GMRES Methods for Least Squares Problems

Ken Hayami, Jun-Feng Yin,
National Institute of Informatics, Tokyo
and
Tokushi Ito
Business Design Laboratory Co.,Ltd.,
Nagoya

HARRACHOV 2007
Computational Linear Algebra with Applications

Harrachov, Czech Republic

August 23rd, 2007

Problem:

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$$

$A \in \mathbf{R}^{m \times n}$

- $m > n$ or $m = n$ or $m < n$
- not necessarily full rank
- large sparse

$$\Updownarrow$$

$$A^\top A\boldsymbol{x} = A^\top \boldsymbol{b} \quad \text{(normal equation)}$$

For $m < n$

$$AA^{\top}\boldsymbol{y} = \boldsymbol{b}, \quad \boldsymbol{x} = A^{\top}\boldsymbol{y}$$

$\downarrow$

minimum norm solution

1. Iterative methods using the normal equation

$A^\top A$: square, symmetric,

positive definite (if rank$A = n$) matrix

$\downarrow$

Apply Conjugate Gradient method.

$\downarrow$

CGLS (Conjugate Gradient Least Squares) method

The CGLS(CGNR) method

Choose $\boldsymbol{x}_0$.

$\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0,\ \boldsymbol{p}_0 = \boldsymbol{s}_0 = A^{\mathsf{T}}\boldsymbol{r}_0,\ \gamma_0 = \|\boldsymbol{s}_0\|_2{}^2$

for $i = 0, 1, 2, \ldots$ until $\gamma_i < \epsilon$

$\boldsymbol{q}_i = A\boldsymbol{p}_i$

$\alpha_i = \gamma_i / \|\boldsymbol{q}_i\|_2{}^2$

$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$

$\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \alpha_i \boldsymbol{q}_i$

$\boldsymbol{s}_{i+1} = A^{\mathsf{T}}\boldsymbol{r}_{i+1}$

$\gamma_{i+1} = \|\boldsymbol{s}_{i+1}\|_2{}^2$

$\beta_i = \gamma_{i+1}/\gamma_i$

$\boldsymbol{p}_{i+1} = \boldsymbol{s}_{i+1} + \beta_i \boldsymbol{p}_i$

endfor

However,

condition number of $A^{\mathsf{T}}A$: square of A.

$\downarrow$

Slow convergence

Preconditioning necessary

- diagonal scaling
- incomplete Cholesky decomposition
  (Meijerink, van der Vorst '77)
- incomplete QR decomposition
  (Jennings, Ajiz '84, Saad '88)
- incomplete Givens orthogonalization
  (Bai et al. '01)
- robust incomplete factorization (RIF)
  (Benzi, Tuma '03)

etc.

2. Iterative methods
<u>not</u> based on the normal equation

CR-LS(k) method (Zhang and Oyanagi, '89)

Apply Orthomin(k) method to

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$$

by introducing a mapping matrix $B \in \mathbf{R}^{n \times m}$, and using the Krylov subspace generated by $AB \in \mathbf{R}^{m \times m}$.

Consider using
GMRES (Generalized Minimal RESidual) method
instead.

GMRES: efficient and robust method for

$$A\boldsymbol{x} = \boldsymbol{b},$$

where

$A \in \mathbf{R}^{n \times n}$: nonsingular.

GMRES(k) method

Choose $\boldsymbol{x}_0$

* $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$; $\boldsymbol{v}_1 = \boldsymbol{r}_0/||\boldsymbol{r}_0||_2$

for $i = 1, 2, \ldots, k$

$\boldsymbol{w}_i = A\boldsymbol{v}_i$

for $j = 1, 2, \ldots, i$

$h_{j,i} = (\boldsymbol{w}_i, \boldsymbol{v}_j)$

$\boldsymbol{w}_i = \boldsymbol{w}_i - h_{j,i}\boldsymbol{v}_j$

end for

$h_{i+1,i} = \|\boldsymbol{w}_i\|_2$

$\boldsymbol{v}_{i+1} = \boldsymbol{w}_i/h_{i+1,i}$

Find $\boldsymbol{y}_i \in \mathbf{R}^i$ which minimizes $\|\boldsymbol{r}_i\|_2 = \| \, \|\boldsymbol{r}_0\|_2 \, \boldsymbol{e}_i - \bar{H}_i \, \boldsymbol{y} \, \|_2 \, .$

if $\|\boldsymbol{r}_i\|_2 < \epsilon$ then

$\boldsymbol{x}_i = \boldsymbol{x}_0 + [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i$; stop

endif

endfor

$\boldsymbol{x}_k = \boldsymbol{x}_0 + [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k]\boldsymbol{y}_k$

$\boldsymbol{x}_0 = \boldsymbol{x}_k$

Go to $*$.

(Here, $\bar{H}_i = (h_{pq}) \in \mathbf{R}^{(i+1)\times i}$,
$\quad\quad \boldsymbol{e}_i = (1, 0, \ldots, 0)^\top \in \mathbf{R}^{i+1}$.)

Minimizes $\|\boldsymbol{r}_k\|_2$ over

$$\boldsymbol{x}_k = \boldsymbol{x}_0 + \langle \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \rangle$$

$$\langle \boldsymbol{v}_1, \cdots, \boldsymbol{v}_k \rangle = \langle \boldsymbol{r}_0, A\boldsymbol{r}_0, \cdots, A^{k-1}\boldsymbol{r}_0 \rangle$$

$$(\boldsymbol{v}_i, \boldsymbol{v}_j) = \delta_{ij}$$

$k = \infty$: (full) GMRES.

$h_{i+1,i} = 0$: breakdown

When A : nonsingular,

GMRES does not break down until it has found
the solution of $A\boldsymbol{x} = \boldsymbol{b}$, $\ {}^\forall\boldsymbol{b}$, ${}^\forall\boldsymbol{x}_0 \in \mathbf{R}^n$
(with in n steps).

When A : singular,

Theorem 1 *(Brown, Walker '97)*

GMRES determines a least squares solution of $\min_{\boldsymbol{x} \in \mathbf{R}^n} ||\boldsymbol{b} - A\boldsymbol{x}||_2$ *without breakdown*
$^\forall \boldsymbol{b}, {}^\forall \boldsymbol{x}_0 \in \mathbf{R}^n$

$$\Updownarrow$$

$$\mathcal{R}(A) = \mathcal{R}(A^\mathsf{T}). \qquad \square$$

$\mathcal{R}(M)$: range space of M.

How can we apply GMRES to
the least squares problem

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$$

where $A \in \mathbf{R}^{m \times n}$?

$A \in \mathbf{R}^{m \times n}, \quad \boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0 \in \mathbf{R}^m.$

$\downarrow$

Cannot create Krylov subspace by $A \times \boldsymbol{r}_0$.

Basically two ways to overcome
by using a mapping matrix $B \in \mathbf{R}^{n \times m}$.

1. AB-GMRES method

Use Krylov subspace:

$$\mathcal{K}_k(AB, \boldsymbol{r}_0) := \langle \boldsymbol{r}_0, AB\boldsymbol{r}_0, \ldots, (AB)^{i-1}\boldsymbol{r}_0 \rangle \text{ in } \mathbf{R}^m.$$

$AB \in \mathbf{R}^{m \times m}$.

(cf. CR-LS(k) method by Zhang, Oyanagi)

First note:

Lemma 2

$$\min_{\boldsymbol{x}\in\mathbf{R}^n} ||\boldsymbol{b} - A\boldsymbol{x}||_2 = \min_{\boldsymbol{z}\in\mathbf{R}^m} ||\boldsymbol{b} - AB\boldsymbol{z}||_2 \quad {}^{\forall}\boldsymbol{b} \in \mathbf{R}^m$$

$$\Updownarrow$$

$$\mathcal{R}(A) = \mathcal{R}(AB) \qquad \square$$

Using

Lemma 3 $\mathcal{R}(AA^{\mathsf{T}}) = \mathcal{R}(A) \quad \square$

we can show

Lemma 4 $\mathcal{R}(A^{\mathsf{T}}) = \mathcal{R}(B) \implies \mathcal{R}(A) = \mathcal{R}(AB). \quad \square$

Thus, assume $\mathcal{R}(A) = \mathcal{R}(AB)$.

Consider solving

$$\min_{\boldsymbol{z} \in \mathbf{R}^m} ||\boldsymbol{b} - AB\boldsymbol{z}||_2 = \min_{\boldsymbol{x} \in \mathbf{R}^n} ||\boldsymbol{b} - A\boldsymbol{x}||_2$$

using GMRES with initial approximation
$\boldsymbol{z}_0 \in \mathbf{R}^m; \;\; AB\boldsymbol{z}_0 = A\boldsymbol{x}_0.$

Then, we have the following.

AB-GMRES(k) method

Choose $\boldsymbol{x}_0$ (; $A\boldsymbol{x}_0 = AB\boldsymbol{z}_0$).

$*$ $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$ ($= \boldsymbol{b} - AB\boldsymbol{z}_0$); $\boldsymbol{v}_1 = \boldsymbol{r}_0/||\boldsymbol{r}_0||_2$

for $i = 1, 2, \ldots, k$

$\quad \boldsymbol{w}_i = AB\boldsymbol{v}_i$

$\quad$ for $j = 1, 2, \ldots, i$

$\quad\quad h_{j,i} = (\boldsymbol{w}_i, \boldsymbol{v}_j)$

$\quad\quad \boldsymbol{w}_i = \boldsymbol{w}_i - h_{j,i}\boldsymbol{v}_j$

$\quad$ end for

$\quad h_{i+1,i} = \|\boldsymbol{w}_i\|_2$

$\quad \boldsymbol{v}_{i+1} = \boldsymbol{w}_i/h_{i+1,i}$

$\quad$ Find $\boldsymbol{y}_i \in \mathbf{R}^i$ which minimizes $\|\boldsymbol{r}_i\|_2 = \| \, \|\boldsymbol{r}_0\|_2 \, \boldsymbol{e}_i - \bar{H}_i \, \boldsymbol{y} \, \|_2$

$\quad \boldsymbol{x}_i = \boldsymbol{x}_0 + B[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i$

$\quad \boldsymbol{r}_i = \boldsymbol{b} - A\boldsymbol{x}_i$

$\quad$ if $\|A^{\mathsf{T}}\boldsymbol{r}_i\|_2 < \epsilon$ stop

endfor

$\boldsymbol{x}_0 = \boldsymbol{x}_k$

Go to $*$.

Does the AB-GMRES method determine the least squares solution without breakdown ?

Recall the following.

