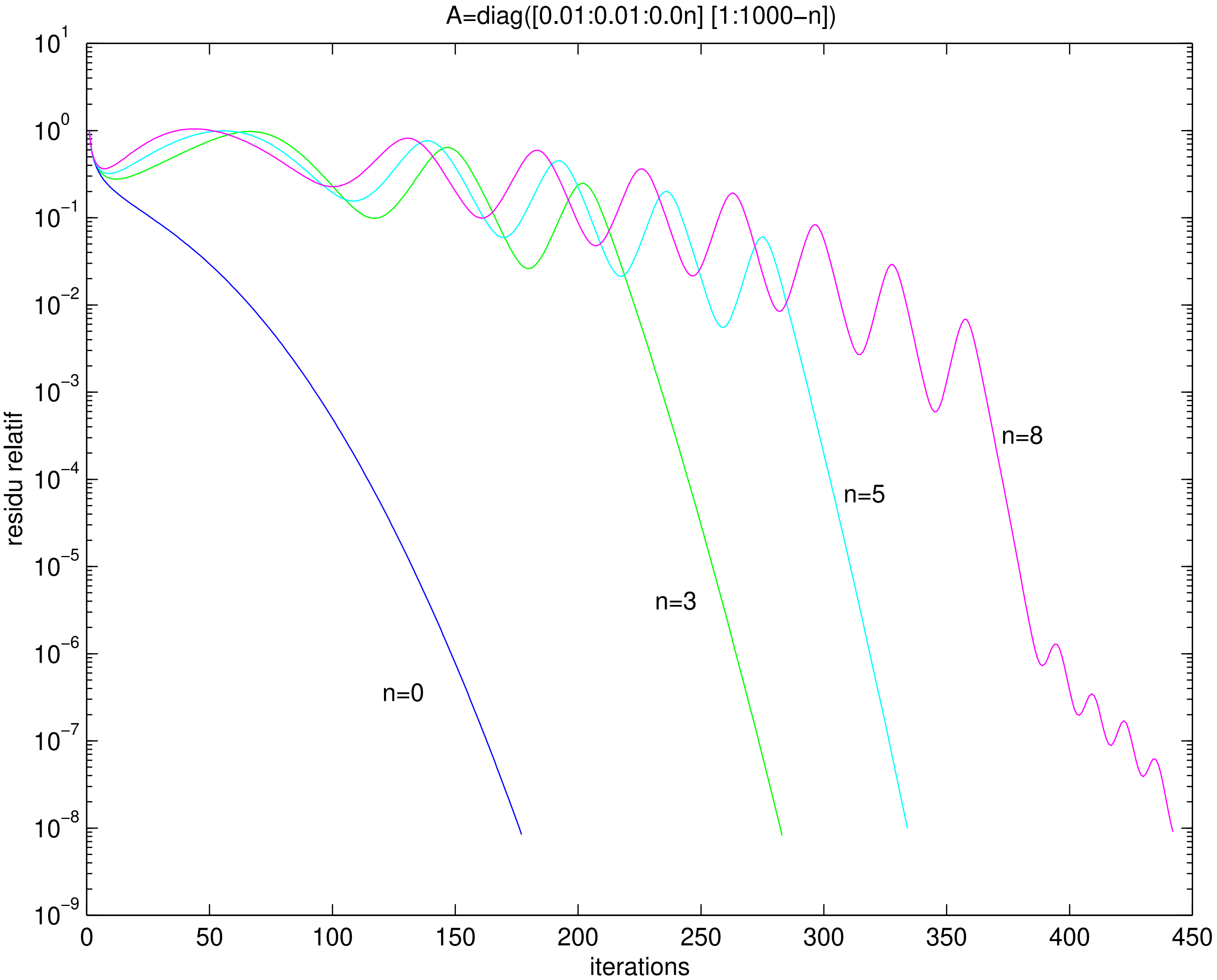
$$\|r_k\|_{A^{-1}} \leq \|r_0\|_{A^{-1}} \max_{\{\lambda_j\}} |P_k(\lambda_j)|$$
$$\leq \|r_0\|_{A^{-1}} \min_{\{P/deg(P)=k \; P(0)=1\}} \max_{\lambda_1 \leq t \leq \lambda_n} |P(t)|$$

$$\|r_k\|_{A^{-1}} \leq 2 \|r_0\|_{A^{-1}} \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k$$

Conjugate Gradient - asymptotic convergence



Conjugate Gradient - superlinear convergence



Preconditioned Conjugate Gradient (PCG)

symmetric positif definite preconditioning matrix M

Algorithm

Initialisation

choose x_0

$r_0 = b - Ax_0$

$z_0 = M^{-1} r_0$

$p_0 = z_0$

For $k = 0,1 \ldots$

$\alpha_k = \frac{(r_k, z_k)}{(Ap_k, p_k)}$

$x_{k+1} = x_k + \alpha_k p_k$

$r_{k+1} = r_k - \alpha_k A p_k$

$z_{k+1} = M^{-1} r_{k+1}$

$\beta_{k+1} = \frac{(r_{k+1}, z_{k+1})}{(r_k, z_k)}$

$p_{k+1} = z_{k+1} + \beta_{k+1} p_k$

End For

Symmetric positive definite preconditionings

- diagonal or Jacobi
- SSOR
- m-step SSOR
- Schwarz
- polynomial
- Incomplete Cholesky IC(k) or ICT
- Approximate inverse
- multigrid
- multilevel
- etc

Jacobi and SSOR preconditionings

decomposition $A = D + L + U$,
D diagonal, L lower triangular, U upper triangular

Jacobi

$M = D$: parallel but slow convergence

SSOR

$M = (D + L)D^{-1}(D + U)$: faster convergence but sequential
If A is symmetric positive definie, M too