Theorem 5 *(Brown, Walker '97)*
Let $\tilde{A} \in \mathbf{R}^{m \times m}$*. Then, the following holds.*

The GMRES method determines
a least squares solution of $\min_{\boldsymbol{z} \in \mathbf{R}^m} ||\boldsymbol{b} - \tilde{A}\boldsymbol{z}||_2$
$^\forall \boldsymbol{b} \in \mathbf{R}^m, {}^\forall \boldsymbol{z}_0 \in \mathbf{R}^m$ *without breakdown*

$$\Updownarrow$$

$$\mathcal{R}(\tilde{A}) = \mathcal{R}(\tilde{A}^\top). \quad \square$$

Let $\tilde{A} := AB$. Noting

Theorem 6 *If* $\mathcal{R}(A^\top) = \mathcal{R}(B)$, *then*

$$\mathcal{R}(AB) = \mathcal{R}(B^\top A^\top) \iff \mathcal{R}(A) = \mathcal{R}(B^\top) \quad \square$$

we obtain

Theorem 7
If $\mathcal{R}(A^\top) = \mathcal{R}(B)$, *then*

AB-GMRES method determines
a least squares solution of $\min_{\boldsymbol{x} \in \mathbf{R}^n} ||\boldsymbol{b} - A\boldsymbol{x}||_2$
${}^\forall \boldsymbol{b} \in \mathbf{R}^m$, ${}^\forall \boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown*

$$\Updownarrow$$

$$\mathcal{R}(A) = \mathcal{R}(B^\top). \quad \square$$

Corollary 8

$\mathcal{R}(A^\top) = \mathcal{R}(B),\ \mathcal{R}(A) = \mathcal{R}(B^\top)$

$$\Downarrow$$

AB-GMRES method determines
a least squares solution of $\min_{\boldsymbol{x}\in\mathbf{R}^n} ||\boldsymbol{b} - A\boldsymbol{x}||_2$
$^\forall\boldsymbol{b} \in \mathbf{R}^m,\ ^\forall\boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown.* $\square$

Remark 1

$\mathcal{R}(A) = \mathcal{R}(B^\top)$

$$\Uparrow$$

$B^\top = AC^\top;\ \ C^\top$: nonsingular

$$\Updownarrow$$

$B = CA^\top;\ \ C$: nonsingular.

2. BA-GMRES method

The other alternative:

Use a matrix $B \in \mathbf{R}^{n \times m}$
to map $\boldsymbol{r}_0 \in \mathbf{R}^m$ to $\tilde{\boldsymbol{r}}_0 = B\boldsymbol{r}_0 \in \mathbf{R}^n$.

Then create Krylov subspace

$$\langle \tilde{\boldsymbol{r}}_0, BA\tilde{\boldsymbol{r}}_0, \ldots, (BA)^{i-1}\tilde{\boldsymbol{r}}_0 \rangle \text{ in } \mathbf{R}^n.$$

$$BA \in \mathbf{R}^{n \times n}$$

First note:

Theorem 9

$$\|\boldsymbol{b} - A\boldsymbol{x}^*\|_2 = \min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2$$

and

$$\|B\boldsymbol{b} - BA\boldsymbol{x}^*\|_2 = \min_{\boldsymbol{x} \in \mathbf{R}^n} \|B\boldsymbol{b} - BA\boldsymbol{x}\|_2$$

are equivalent for all $\boldsymbol{b} \in \mathbf{R}^m$

$$\Updownarrow$$

$$\mathcal{R}(A) = \mathcal{R}(B^\top BA).$$

Also note

Lemma 10
$\mathcal{R}(A) = \mathcal{R}(B^\top) \Longrightarrow \mathcal{R}(BA) = \mathcal{R}(B).$

Lemma 11
$\mathcal{R}(BA) = \mathcal{R}(B) \Longrightarrow \mathcal{R}(B^\top BA) = \mathcal{R}(B^\top).$

Lemma 12
$\mathcal{R}(A) = \mathcal{R}(B^\top) \Longrightarrow \mathcal{R}(A) = \mathcal{R}(B^\top BA).$

Thus, assume $\mathcal{R}(A) = \mathcal{R}(B^\top BA)$,

and consider applying GMRES(k) method to

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} ||B\boldsymbol{b} - BA\boldsymbol{x}||_2$$

with initial approximation $\boldsymbol{x}_0$.

BA-GMRES(k) method

Choose $\boldsymbol{x}_0$.

* $\tilde{\boldsymbol{r}}_0 = B(\boldsymbol{b} - A\boldsymbol{x}_0)$

$\boldsymbol{v}_1 = \tilde{\boldsymbol{r}}_0 / ||\tilde{\boldsymbol{r}}_0||_2$

for $i = 1, 2, \ldots, k$ until convergence

$\boldsymbol{w}_i = BA\boldsymbol{v}_i$

for $j = 1, 2, \ldots, i$

$h_{j,i} = (\boldsymbol{w}_i, \boldsymbol{v}_j)$

$\boldsymbol{w}_i = \boldsymbol{w}_i - h_{j,i}\boldsymbol{v}_j$

end for

$h_{i+1,i} = \|\boldsymbol{w}_i\|_2$

$\boldsymbol{v}_{i+1} = \boldsymbol{w}_i / h_{i+1,i}$

Find $\boldsymbol{y}_i \in \mathbf{R}^i$ which minimizes $\|\tilde{\boldsymbol{r}}_i\|_2 = \| \, \|\tilde{\boldsymbol{r}}_0\|_2 \, \boldsymbol{e}_i - \bar{H}_i \, \boldsymbol{y} \, \|_2$

$\boldsymbol{x}_i = \boldsymbol{x}_0 + [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i]\boldsymbol{y}_i$

$\boldsymbol{r}_i = \boldsymbol{b} - A\boldsymbol{x}_i$

if $\|A^\mathsf{T}\boldsymbol{r}_i\|_2 < \epsilon$ stop

end for

Go to $*$.

Does BA-GMRES method give the least squares solution without breakdown ?

Noting

Theorem 13 *If* $\mathcal{R}(A) = \mathcal{R}(B^\top)$, *then*

$$\mathcal{R}(BA) = \mathcal{R}(A^\top B^\top) \iff \mathcal{R}(A^\top) = \mathcal{R}(B). \quad \square$$

Theorem 14

If $\mathcal{R}(A) = \mathcal{R}(B^\top)$, *then*

BA-GMRES method determines

a least squares solution of $\min_{\boldsymbol{x} \in \mathbf{R}^m} ||\boldsymbol{b} - A\boldsymbol{x}||_2$

${}^\forall \boldsymbol{b} \in \mathbf{R}^m$, ${}^\forall \boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown*

$\Updownarrow$

$\mathcal{R}(A^\top) = \mathcal{R}(B)$. $\square$

Corollary 15

$\mathcal{R}(A) = \mathcal{R}(B^\top)$, $\mathcal{R}(A^\top) = \mathcal{R}(B)$

$\Downarrow$

BA-GMRES method determines

a least squares solution of $\min_{\boldsymbol{x} \in \mathbf{R}^m} ||\boldsymbol{b} - A\boldsymbol{x}||_2$

${}^\forall \boldsymbol{b} \in \mathbf{R}^m$, ${}^\forall \boldsymbol{x}_0 \in \mathbf{R}^n$ *without breakdown.* $\square$

Summary on condition for B

General case: rank$A \leq$ min(m, n),

$\mathcal{R}(A) = \mathcal{R}(B^\top),\ \mathcal{R}(A^\top) = \mathcal{R}(B)$

$$\Downarrow$$

AB-GMRES, BA-GMRES methods give
a least squares solution of $\min_{\boldsymbol{x} \in \mathbf{R}^m} ||\boldsymbol{b} - A\boldsymbol{x}||_2$
$^\forall \boldsymbol{b} \in \mathbf{R}^m$, $^\forall \boldsymbol{x}_0 \in \mathbf{R}^n$ without breakdown.

$\cdots$ e.g. Let $B = \alpha A^\top$ where $0 \neq \alpha \in \mathbf{R}$.

Full rank case

$m \geq n = \text{rank} A$ (**over-determined case**)

Let $B = CA^{\mathsf{T}}$ where $C \in \mathbf{R}^{n \times n}$:nonsingular
$(B, A^{\mathsf{T}} \in \mathbf{R}^{n \times m})$.

$$\Downarrow$$

$$B^{\mathsf{T}} = AC^{\mathsf{T}}$$

$$\Downarrow$$

$$\mathcal{R}(A) = \mathcal{R}(B^{\mathsf{T}}).$$

$$\Downarrow$$

$$n = \text{rank} A^{\mathsf{T}} = \text{rank} A = \text{rank} B^{\mathsf{T}} = \text{rank} B$$

$$\Downarrow$$

$$\mathcal{R}(A^{\mathsf{T}}) = \mathcal{R}(B) = \mathbf{R}^n.$$

Since $AB \in \mathbf{R}^{m \times m}$, $BA \in \mathbf{R}^{n \times n}$, $m \geq n$,
use BA-GMRES method.
e.g. Let $C := \{\text{diag}(A^{\mathsf{T}} A)\}^{-1}$
i.e. $BA = \{\text{diag}(A^{\mathsf{T}} A)\}^{-1} A^{\mathsf{T}} A$.

Convergence Analysis

Theorem 16 *Let* $A \in \mathbf{R}^{m \times n}$, $m \geq n$,
$B := CA^{\mathsf{T}}$, $C \in \mathbf{R}^{n \times n}$ *: sym.pos.def.,*
$\sigma_i\,(1 \leq i \leq n)$*: singular values of* $\tilde{A} := AC^{\frac{1}{2}}$*.*
Then,
$\sigma_i^2\,(1 \leq i \leq n)$ *: eigenvalues of* AB *and* BA*.*
If $m > n$*, all other eigenvalues of* AB *are* 0*.*

[Proof]

Let $\tilde{A} := AC^{\frac{1}{2}} = U\Sigma V^\top$: SVD,

$U \in \mathbf{R}^{m\times m}, V \in \mathbf{R}^{n\times n}$: orthogonal matrices,

$$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & 0 \\ & & \sigma_n \\ & & \\ 0 & & \end{bmatrix} \in \mathbf{R}^{m\times n},$$

$\sigma_1 \geq \ldots \geq \sigma_n \geq 0$: singular values of $\tilde{A}$.