Block versions

D,L,U Block-diagonal and block-triangular matrices
faster convergence

Incomplete Cholesky factorisations

$A = LDL^T + R$

various strategies to choose R

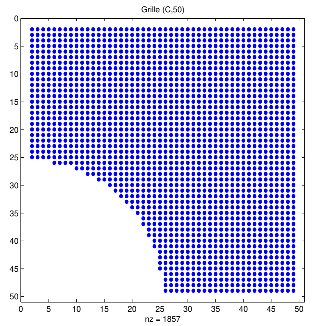
IC(0) : no fill-in

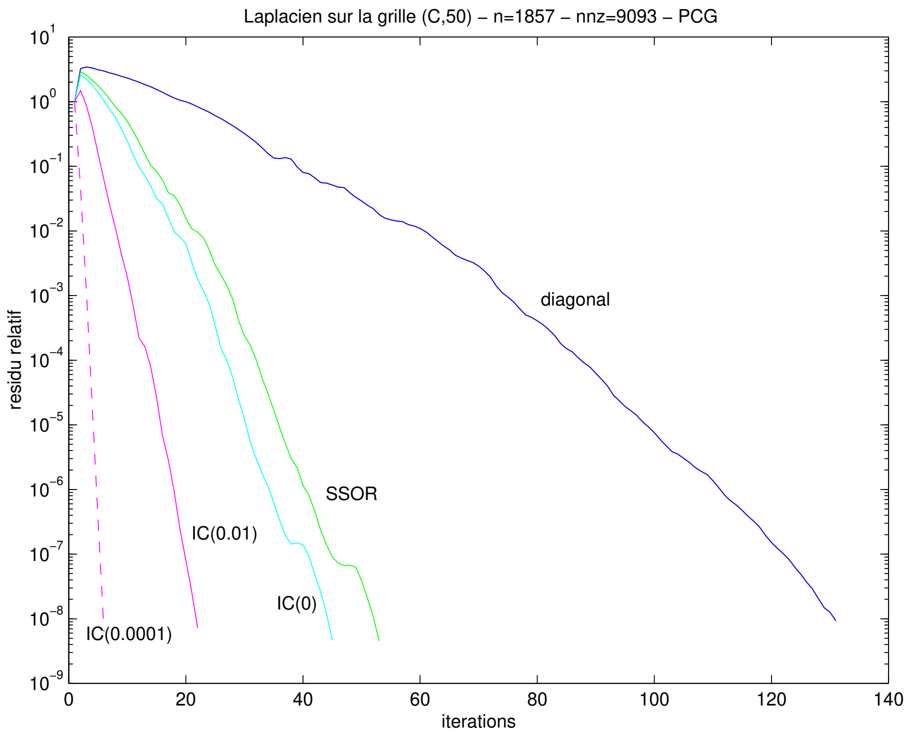
IC(k) : fill-in up to level k

ICT(α) : fill-in with threshold value α

PCG - comparisons

Laplacian on a Finite differences grid





Symmetric Lanczos

Symmetric Lanczos method

$Span(V_k) = \mathcal{K}_k(A,v_1)$ Base de l'espace de Krylov

$V_k^T V_k = I$ Système orthonormé

$$AV_k = V_k T_k + \delta_{k+1} v_{k+1} e_k^T \quad T_k = \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_k \\ & & \delta_k & \gamma_k \end{pmatrix}$$

Relation with CG

$r_0 = \|r_0\|_2 v_1 = \beta v_1$ Krylov space

$x_k = x_0 + V_k y \quad y \in \mathbb{R}^k$ Space condition

$r_k = r_0 - AV_k y = V_k(\beta e_1 - T_k y) - \delta_{k+1}(e_k^T y) v_{k+1}$

$V_k^T r_k = 0 \Leftrightarrow T_k y = \beta e_1$ Galerkin condition

CG : Cholesky Factorisation of T_k

and minimisation of $\|r_k\|_{A^{-1}}$

A symmetric indefinite

SYMMLQ, MINRES

Symmetric Lanczos with least-squares

Preconditioning : open problem

A non symmetric

More difficult ...

Not possible to get
a short recurrence together with
a minimisation of the residual

- Minimisation : GMRES
- Short recurrence : Bi-Conjugate Gradient, QMR, etc
- Preconditioning

Arnoldi and GMRES

Arnoldi process

$Span(V_k) = \mathcal{K}_k(A, v_1)$	Basis of Krylov space
$V_k^T V_k = I$	Orthonormal system
$AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T$	H_k Hessenberg matrix

GMRES algorithm

$r_0 = \|r_0\|_2 v_1 = \beta v_1$	Krylov space
$x_k = x_0 + V_k y \quad y \in \mathbb{R}^k$	Space condition
$r_k \perp AV_k$	Galerkin condition

$$r_k = r_0 - AV_k y = V_{k+1}(\beta e_1 - \overline{H}_k y) \quad \overline{H}_k = \begin{pmatrix} H_k \\ h_{k+1,k} e_k^T \end{pmatrix}$$

$$(AV_k)^T r_k = \overline{H}_k^T (\beta e_1 - \overline{H}_k y)$$

Solve $\min_{y \in \mathbb{R}^k} \|\beta e_1 - \overline{H}_k y\|_2$ Galerkin condition

GMRES - Convergence

Polynomial method

$$x_k = x_0 + Q_{k-1}(A)r_0$$
$$r_k = (I - AQ_{k-1}(A))r_0 = P_k(A)r_0 \text{ avec } P_k(0) = 1$$

Minmax property

IF $A = U\Delta U^{-1}$ with $\Delta = diag(\lambda_1, \ldots, \lambda_n)$

$$\|r_k\|_2 \leq \|r_0\|_2 \kappa(U) \min_{\{P/deg(P)=k \; P(0)=1\}} \max_{1 \leq j \leq n} |P(\lambda_j)|$$

GMRES - restarting

Memory space and complexity

k itérations

Arnoldi : $O(k \times nz) + O(k^2 \times n)$ operations

Least-squares : negligible

Storage of $k+3$ vectors of length n

Restarted GMRES(k)

Initialisation

choix de x_0

Until convergence

k itérations de GMRES

$x_k = x_0 + V_k y$

$x_0 = x_k$

End Until

Risk of stagnation

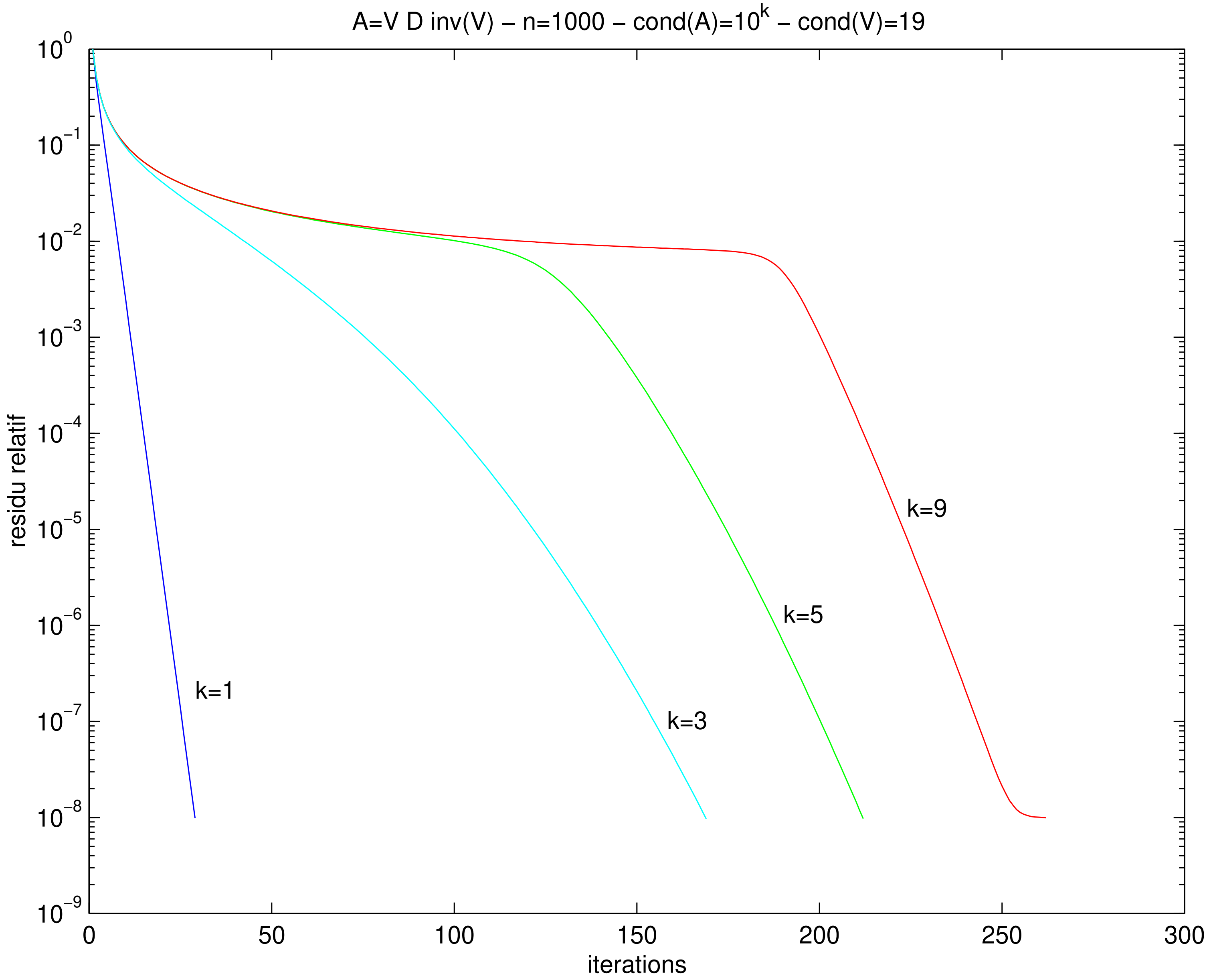
GMRES - algorithm

Arnoldi

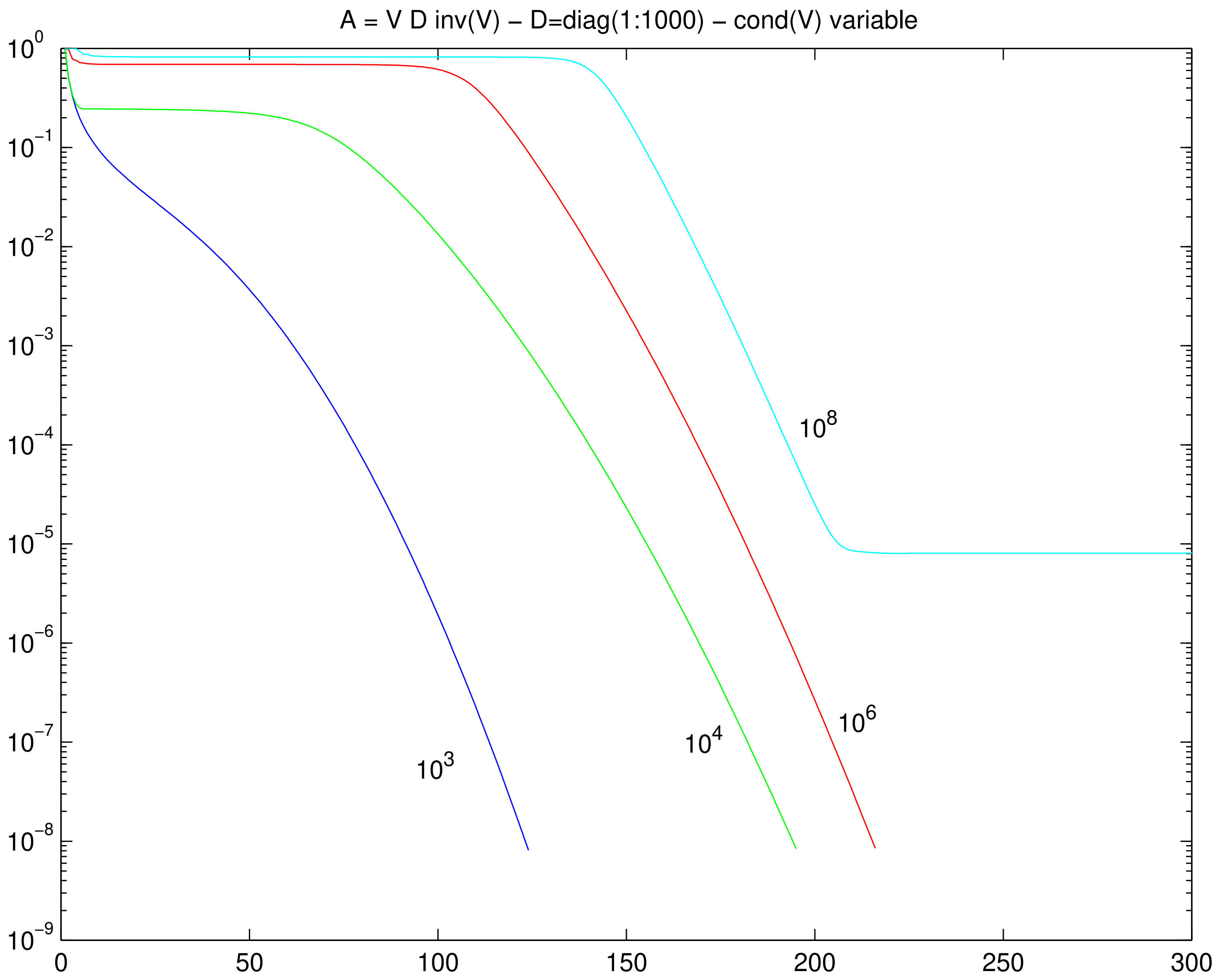
Givens rotations for least-squares

Restarting

GMRES - convergence with varying eigenvalues



GMRES - convergence with varying eigenvectors



Bi-Conjugate Gradient (Bi-CG)

Algorithm

Initialisation

choose x_0 and $\tilde{x}_0$

$r_0 = b - Ax_0$ et $\tilde{r}_0 = b - A^T\tilde{x}_0$

$p_0 = r_0$ et $\tilde{p}_0 = \tilde{r}_0$

For $k = 0,1\dots$

$\alpha_k = \frac{(r_k,\tilde{r}_k)}{(Ap_k,\tilde{p}_k)}$

$x_{k+1} = x_k + \alpha_k p_k$

$\tilde{x}_{k+1} = \tilde{x}_k + \alpha_k \tilde{p}_k$

$r_{k+1} = r_k - \alpha_k A p_k$

$\tilde{r}_{k+1} = \tilde{r}_k - \alpha_k A^T \tilde{p}_k$

$\beta_{k+1} = \frac{(r_{k+1},\tilde{r}_{k+1})}{(r_k,\tilde{r}_k)}$

$p_{k+1} = r_{k+1} + \beta_{k+1} p_k$

$\tilde{p}_{k+1} = \tilde{r}_{k+1} + \beta_{k+1} \tilde{p}_k$

End For

Bi-CG - properties

Orthogonality

$$(r_k,\tilde{r}_i) = 0, \;\; i \leq k-1$$
$$(\tilde{p}_k,Ap_i) = 0, \;\; i \leq k-1$$

No minimisation

Krylov method

$\mathcal{K}_k(A,r_0) = Span(r_0,Ar_0,\ldots,A^{k-1}r_0)$ Krylov space with A
$\mathcal{K}_k(A^T,\tilde{r}_0) = Span(\tilde{r}_0,A^T r_0,\ldots,(A^T)^{k-1}\tilde{r}_0)$ Krylov space with A^T

Projection method

$x_k \in x_0 + \mathcal{K}_k(A,r_0)$ Space condition
$r_k \perp \mathcal{K}_k(A^T,\tilde{r}_0)$ Galerkin condition

Non symmetric Lanczos

Method

$Span(V_k) = \mathcal{K}_k(A, v_1)$ Basis of Krylov space

$Span(W_k) = \mathcal{K}_k(A^T, w_1)$ Basis of dual Krylov space

$W_k^T V_k = I$ Bi-orthogonal system

$$AV_k = V_k T_k + \delta_{k+1} v_{k+1} e_k^T \qquad T_k = \begin{pmatrix} \gamma_1 & \eta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \eta_k \\ & & \delta_k & \gamma_k \end{pmatrix}$$

$$A^T W_k = W_k T_k^T + \eta_{k+1} w_{k+1} e_k^T$$

Relation with Bi-CG

$r_0 = \|r_0\|_2 v_1 = \beta v_1$ Krylov space

$x_k = x_0 + V_k y \quad y \in \mathbb{R}^k$ Space condition

$r_k = r_0 - AV_k y = V_k(\beta e_1 - T_k y) - \delta_{k+1}(e_k^T y) v_{k+1}$

$W_k^T r_k = 0 \Leftrightarrow T_k y = \beta e_1$ Galerkin condition

Bi-CG : Gauss Factorisation of T_k

Bi-CG - Convergence - Variants

risk of Breakdown in Lanczos : $(r_k, \tilde{r}_k) = 0$
Lanczos version with Look-Ahead

irregular convergence
Product by A^T : Transpose-Free version BICGSTAB
Smoother convergence of BICGSTAB

risk of Breakdown in LU : $(\tilde{p}_k, Ap_k) = 0$
QMR Algorithm

Quasi-minimum Residual QMR

Algorithm

Non symmetric Lanczos with look-ahead

$$r_k = V_{k+1}(\beta e_1 - \overline{T}_k y) \qquad \overline{T}_k = \begin{pmatrix} T_k \\ \delta_{k+1} e_k^T \end{pmatrix}$$

$$\Omega_{k+1} = diag(\|v_1\|_2, \ldots, \|v_{k+1}\|_2)$$

Solve $\min_{y \in \mathbb{R}^k} \|\beta e_1 - \Omega_{k+1} \overline{T}_k y\|_2$