Then,

$$AB = ACA^\top = \tilde{A}\,\tilde{A}^\top = U\Sigma\Sigma^\top U^\top,$$

$$BA = CA^\top A = C^{\frac{1}{2}}\tilde{A}^\top\tilde{A}C^{-\frac{1}{2}} = C^{\frac{1}{2}}V\Sigma^\top\Sigma(C^{\frac{1}{2}}V)^{-1}. \quad \blacksquare$$

cf. If rank$A = n$, $C := \{\text{diag}(A^\top A)\}^{-1}$: sym.pos.def.

Similarly for RIF.

Theorem 17 *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *achieved by the* k*-th step of AB-GMRES satisfies*

$$\|\boldsymbol{r}_k|_{\mathcal{R}(A)}\|_2 \leq 2 \left(\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} \right)^k \|\boldsymbol{r}_0|_{\mathcal{R}(A)}\|_2. \quad \blacksquare$$

Theorem 18 *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *achieved by the* k*-th step of BA-GMRES satisfies*

$$\|B\boldsymbol{r}_k\|_2 = \|CA^{\mathsf{T}}\boldsymbol{r}_k\|_2 \leq 2\sqrt{\kappa(C)} \left(\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} \right)^k \|B\boldsymbol{r}_0\|_2. \quad \blacksquare$$

CGLS method

Preconditionings
$LDL^\top \sim A^\top A, \tilde{L} = LD^{\frac{1}{2}}$ (e.g. RIF)
$\mathrm{diag}(A^\top A) \sim A^\top A, \tilde{L} = (\mathrm{diag}(A^\top A))^{\frac{1}{2}}$.

Apply CG to

$$A'\boldsymbol{x}' = \boldsymbol{b}' \quad (*)$$

where

$$A' = \tilde{L}^{-1} A^\top A \tilde{L}^{-\top}$$

$$\boldsymbol{x}' = \tilde{L}^\top \boldsymbol{x}$$

$$\boldsymbol{b} = \tilde{L}^{-1} A^\top \boldsymbol{b}.$$

Since

$$(\tilde{L}^\top)^{-1} A' (\tilde{L}^\top) = C A^\top A = BA,$$

$$\lambda_i(A') = \lambda_i(BA) = {\sigma_i}^2,\ i = 1, \ldots, n$$

Theorem 19 *The residual $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ of the k-th step of preconditioned CGLS (CG applied to $(*)$) satisfies*

$$\|A^\top \boldsymbol{r}_k\|_{(A^\top A)^{-1}} = \|e_m\|_{A^\top A} \le 2 \left(\frac{\sigma_1 - \sigma_n}{\sigma_1 + \sigma_n} \right)^k \|A^\top \boldsymbol{r}_0\|_{(A^\top A)^{-1}}. \quad \blacksquare$$

Similar convergence behaviours expected for
AB-GMRES, BA-GMRES,
and preconditioned CGLS methods,
using the same preconditioner C.

rank$A = m \leq n$ (under-determined case)

Let $B = A^\top C$ where $C \in \mathbf{R}^{m\times m}$:nonsingular
$(B, A^\top \in \mathbf{R}^{n\times m})$.

$$\Downarrow$$

$$\mathcal{R}(A^\top) = \mathcal{R}(B).$$

$$\Downarrow$$

$m = \text{rank}A = \text{rank}A^\top = \text{rank}B = \text{rank}B^\top$

$$\Downarrow$$

$$\mathcal{R}(A) = \mathcal{R}(B^\top) = \mathbf{R}^m.$$

Since $AB \in \mathbf{R}^{m\times m}$, $BA \in \mathbf{R}^{n\times n}$, $m \leq n$,
use AB-GMRES method.

e.g. Let $C := \{\text{diag}(AA^\top)\}^{-1}$.
i.e. $AB = AA^\top\{\text{diag}(AA^\top)\}^{-1}$.

Note that when rank$A = m, \mathcal{R}(A) = \mathbf{R}^m \ni \boldsymbol{b}$,
so that

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\| = \min_{\boldsymbol{z} \in \mathbf{R}^m} \|\boldsymbol{b} - AB\boldsymbol{z}\| = \min_{\boldsymbol{z} \in \mathbf{R}^m} \|\boldsymbol{b} - AA^\mathsf{T} C\boldsymbol{z}\| = 0.$$

Hence, AB-GMRES method with $B = A^\mathsf{T} C$
gives the minimum norm least squares
solution

$$\boldsymbol{x}^* = B\boldsymbol{z}^* = A^\mathsf{T}(C\boldsymbol{z}^*)$$

of

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|$$

since

$$AA^\mathsf{T}(C\boldsymbol{z}^*) = \boldsymbol{b}.$$

Theorem 20

Let $A \in \mathbf{R}^{m\times n}$, $m \le n$,

$B := A^\top C,\ C \in \mathbf{R}^{m\times m}$ *: sym.pos.def.,*

$\sigma_i\,(1 \le i \le m)$*: singular values of* $\tilde{A} := C^{\frac{1}{2}}A$*.*

Then,

$\sigma_i^2\,(1 \le i \le m)$ *: eigenvalues of* AB *and* BA*.*

If $m < n$*, all other eigenvalues of* BA *are* 0*.*

[Proof]

Let $\tilde{A} := C^{\frac{1}{2}} A = U \mathbf{\Sigma} V^{\mathsf{T}}$: SVD,

$U \in \mathbf{R}^{m \times m}, V \in \mathbf{R}^{n \times n}$: orthogonal matrices,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{bmatrix} \in \mathbf{R}^{m \times n},$$

$\sigma_1 \geq \ldots \geq \sigma_m \geq 0$: singular values of $\tilde{A}$.

Then,

$$AB = AA^{\mathsf{T}}C = C^{-\frac{1}{2}} \tilde{A}\, \tilde{A}^{\mathsf{T}} C^{\frac{1}{2}} = C^{-\frac{1}{2}} U \mathbf{\Sigma} \mathbf{\Sigma}^{\mathsf{T}} (C^{-\frac{1}{2}} U)^{-1},$$

$$BA = A^{\mathsf{T}} C A = \tilde{A}^{\mathsf{T}} \tilde{A} = V \mathbf{\Sigma}^{\mathsf{T}} \mathbf{\Sigma} V^{\mathsf{T}}. \quad \blacksquare$$

cf. If rank$A = m$,

$C := \{\text{diag}(AA^{\mathsf{T}})\}^{-1}$: sym.pos.def.

Similarly for RIF.

Theorem 21 *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *achieved by the* k*-th step of AB-GMRES satisfies*

$$\|\boldsymbol{r}_k\|_2 \leq 2\sqrt{\kappa_2(C)} \left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \|\boldsymbol{r}_0\|_2. \quad \blacksquare$$

Theorem 22 *The residual* $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$ *achieved by the* k*-th step of BA-GMRES satisfies*

$$\|B\boldsymbol{r}_k|_{\mathcal{R}(B)}\|_2 \leq 2 \left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \|B\boldsymbol{r}_0|_{\mathcal{R}(B)}\|_2. \quad \blacksquare$$

Preconditioned CGLS

$$C = (\tilde{L}\tilde{L}^\top)^{-1} \in \mathbf{R}^{m\times m}, SPD$$

Precondition

$$AA^\top \boldsymbol{y} = \boldsymbol{b}$$

as

$$A'\boldsymbol{y}' = \boldsymbol{b}' \quad (**)$$

where

$$A' = \tilde{L}^{-1}AA^\top\tilde{L}^{-\top},$$

$$\boldsymbol{y}' = \tilde{L}^\top\boldsymbol{y},$$

$$\boldsymbol{b}' = \tilde{L}^{-1}\boldsymbol{b}.$$

Since

$$\tilde{L}A'\tilde{L}^{-1} = AA^\top C = AB,$$

$$\lambda_i(A') = \lambda_i(AB) = {\sigma_i}^2,\ i = 1, \ldots, m.$$

Theorem 23 *The residual* $\boldsymbol{r} = \boldsymbol{b} - AA^\top \boldsymbol{y}$ *of the* k*-th step of preconditioned CGLS (CG applied to* $(**)$ *) satisfies*

$$\|\boldsymbol{r}_k\|_{(AA^\top)^{-1}} = \|e_m\|_{AA^\top} \leq 2\left(\frac{\sigma_1 - \sigma_m}{\sigma_1 + \sigma_m}\right)^k \|\boldsymbol{r}_0\|_{(AA^\top)^{-1}}. \quad \blacksquare$$

Similar convergence behaviours expected for
AB-GMRES, BA-GMRES,
and preconditioned CGLS methods,
using the same preconditioner C.

Choice of B

Besides satisfying
$\mathcal{R}(A) = \mathcal{R}(B^\top)$ and $\mathcal{R}(A^\top) = \mathcal{R}(B)$,

B should satisfy $AB \approx \mathrm{I}_m$ or $BA \approx \mathrm{I}_n$.

Simple choice:

$$B = CA^\top,\ C = \{\mathrm{diag}(A^\top A)\}^{-1}$$
when $m \geq n = \mathrm{rank} A$,

$$B = A^\top C, C := \{\mathrm{diag}(AA^\top)\}^{-1}$$
when $\mathrm{rank} A = m \leq n$.

Application of
Robust Incomplete Factorization (RIF)
(Benzi and Tuma, '03)

$$A^\top A \approx LDL^\top,$$

$$A^\top A \approx Z^{-\top} D Z^{-1},$$

where $A \in \mathbf{R}^{m\times n}$ with $m \geq n = \mathrm{rank} A$,
Z: upper triangular matrix,
L: lower triangular matrix.
Use of (incomplete) $A^\top A$-orthogonalization.
Low memory requirement.

RIF Method

```
Let Z = F = [e_1, e_2, ..., e_n]
For j = 1, ..., n Do
  Compute u_j = A z_j
  Compute d_j = (u_j, u_j)
  For i = j + 1, ..., n Do:
    Compute v_i = A e_i
    Compute θ_ij = (v_i, u_j) / d_j
    IF θ_ij > τ
      Store L(i, j) = θ_ij
    EndIF
    Compute z_i = z_i − θ_ij z_j
    Drop the elements in z_i smaller than τ
  EndDo
EndDo
```

Where:

- Let $Z = F = [\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}]$
- Compute $\mathbf{u_j} = \mathbf{A}\mathbf{z_j}$
- Compute $d_j = (\mathbf{u_j}, \mathbf{u_j})$
- For $i = j + 1, \ldots, n$ Do:
- Compute $\mathbf{v_i} = \mathbf{A}\mathbf{e_i}$
- Compute $\theta_{ij} = \frac{(\mathbf{v_i}, \mathbf{u_j})}{d_j}$
- IF $\theta_{ij} > \tau$
- Store $L(i, j) = \theta_{ij}$
- Compute $\mathbf{z_i} = \mathbf{z_i} - \theta_{\mathbf{ij}}\mathbf{z_j}$
- Drop the elements in $\mathbf{z_i}$ smaller than τ

Numerical experiments

PC: Dell Precision 690

CPU: 3.00 GHz, Memory: 16 GB

Program, compiler: GNU C/C++ 3.4.3

Linear least squares problems

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2, \quad A \in \mathbf{R}^{m \times n}$$

1. Over-determined case ($m \geq n$)

Test matrices A :

- Generated by MATLAB command "sprandn" .
- Specified density and condition number.
- Value of nonzero elements:
  random (normal distribution)
- Pattern of nonzero elements: random.