Convergence

No breakdown

IF T_k is diagonalisable

$$\|r_k\|_2 \leq \|r_0\|_2 \sqrt{k+1}\, \kappa(T_k) \min_{\{P/deg(P)=k\ P(0)=1\}} \max_{1 \leq j \leq n} |P(\lambda_j)|$$

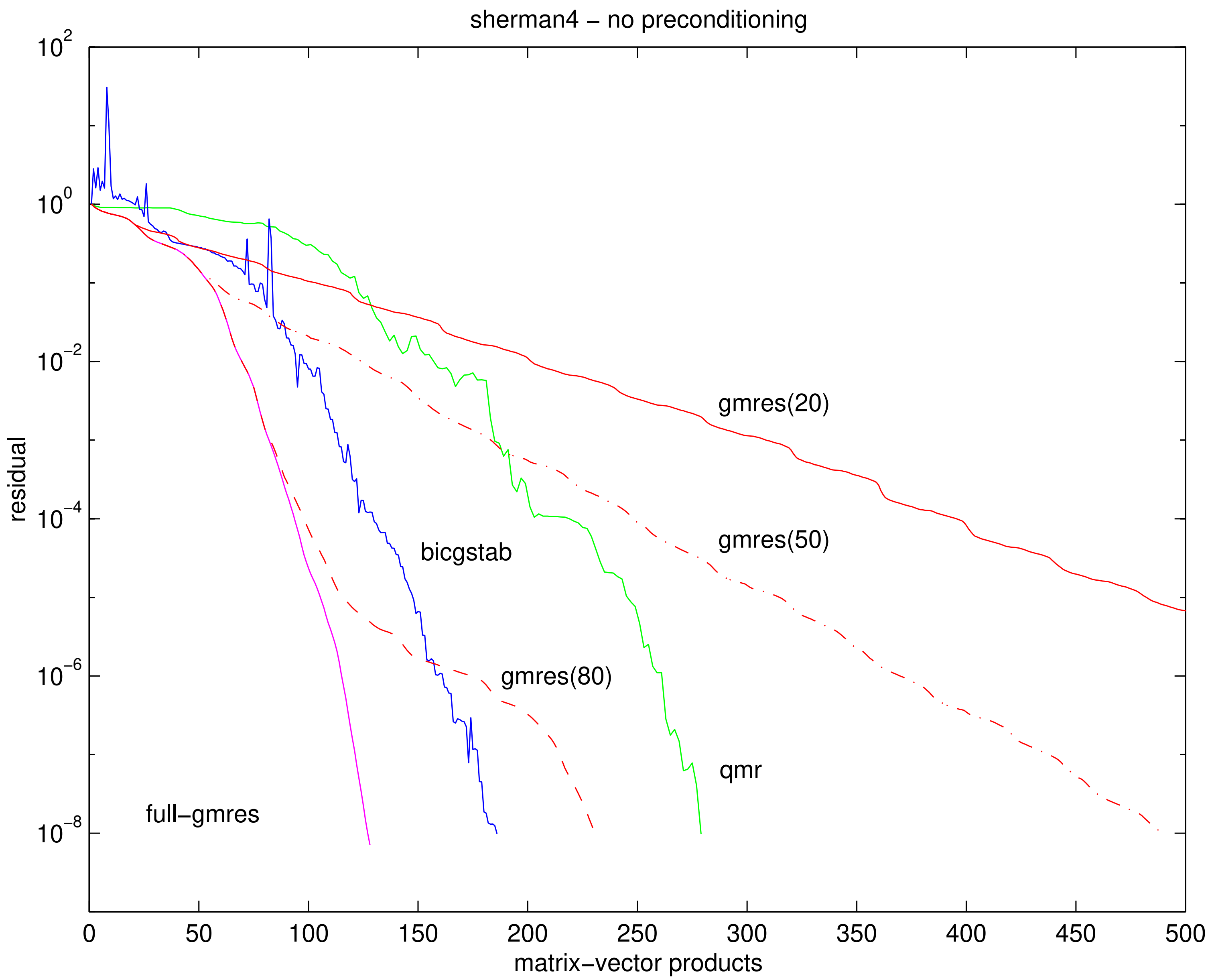
Some Preconditionings

- diagonal or Jacobi
- Gauss-Seidel, SOR
- m-step SOR
- Schwarz
- polynomial
- Incomplete Gauss ILU(k) or ILUT
- Approximate inverse
- multigrid
- multilevel
- etc