$\boldsymbol{b}$: also random, inconsistent ($\boldsymbol{b} \notin \mathcal{R}(A)$).

Convergence judged by $\dfrac{\|A^\top \boldsymbol{r}\|_2}{\|A^\top \boldsymbol{b}\|_2}$, $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$.

$\boldsymbol{x}_0 = \boldsymbol{0}$.

Test matrices (L)

$m = 30,000$, $n = 3,000$

density: 0.1%

Name	Condition number
RANDL1	1.9×10
RANDL2	1.6×10^2
RANDL3	1.3×10^3
RANDL4	2.0×10^4
RANDL5	1.3×10^5
RANDL6	1.3×10^6
RANDL7	1.3×10^7

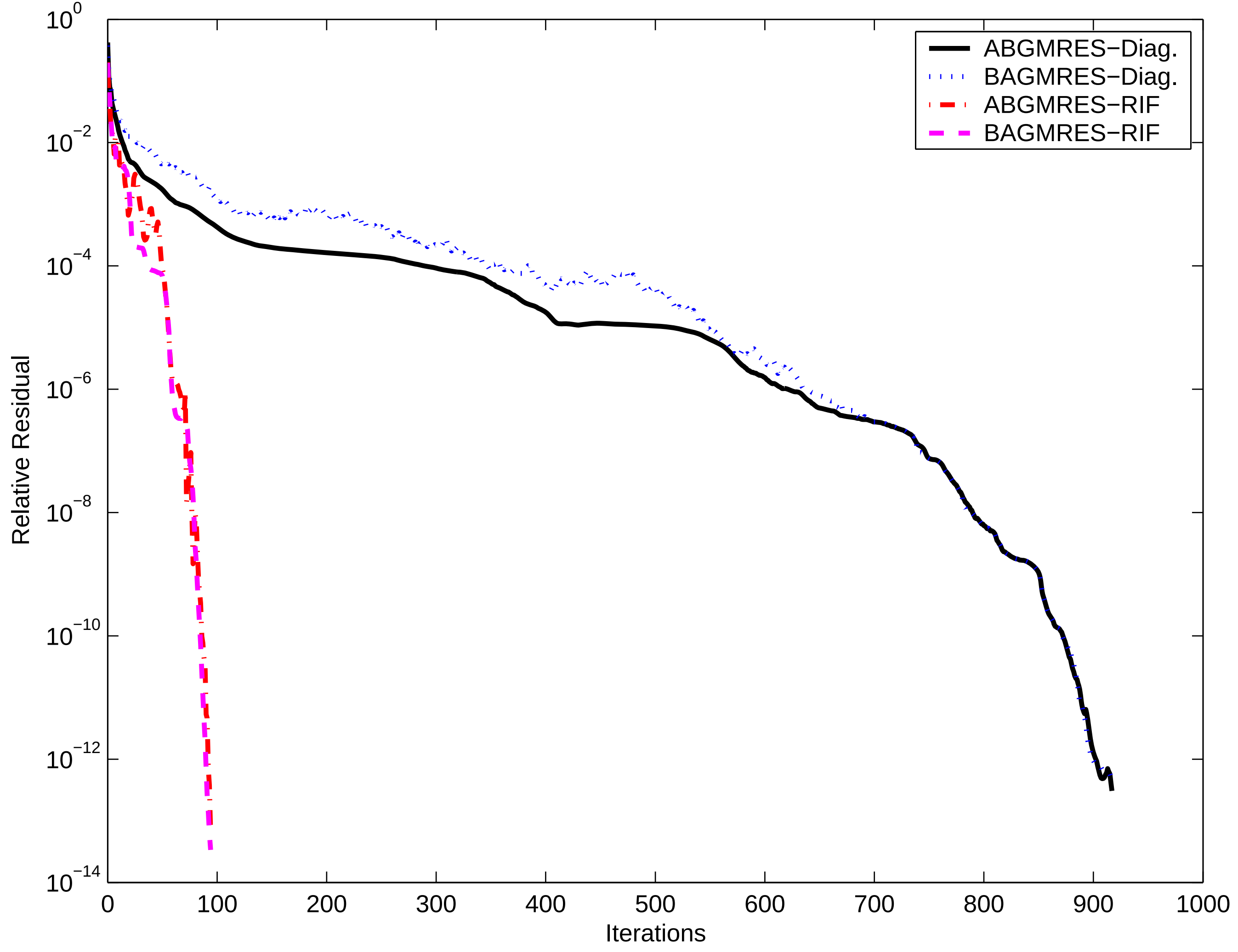
Fig. 1

Comparison of full AB-GMRES and BA-GMRES

with diagonal-scaling
and RIF preconditioner ($\tau = 0.8$)

(RANDL3)

($\|A^\mathsf{T}\boldsymbol{r}\|_2/\|A^\mathsf{T}\boldsymbol{b}\|_2$ vs. iterations)



ABGMRES−Diag.
BAGMRES−Diag.
ABGMRES−RIF
BAGMRES−RIF
Relative Residual
Iterations
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
0
100
200
300
400
500
600
700
800
900
1000

The effect of the restart period k in the BA-GMRES(k)-RIF method (over-determined problem).

RANDL5	k	100	140	180	220	260	$\geq$ 295
($\tau = 0.8$)	iter	755	668	587	498	453	295
	time	25.05	24.81	24.14	22.62	22.15	$*$14.27
RANDL6	k	160	200	240	280	300	$\geq$ 318
($\tau = 0.07$)	iter	2,470	2,085	1,856	1,311	599	318
	time	144.68	130.08	121.72	92.88	48.92	$*$26.73
RANDL7	k	200	280	320	340	360	$\geq$ 362
($\tau = 0.02$)	iter	2,594	2,168	1,567	900	506	362
	time	179.83	165.79	129.07	80.13	50.99	$*$37.26

k: restart period, iter: number of iterations

time: computation time (sec.)

Convergence criterion: $\|A^{\mathsf{T}}\boldsymbol{r}\|_2 / \|A^{\mathsf{T}}\boldsymbol{b}\|_2 < 10^{-6}$.

The effect of the RIF parameter τ
for problem RANDL3.

	τ	0.9	0.8	0.7	0.6	0.5	0
CGLS-RIF	iter	83	72	77	65	67	1
	time	5.39	$*$5.38	5.42	5.43	5.59	68.00
LSQR-RIF	iter	83	73	78	66	68	1
	time	5.39	$*$5.38	5.42	5.43	5.60	68.00
RCGLS-RIF	iter	79	70	73	64	66	1
	time	5.99	5.86	5.96	5.85	6.08	67.99
BA-GMRES-RIF	iter	70	60	73	64	66	1
	time	5.54	5.47	5.64	5.59	5.63	67.99

iter: number of iterations

time: computation time (sec.)

Convergence criterion: $\|A^\mathsf{T}\boldsymbol{r}\|_2/\|A^\mathsf{T}\boldsymbol{b}\|_2 < 10^{-6}$.

The effect of the RIF parameter τ for problem RANDL6.

	τ	0.09	0.08	0.07	0.06	0.05	0
CGLS-RIF	iter	670	674	615	706	658	1
	time	34.38	35.39	33.67	38.17	37.31	60.00
LSQR-RIF	iter	680	685	645	736	701	1
	time	34.87	35.50	34.15	37.21	37.41	60.00
RCGLS-RIF	iter	333	324	317	313	307	1
	time	34.11	33.54	33.22	33.39	33.48	60.01
BA-GMRES-RIF	iter	335	325	318	315	309	1
	time	27.03	26.77	*26.73	27.16	27.45	60.01

iter: number of iterations

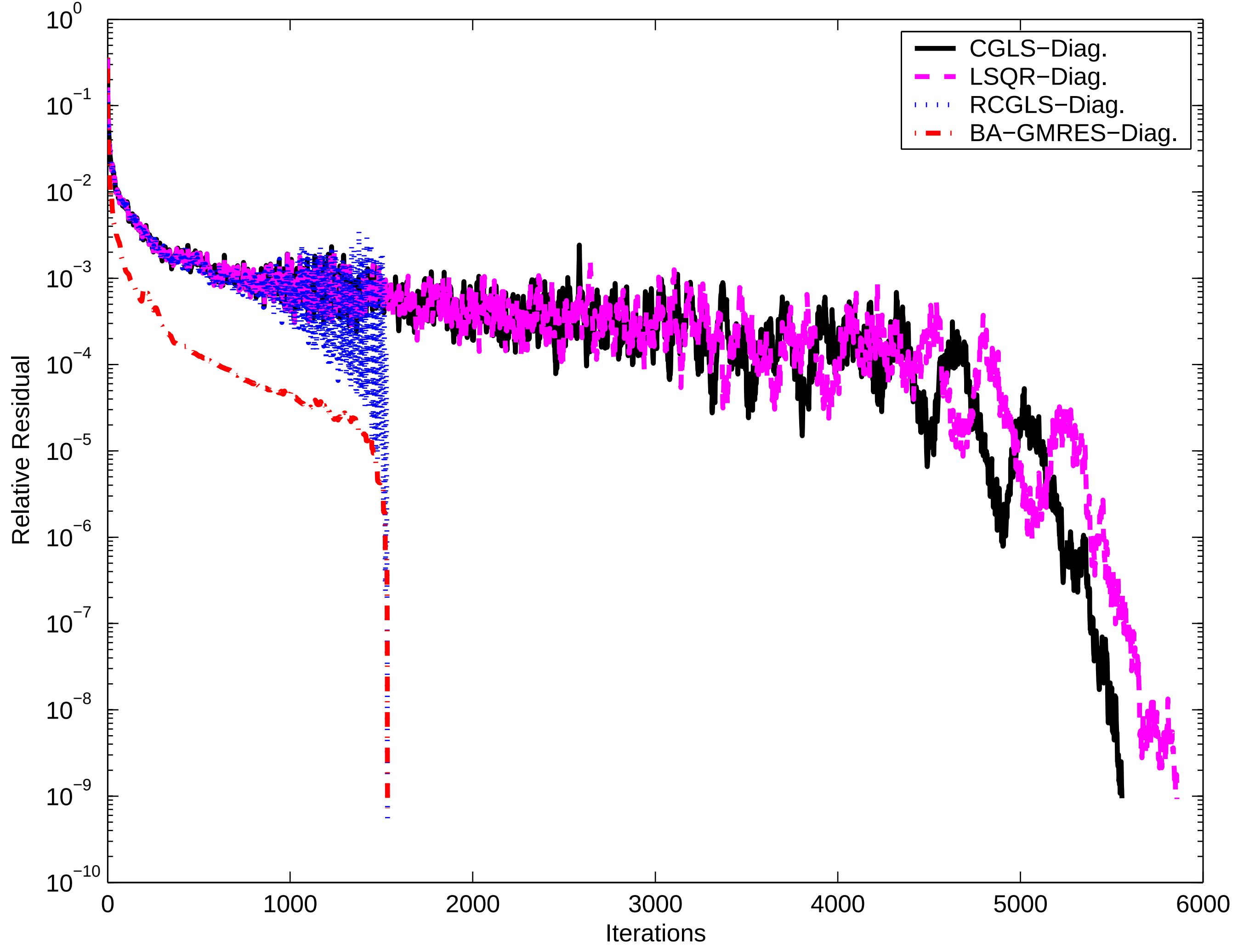
time: computation time (sec.)