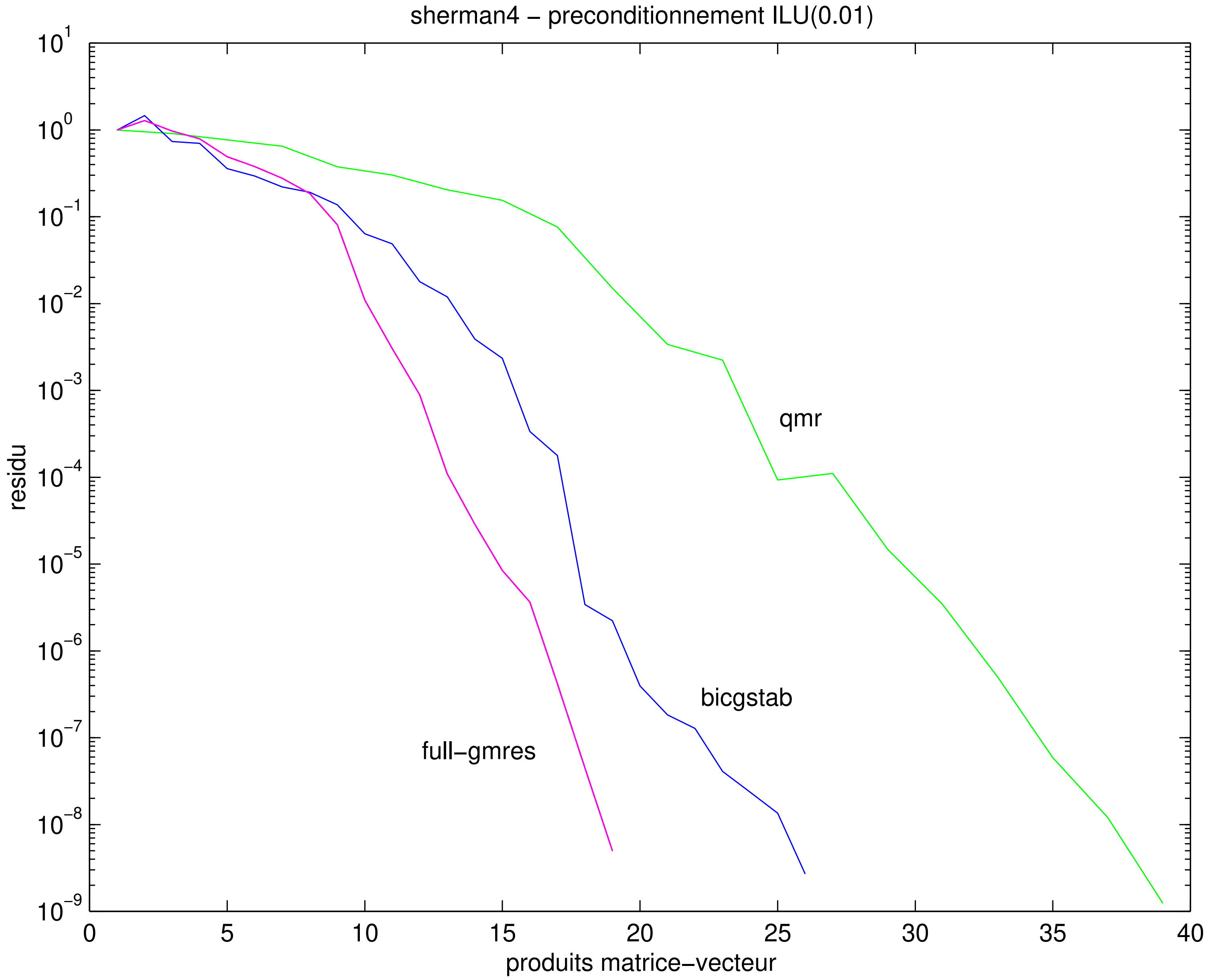
Comparison between Bi-CG, QMR, GMRES

matrices from Harwell-Boeing (HB) collection and from matrix market

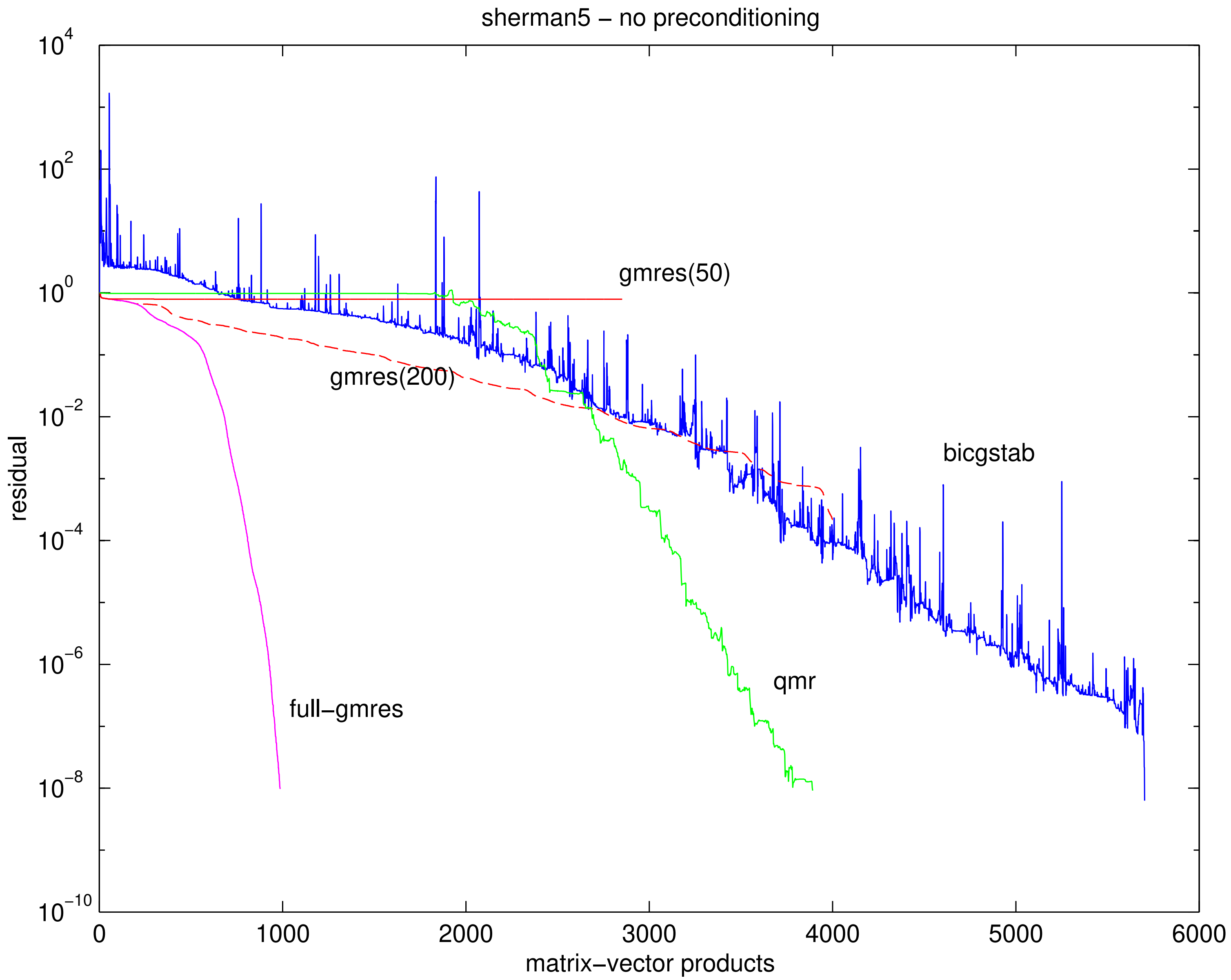
Matrix Sherman4 (HB) - n=1104-nz=3786



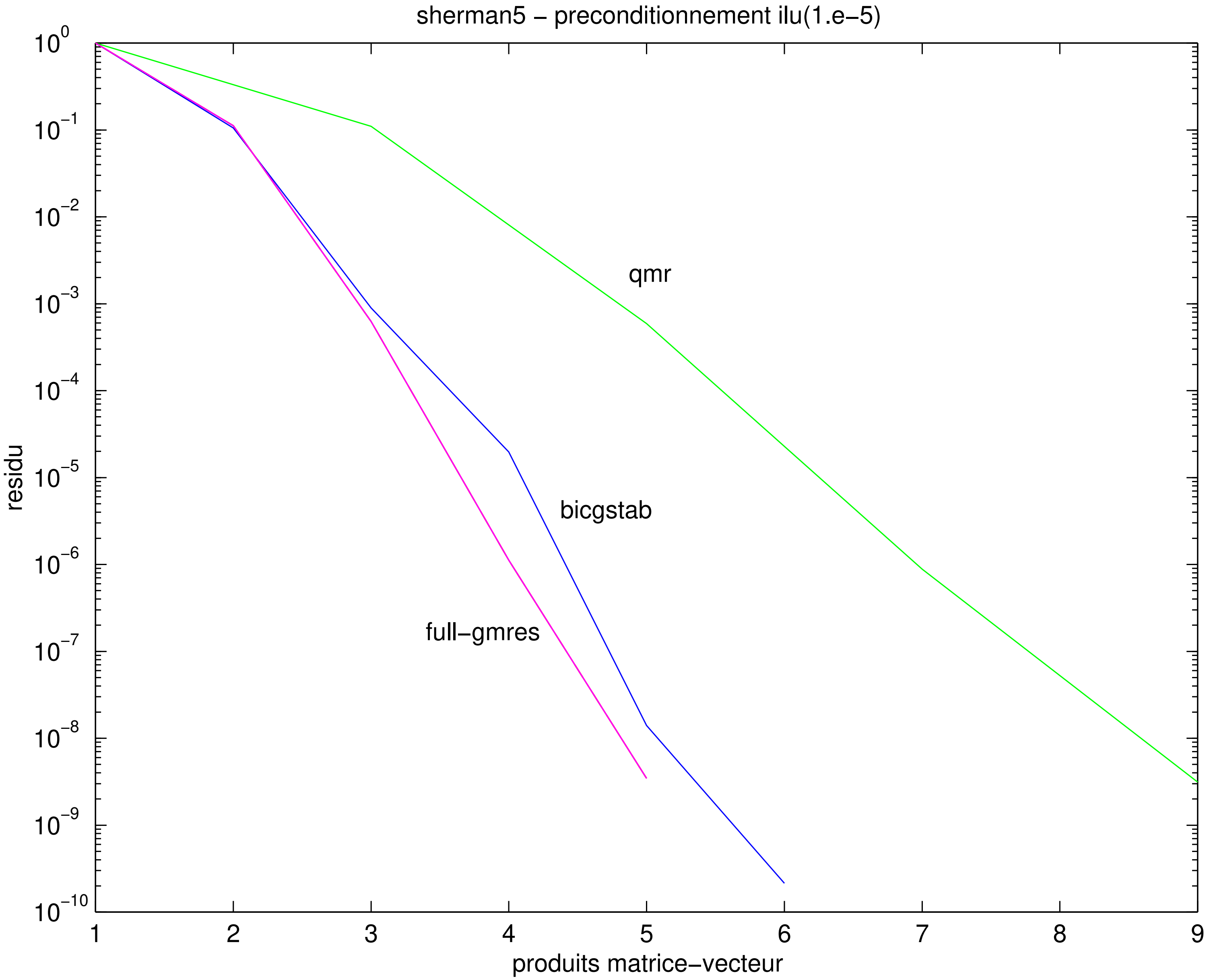
Matrix Sherman4 (HB) - n=1104 - nz=3786



Matrix Sherman5 (HB) - n=3312 - nz=20793



Matrix Sherman5 (HB)- n=3312 - nz=20793



3D Biharmonic equations

Joint work with I. Altas and M. Gupta

Ω is a closed convex domain in three- dimensions
and $\partial\Omega$ is its boundary.
$(x,y,z) \in \Omega$
Biharmonic equation

$$\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4} + \frac{\partial^4 u}{\partial z^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + 2\frac{\partial^4 u}{\partial x^2 \partial z^2} + 2\frac{\partial^4 u}{\partial y^2 \partial z^2} = f(x,y,z),$$

with Dirichlet boundary conditions

$$u = g_1(x,y,z), \qquad \frac{\partial u}{\partial n} = g_2(x,y,z) \qquad (x,y,z) \in \partial\Omega.$$

Choice of Finite Difference scheme

3D uniform grid

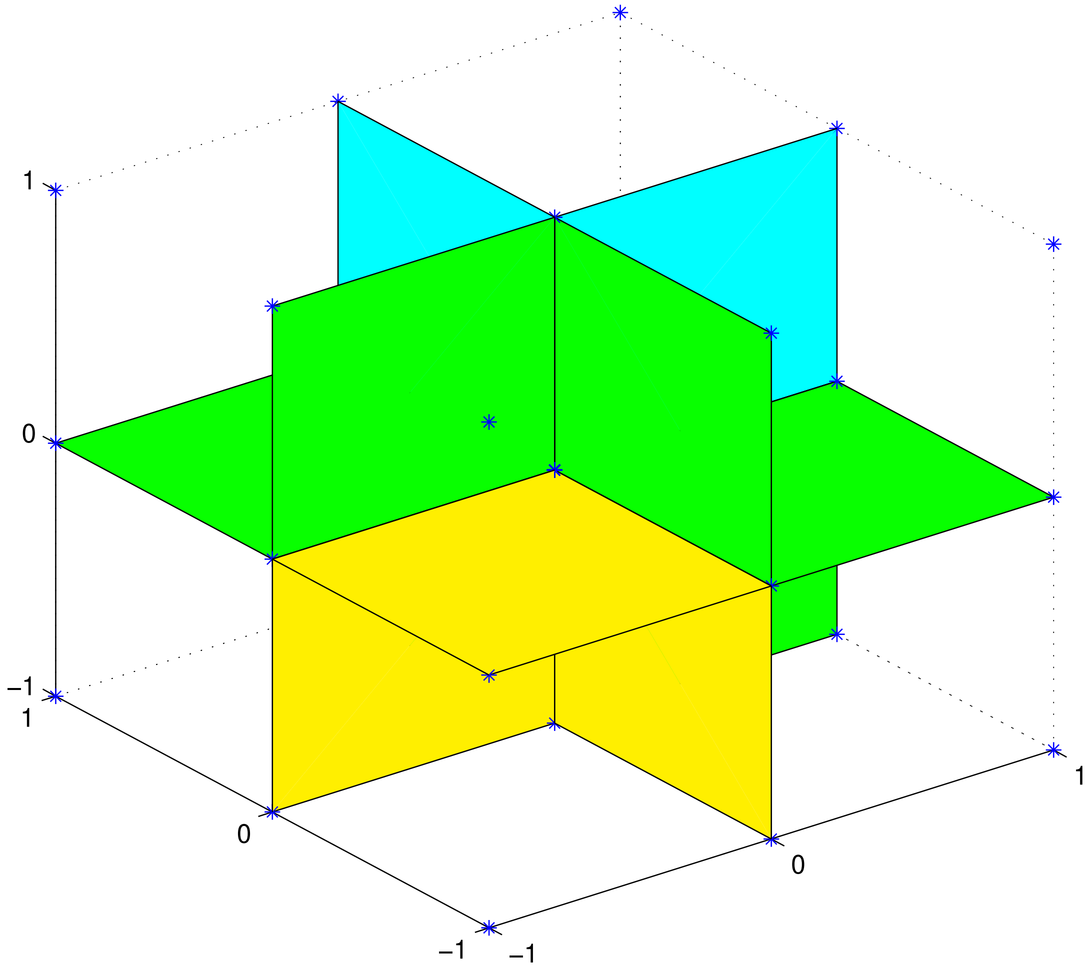
4 unknowns at each grid point (x_i, y_j, z_k) :

$$u(x_i, y_j, z_k), \frac{\partial u}{\partial x}(x_i, y_j, z_k), \frac{\partial u}{\partial y}(x_i, y_j, z_k), \frac{\partial u}{\partial z}(x_i, y_j, z_k)$$

Compact Finite Difference approximation

- Exact boundary conditions
- no modification near the boundary
- values of gradients already available

Compact 27 points 3D Finite Difference cell



Linear solving methods

Large sparse non symmetric matrices
Grid $n \times n \times n$ - Matrix order $N = (n+1)^3 * 4$
examples : $n = 32 \Rightarrow N = 143{,}748$ and $n = 64 \Rightarrow N = 1{,}098{,}500$