Convergence criterion: $\|A^{\mathsf{T}}\boldsymbol{r}\|_2/\|A^{\mathsf{T}}\boldsymbol{b}\|_2 < 10^{-6}$.

Comparison of convergence

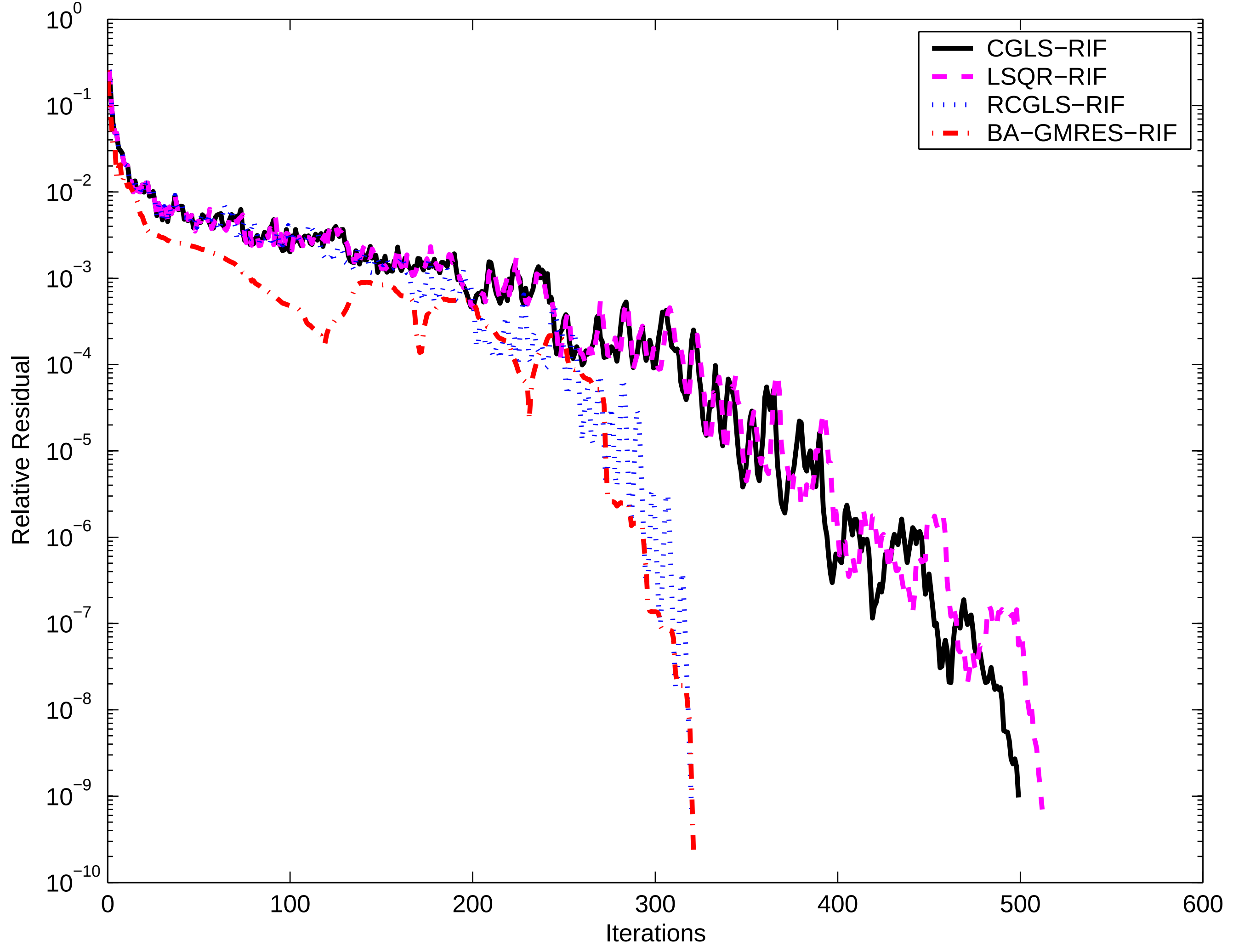
$\|A^{\mathsf{T}}\boldsymbol{r}\|_2 / \|A^{\mathsf{T}}\boldsymbol{b}\|_2$ vs. number of iterations

Fig.2: Diagonal scaling (RANDL5)



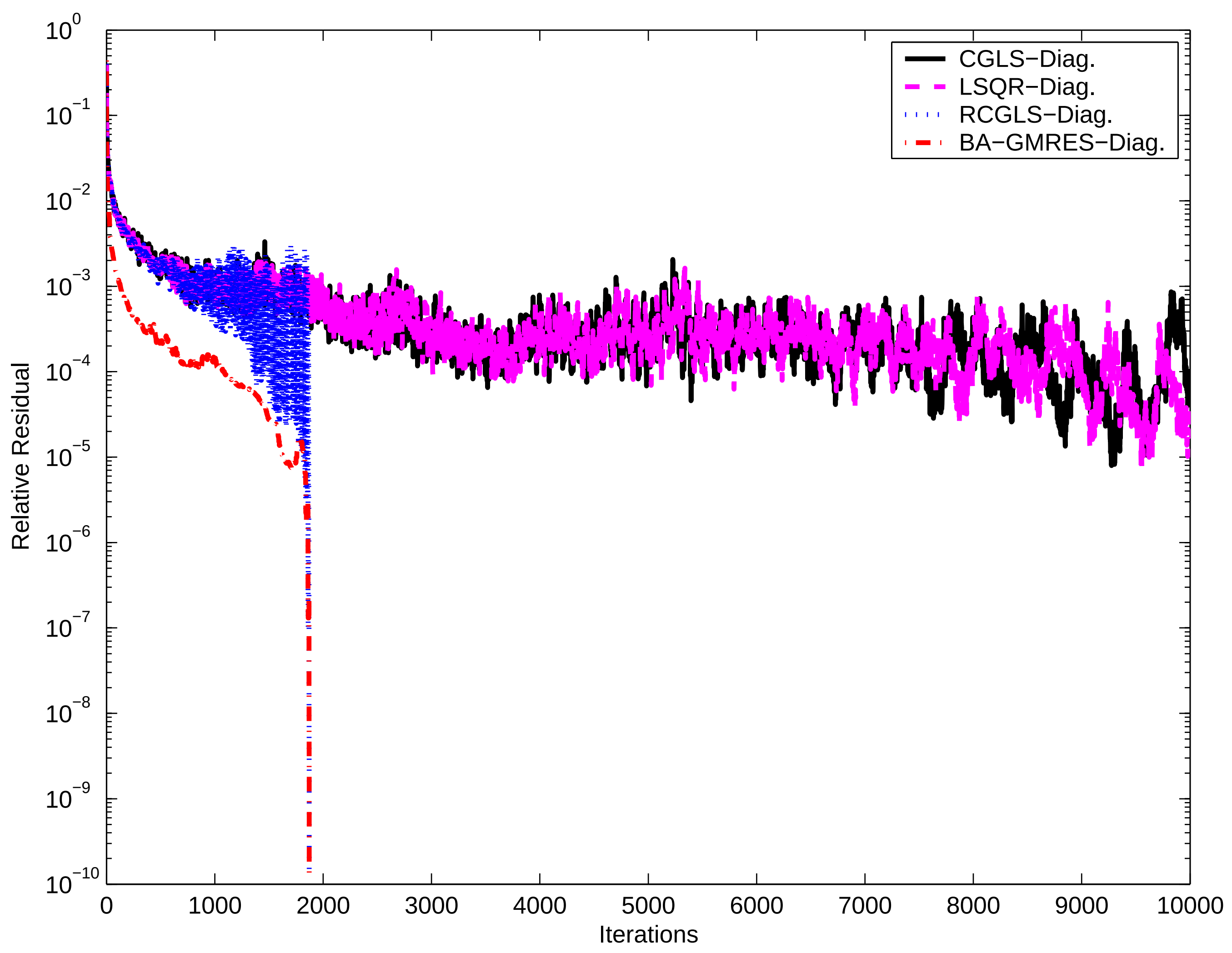
CGLS−Diag.
LSQR−Diag.
RCGLS−Diag.
BA−GMRES−Diag.
Relative Residual
Iterations
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
0
1000
2000
3000
4000
5000
6000

Fig.3: RIF (RANDL5)