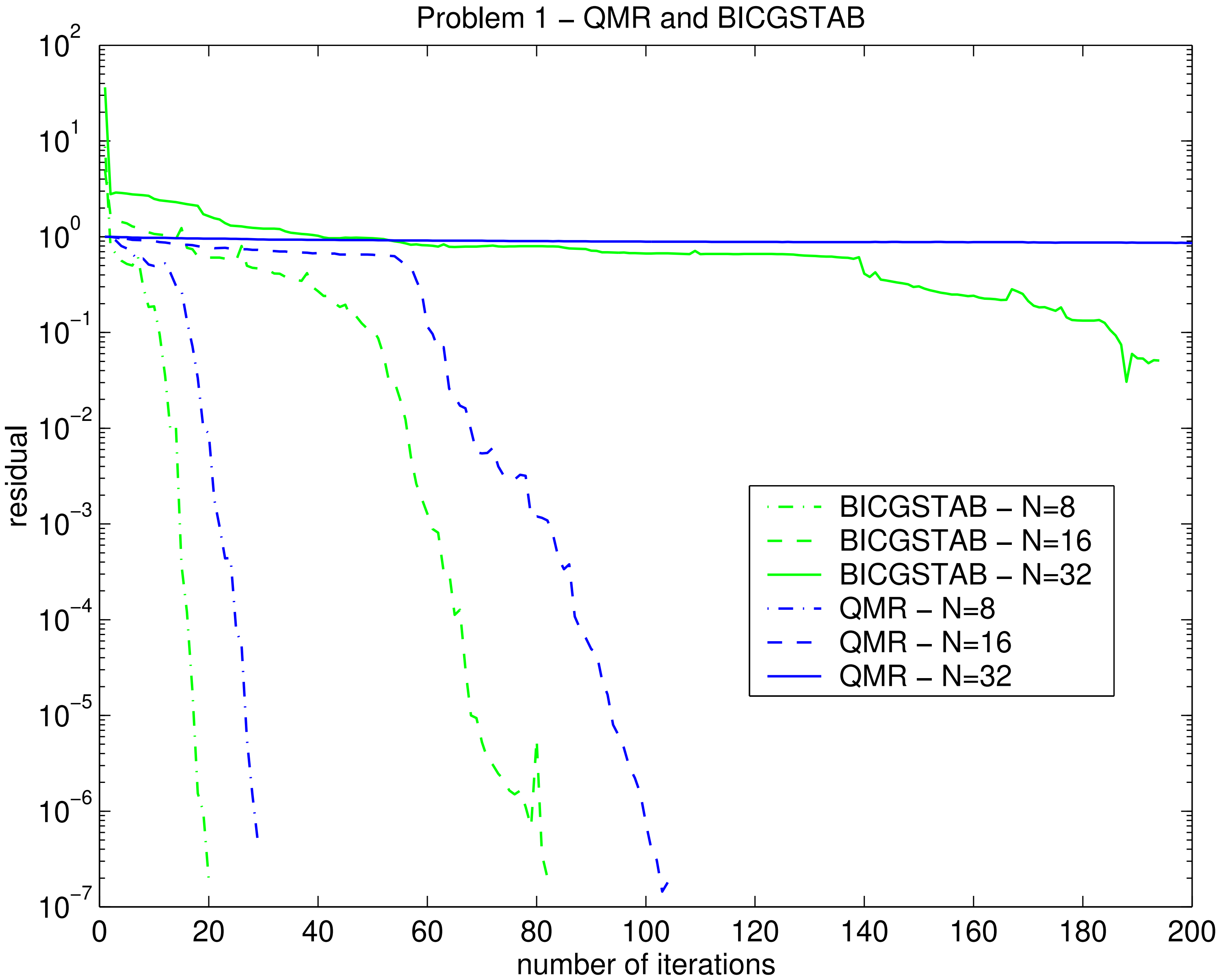
Matrix-free iterative solvers
Matrix-free preconditioners

No fast solver
SSOR too slow

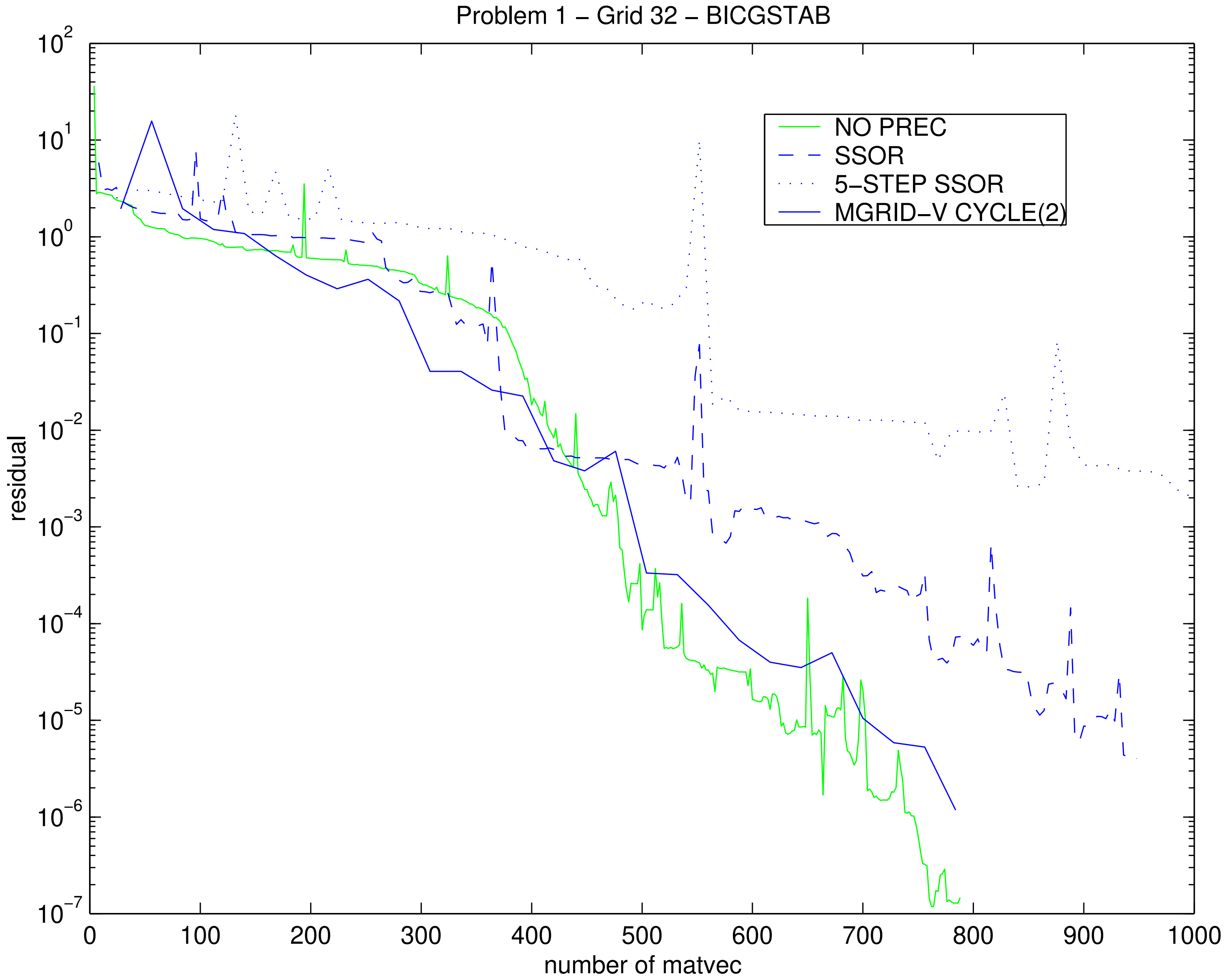
Either MULTIGRID or KRYLOV methods

Comparison between BICGSTAB, QMR, MULTIGRID
With None, SSOR, m-step SSOR, Multigrid preconditioners

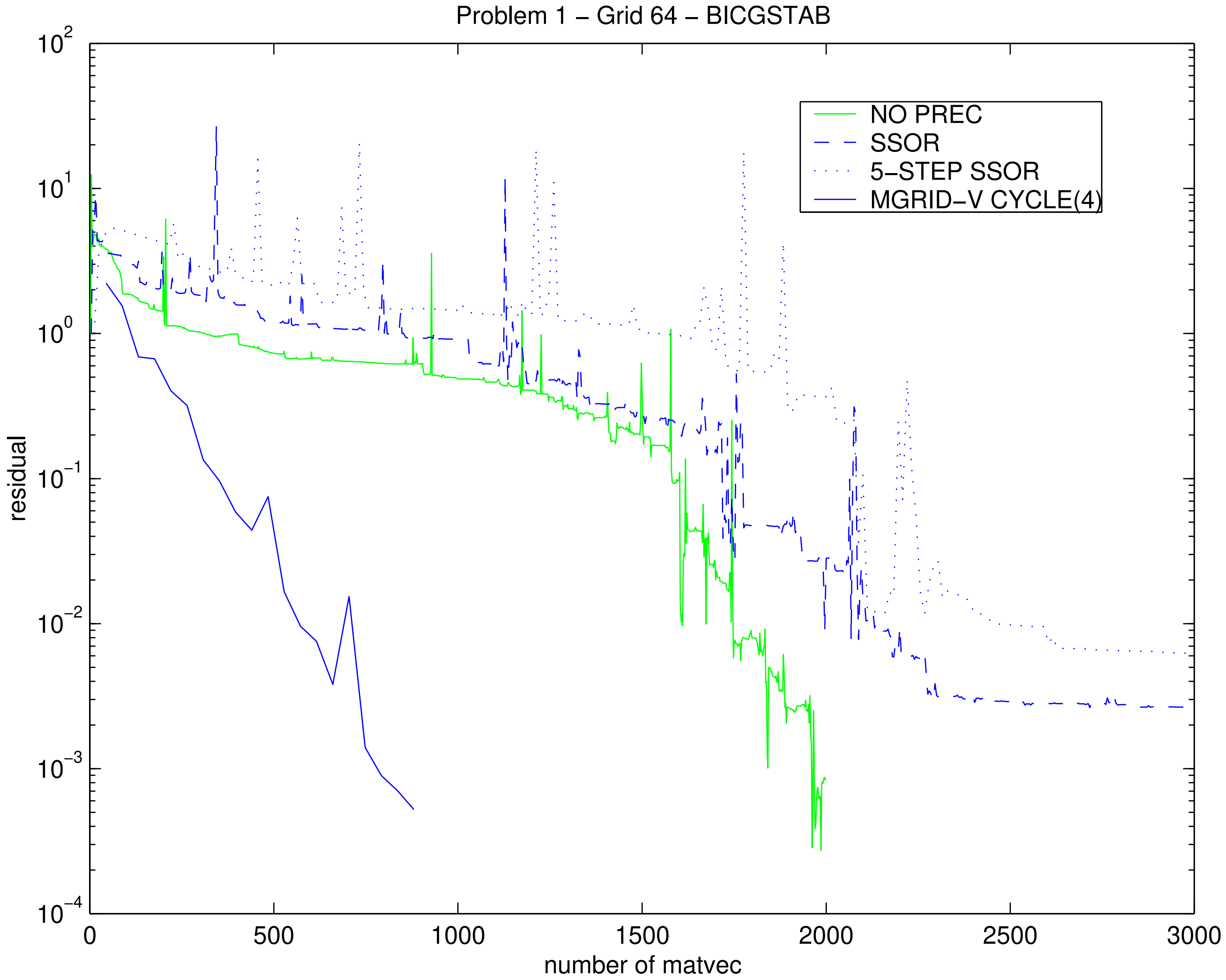
Comparison between BICGSTAB and QMR



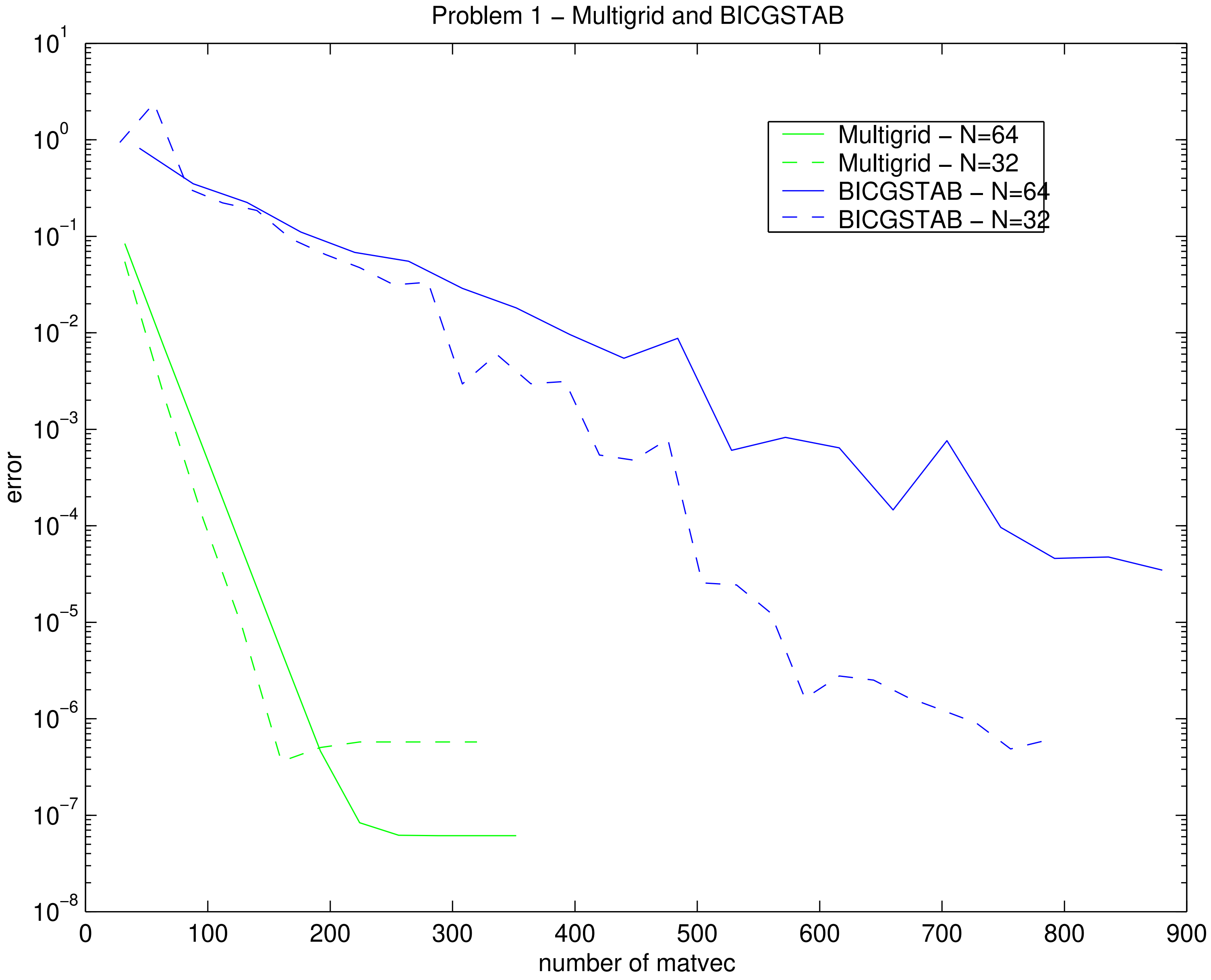
BICGSTAB - Comparison of preconditioners



BICGSTAB - Comparison of preconditioners



Comparison between BICGSTAB and MULTIGRID



Some bibliography and software

- Bibliography
- Software

Bibliography

- Y. Saad, Iterative methods for sparse linear systems, PWS Publishing Company, 1996
  http://www-users.cs.umn.edu/ saad/books.html
- G. Meurant, Computer solution of large linear systems, North Holland, 1999
  http://perso.wanadoo.fr/gerard.meurant/#GACC
- Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition, R. Barrett et al., SIAM, 1994
  http://www.netlib.org/templates/index.html
- J. Erhel, chapter of book in preparation (in french)
- etc

Software

- list : http://www.netlib.org/utk/papers/iterative-survey/
- PETSc : http://www-fp.mcs.anl.gov/petsc/
  Partial, Ordinary and Algebraic Differential Equations,
  Nonlinear and linear systems,
  on distributed memory computers
- Parpre : http://www.cs.utk.edu/ eijkhout/parpre.html
  Parallel preconditionings for iterative methods
  uses Petsc, part of Scalapack project, V. Eijkhout and T. Chan
- Psparslib : http://www.cs.umn.edu/Research/arpa/p_sparslib/psp-abs.html
  Parallel iterative methods, Y. Saad and al.
- Matlab : http://www.mathworks.com
- Scilab and extension Scilin
  iterative methods, Aladin group
- etc

Conclusion

- Multigrid methods efficient in regular cases
- Krylov methods efficient in general cases
- Preconditioning : necessary and difficult
- Multigrid or multilevel preconditioning : efficient for large problems