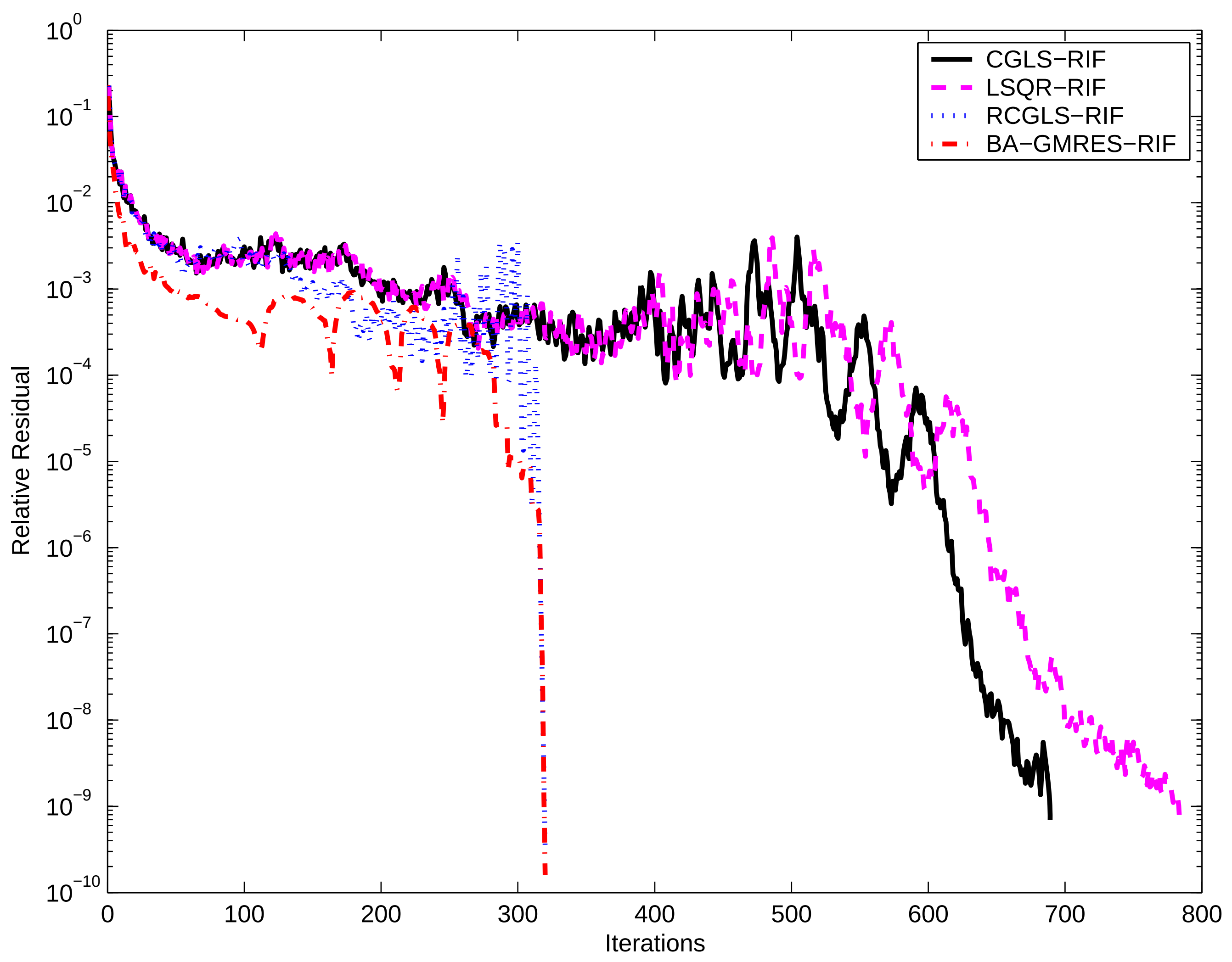
CGLS−RIF
LSQR−RIF
RCGLS−RIF
BA−GMRES−RIF
Relative Residual
Iterations
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
0
100
200
300
400
500
600

Fig.4: Diagonal scaling (RANDL6)



Relative Residual
Iterations
CGLS−Diag.
LSQR−Diag.
RCGLS−Diag.
BA−GMRES−Diag.

Fig.5: RIF (RANDL6)



CGLS−RIF
LSQR−RIF
RCGLS−RIF
BA−GMRES−RIF
Relative Residual
Iterations

Comparison of CPU time

	CGLS		LSQR		RCGLS		BA-GMRES	
	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF
RANDL1	35	14	35	14	35	14	35	14
($\tau = 0.5$)	∗0.10	4.98	∗0.10	4.98	0.16	4.99	0.14	4.99
RANDL2	214	21	214	21	208	21	193	21
($\tau = 0.7$)	∗0.61	5.10	∗0.61	5.10	2.77	5.13	2.28	5.11
RANDL3	742	72	740	73	697	70	622	60
($\tau = 0.8$)	2.08	5.38	∗2.07	5.38	26.30	5.62	20.70	5.47
RANDL4	1,147	85	1,154	85	1,062	84	1,069	82
($\tau = 0.5$)	∗3.22	6.38	3.38	6.38	59.35	7.17	59.41	6.62
RANDL5	4,897	470	5,064	401	1,521	305	1,522	299
($\tau = 0.9$)	13.74	∗12.78	14.03	11.79	119.70	14.42	118.87	13.91
RANDL6	10,551	615	11,088	645	1,861	317	1,862	318
($\tau = 0.07$)	29.65	33.67	30.21	34.15	177.94	26.93	176.93	∗26.73
RANDL7	32,143	1,951	35,034	2,443	1,914	371	1,899	362
($\tau = 0.02$)	89.93	102.28	91.31	128.40	195.63	40.07	183.90	∗37.26

First row: number of iterations

Second row: computation time (seconds).

Convergence criterion: $\|A^T r\|_2 / \|A^T b\|_2 < 10^{-6}$.

Problems from animal breeding studies
and meteorology (HIRLAM)

Name	m	n	nnz	density
SMALL	3,140	1,988	8,510	1.4%
MEDIUM	9,397	6,119	25,013	0.04%
LARGE	28,524	17,264	75,018	0.02%
VLARGE	174,193	105,882	463,303	0.003%
HIRLAM	1,385,270	452,200	2,718,200	0.0004%

	CGLS		LSQR		BAGMRES	
	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF
SMALL ($\tau = 0.2$)	125 $*$0.06	59 0.33	125 $*$0.06	59 0.33	124 0.54	58 0.49
MEDIUM ($\tau = 0.3$)	128 $*$0.20	64 2.10	128 $*$0.20	64 2.10	123 1.67	62 2.45
LARGE ($\tau = 0.5$)	133 $*$0.67	73 15.83	133 $*$0.67	73 15.83	131 5.81	71 17.93
VLARGE ($\tau = 0.8$)	171 6.00	137 36.56	170 $*$5.97	137 36.57	166 58.45	136 73.45
HIRLAM ($\tau = 0.9$)	180 $*$48.16	164 204.44	180 48.20	164 204.46	170 294.05	163 464.93

First row: number of iterations,

Second row: computation time (seconds).

Convergence criterion: $\|A^\top \boldsymbol{r}\|_2 / \|A^\top \boldsymbol{b}\|_2 < 10^{-6}$.

Explanation for better convergence of GMRES:

Modified Gram-Schmidt process in GMRES (explicit orthogonalization):
Robust against rounding error
esp. for ill-conditioned case

vs.

three-term recurrence in CG
(implicit orthogonalization)

2. Under-determined case ($m < n$)

$$\min_{\boldsymbol{x} \in \mathbf{R}^n} \|\boldsymbol{b} - A\boldsymbol{x}\|_2, \quad A \in \mathbf{R}^{m \times n} \quad (m < n).$$

Minimum norm least squares (pseudo-inverse) solution:

$$AA^\mathsf{T}\boldsymbol{y} = \boldsymbol{b}, \quad \boldsymbol{x} = A\boldsymbol{y}$$

Previous approach:

Apply (preconditioned) CG: CGNE

Test random matrices

RANDLnT: 3,000 $\times$30,000, density:1.5%

(transpose of RANDLn).

$\boldsymbol{b} = A\boldsymbol{x}^*$ where $\boldsymbol{x}^* = (1, \ldots, 1)^\mathsf{T}$.

Convergence judged by $\|\boldsymbol{r}\|_2 / \|\boldsymbol{b}\|_2$

(rank$A = m$: consistent system)

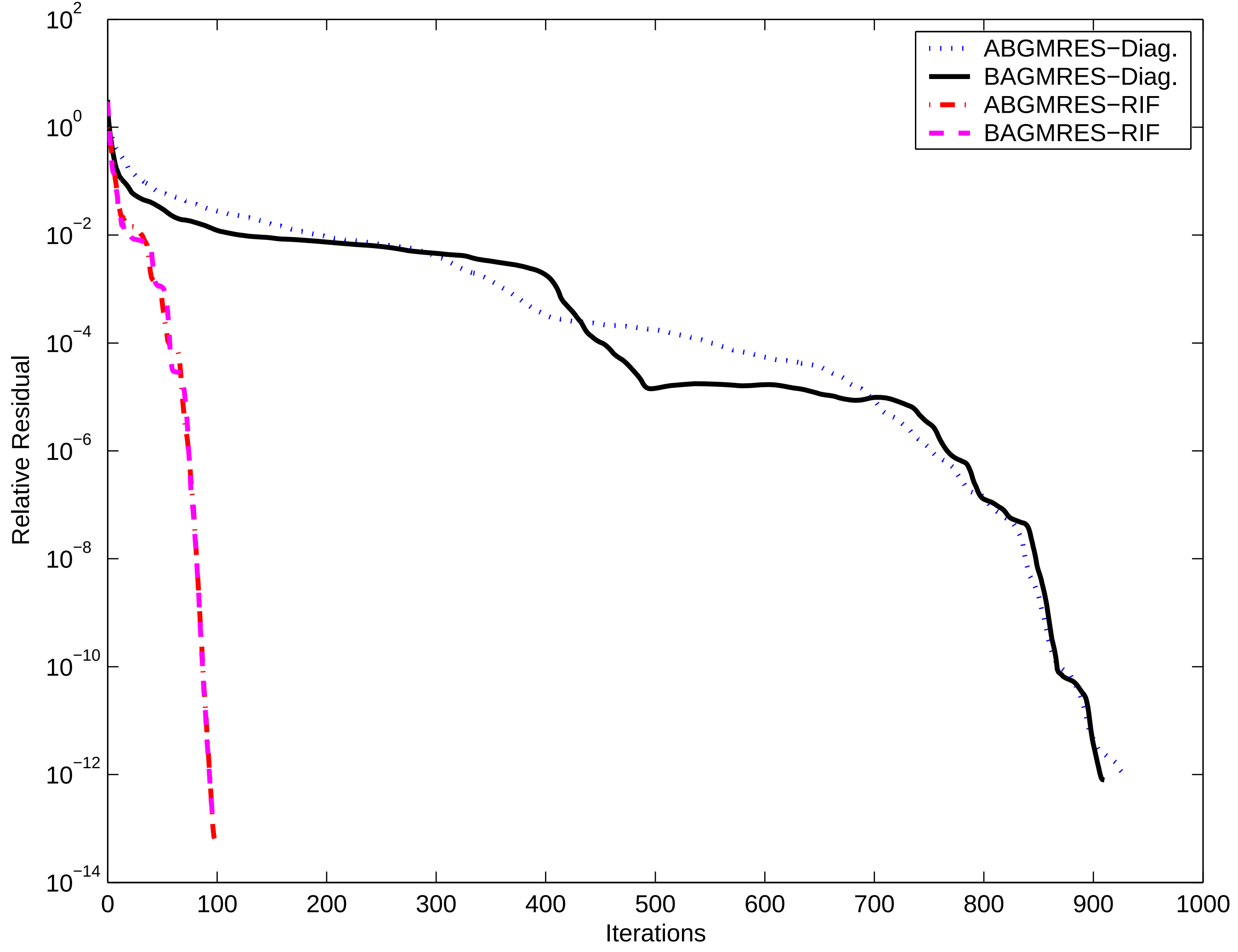
Fig. 6

Comparison of full AB-GMRES and BA-GMRES

with diagonal-scaling
and RIF preconditioner

(RANDL3T)

($\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ vs. iterations)

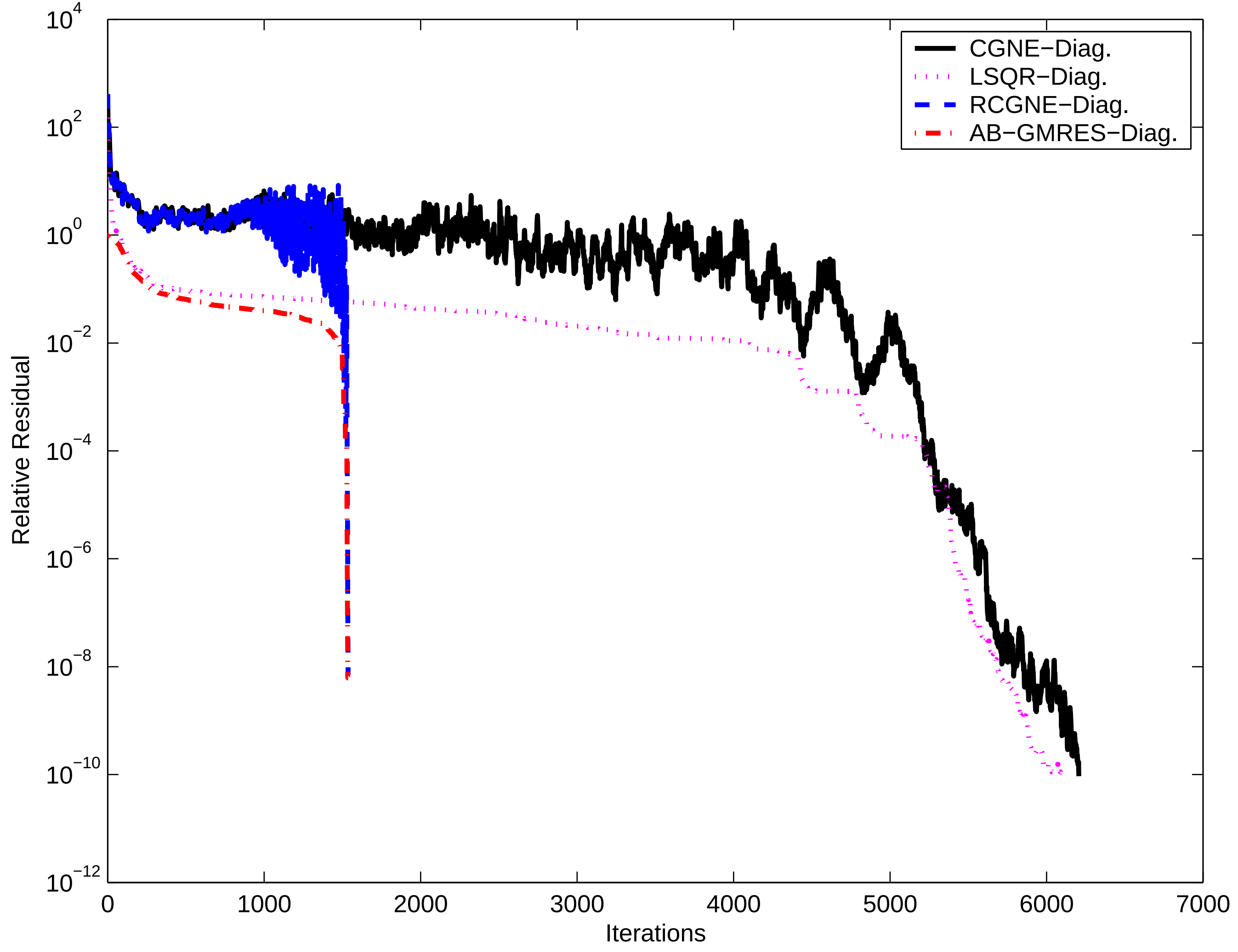


ABGMRES−Diag.
BAGMRES−Diag.
ABGMRES−RIF
BAGMRES−RIF
Relative Residual
Iterations
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
0
100
200
300
400
500
600
700
800
900
1000

Comparison of convergence

$\|\boldsymbol{r}\|_2/\|\boldsymbol{b}\|_2$ vs. number of iterations

Fig.7: Diagonal scaling (RANDL5T)



CGNE−Diag.
LSQR−Diag.
RCGNE−Diag.
AB−GMRES−Diag.
Relative Residual
Iterations
10^4
10^2
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
0
1000
2000
3000
4000
5000
6000
7000

Fig.8: RIF (RANDL5T)

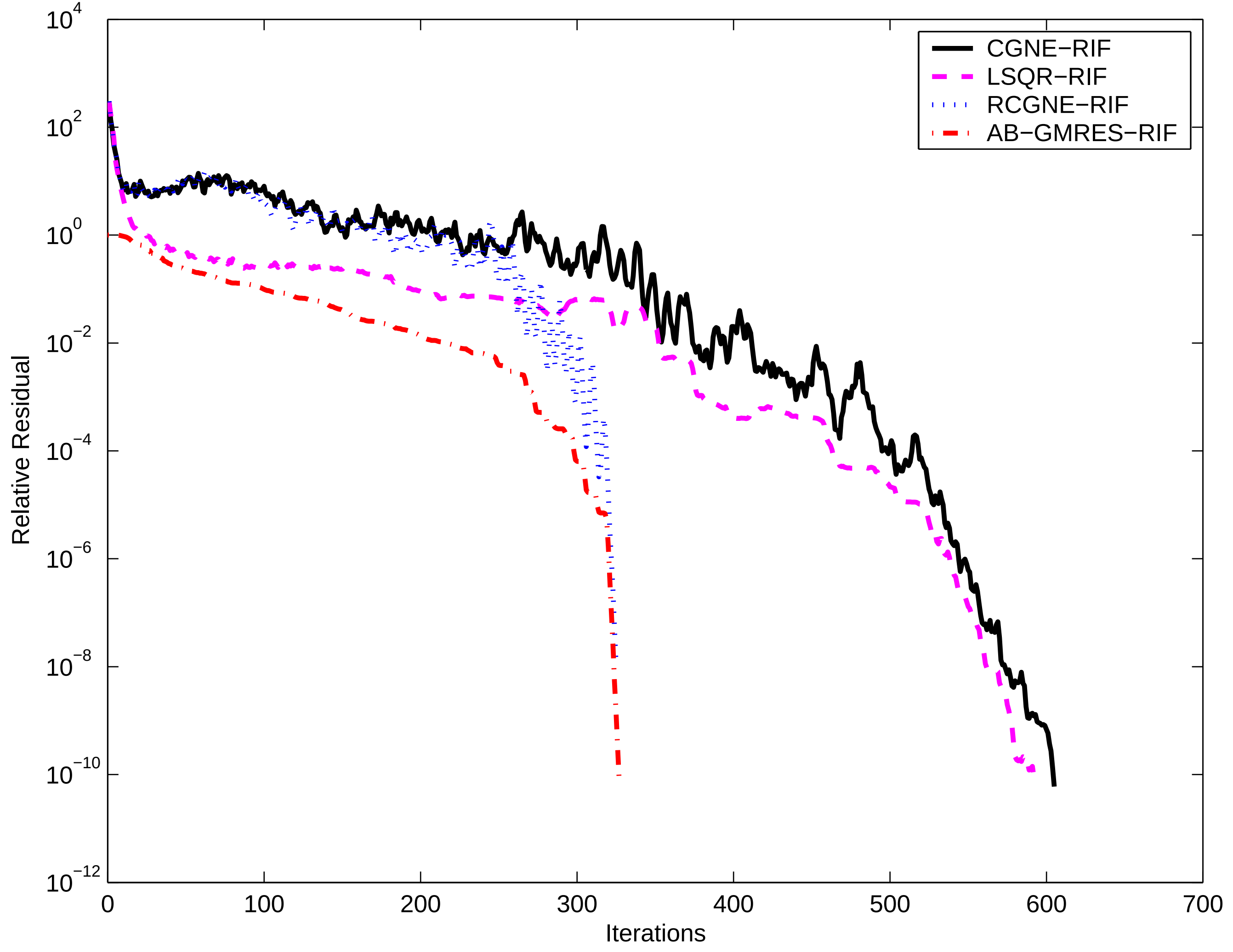
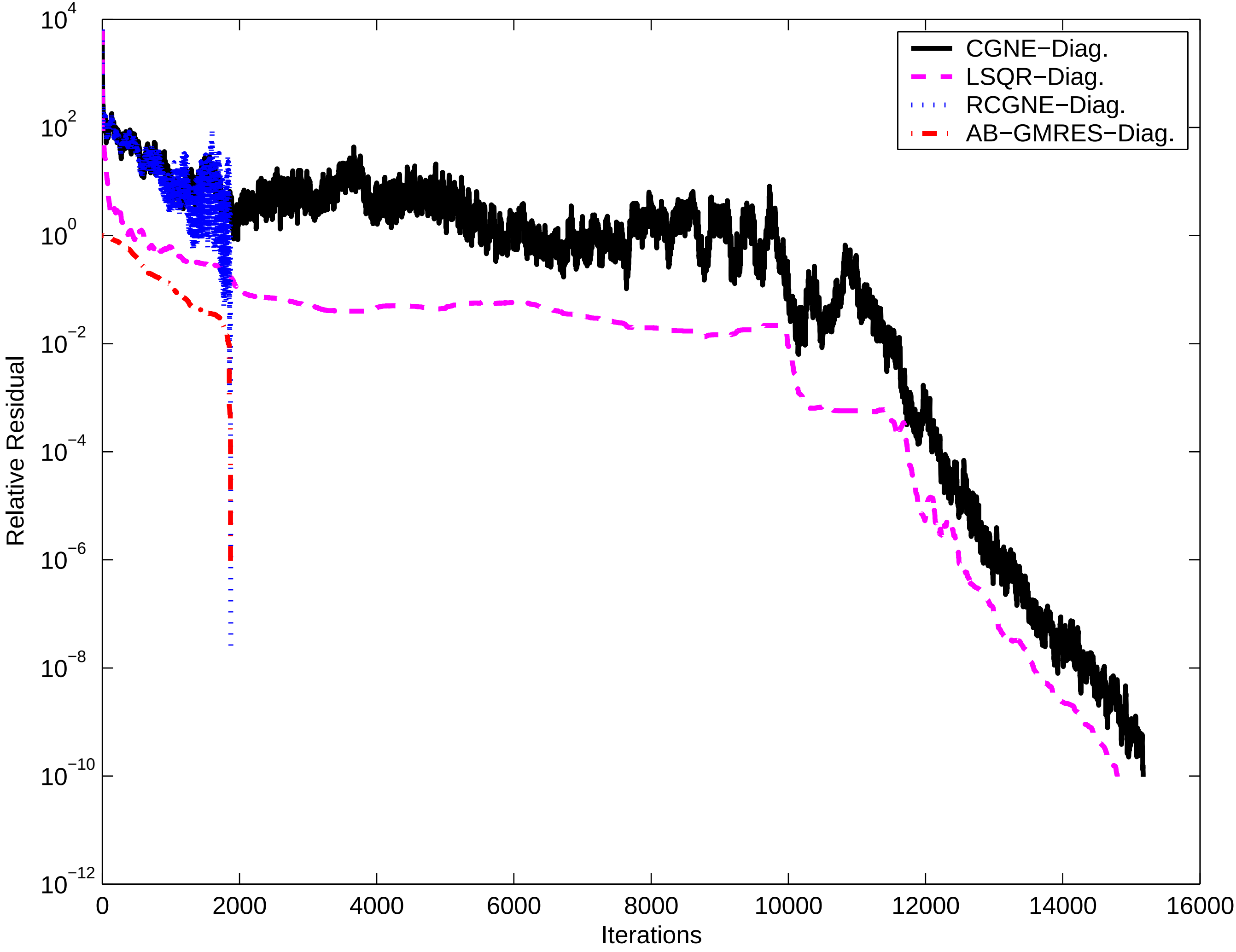
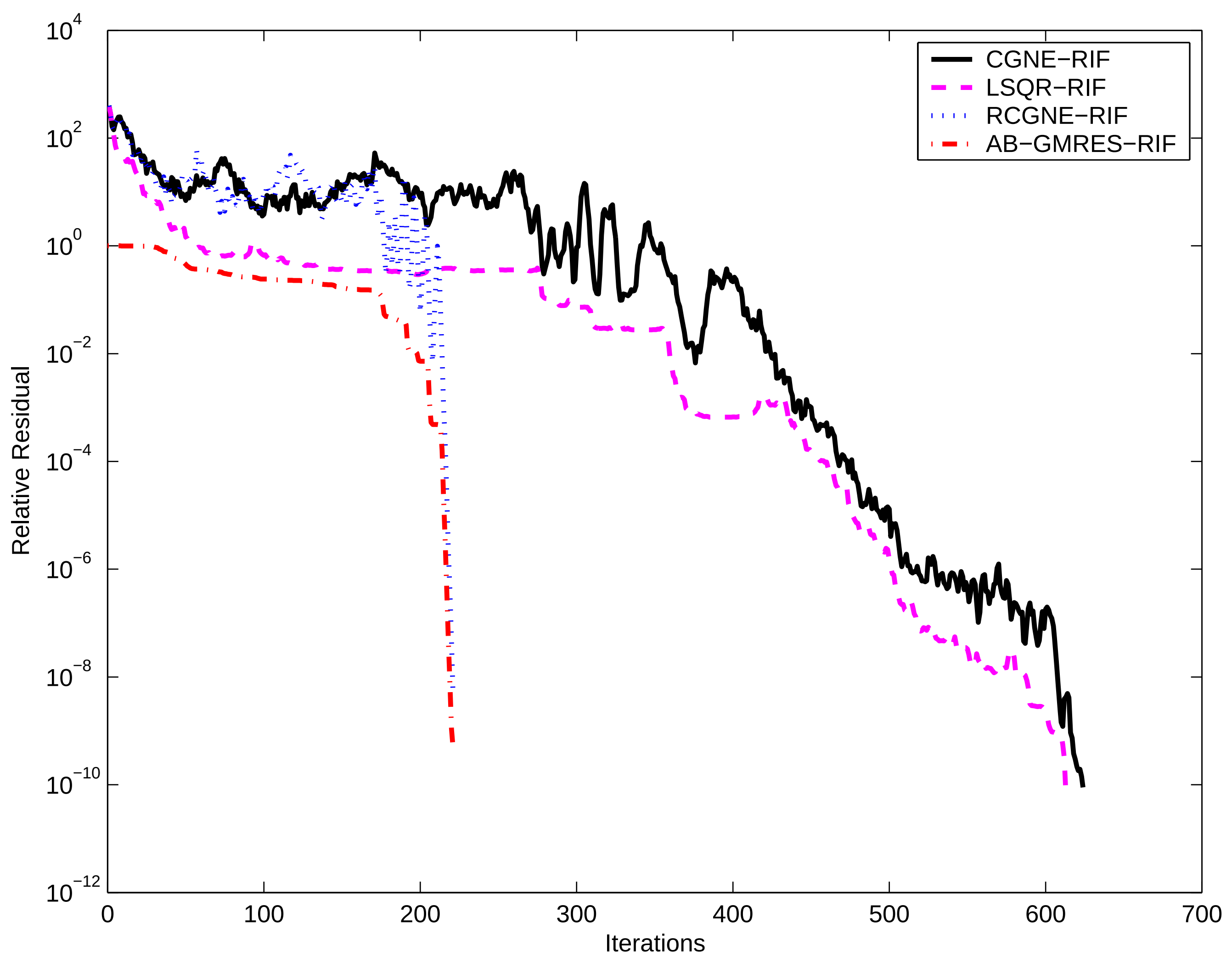

CGNE−RIF
LSQR−RIF
RCGNE−RIF
AB−GMRES−RIF
Relative Residual
Iterations
10^4
10^2
10^0
10^−2
10^−4
10^−6
10^−8
10^−10
10^−12
0
100
200
300
400
500
600
700

Fig.9: Diagonal scaling (RANDL6T)



CGNE−Diag.
LSQR−Diag.
RCGNE−Diag.
AB−GMRES−Diag.
Relative Residual
Iterations
10^{4}
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
0
2000
4000
6000
8000
10000
12000
14000
16000

Fig.10: RIF (RANDL6T)



CGNE−RIF
LSQR−RIF
RCGNE−RIF
AB−GMRES−RIF
Relative Residual
Iterations
10^{4}
10^{2}
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
0
100
200
300
400
500
600
700

Comparison of CPU time

	CGNE		LSQR		RCGNE		AB-GMRES	
	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF	-diag.	-RIF
RANDL1T	37	12	37	12	37	12	37	12
($\tau = 0.4$)	∗0.15	5.01	∗0.15	5.02	0.23	5.02	0.17	5.01
RANDL2T	259	25	256	25	252	25	247	25
($\tau = 0.7$)	∗1.06	5.16	∗1.06	5.16	4.22	5.17	3.75	5.17
RANDL3T	838	81	823	80	783	77	754	75
($\tau = 0.8$)	∗3.43	5.53	∗3.43	5.53	33.87	5.81	30.56	5.67
RANDL4T	1,464	116	1,407	114	1,223	106	1,187	106
($\tau = 0.5$)	5.97	6.94	∗5.96	6.94	79.67	7.45	73.67	7.21
RANDL5T	5,548	544	5,414	539	1,535	322	1,533	322
($\tau = 0.9$)	22.61	13.88	22.71	∗13.86	123.80	15.58	121.58	15.11
RANDL6T	12,837	514	12,486	502	1,873	219	1,871	218
($\tau = 0.01$)	52.38	39.04	50.10	38.99	182.51	27.43	179.85	∗26.99
RANDL7T	38,397	3,078	37,792	2,979	2,240	451	2,238	450
($\tau = 0.04$)	156.01	94.19	152.07	91.69	366.88	44.49	353.92	∗43.76

First row: number of iterations

Second row: computation time (seconds).

Convergence criterion: $\|r\|_2/\|b\|_2 < 10^{-6}$.

Memory requirements

Intermediate memory required for each method for k iterations.

	dim(m)	dim(n)	dim(k)	total
CGLS	$\boldsymbol{r}, A\boldsymbol{p}$	$\boldsymbol{x}, \boldsymbol{p}, A^\top \boldsymbol{r}$		$2m + 3n$
CGNE	$\boldsymbol{r}, A\boldsymbol{p}$	$\boldsymbol{x}, \boldsymbol{p}, A^\top \boldsymbol{r}$		$2m + 3n$
LSQR	$\boldsymbol{u}$	$\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}$		$m + 3n$
AB-GMRES(k)	$V, \boldsymbol{w}$	$\boldsymbol{x}$	$\boldsymbol{y}, \boldsymbol{e}$	$(k+1)m + n + 2k + k^2/2$
BA-GMRES(k)		$V, \boldsymbol{w}, \boldsymbol{x}$	$\boldsymbol{y}, \boldsymbol{e}$	$(k+2)n + 2k + k^2/2$

Conclusions

Proposed two methods

for least squares problems

using GMRES on

1. $\min_{\boldsymbol{z}\in\mathbf{R}^m} ||\boldsymbol{b} - AB\boldsymbol{z}||_2$,
2. $\min_{\boldsymbol{x}\in\mathbf{R}^n} ||B\boldsymbol{b} - BA\boldsymbol{x}||_2$.

Sufficient conditions for B:

$\mathcal{R}(B) = \mathcal{R}(A^\mathsf{T}),\ \ \mathcal{R}(B^\mathsf{T}) = \mathcal{R}(A)$.

Convergence analysis.

Numerical experiments:

GMRES faster than CGLS type methods

when used with RIF

for ill-conditioned problems.

Thank you!

Preprint:

http://research.nii.ac.jp/TechReports/07-009E.pdf