Namur, May 5, 1995

Reliable residual updates in CGS and other hybrid Bi-CG methods

by

Gerard L.G. Sleijpen

sleijpen@math.ruu.nl

joint work with

Henk A. van der Vorst

Department of Mathematics
University Utrecht
Utrecht, The Netherlands

Solve x from $\boxed{A\,x = b.}$

Residual update in hybrid Bi-CG

Choose $\mathbf{x}_0$, ...
Compute $\mathbf{r}_0 = b - A\,\mathbf{x}_0$.
$\mathbf{x} = \mathbf{x}_0$, $\mathbf{r} = \mathbf{r}_0$.
$\mathbf{u} = 0$, ...
While $\|\mathbf{r}\| > tol$ do

Use $\cancel{\mathbf{x}}$, $\mathbf{r}$, $\mathbf{u}$, ...
to update $\mathbf{p}$

$\mathbf{q} \leftarrow A\mathbf{p}$
⋮
$\mathbf{p} \leftarrow \alpha\,\mathbf{p}$, $\mathbf{q} \leftarrow \alpha\,\mathbf{q}$
$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{p}$
$\mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}$

end while

Shifting the problem

$$A\,x = b$$
$$\mathbf{r} = b - A\mathbf{x}$$

$$A\,y = \mathbf{b}_0 := b - A\mathbf{x}_0$$
$$\mathbf{x} = \mathbf{x}_0 + \mathbf{y}$$
$$\mathbf{r} = \mathbf{b}_0 - A\mathbf{y}$$

Choose $\mathbf{x}$, ...
Compute $\mathbf{b} = b - A\,\mathbf{x}$.
$\mathbf{y} = 0$, $\mathbf{r} = \mathbf{b}$.
$\mathbf{u} = 0$, ...
While $\|\mathbf{r}\| > tol$ do

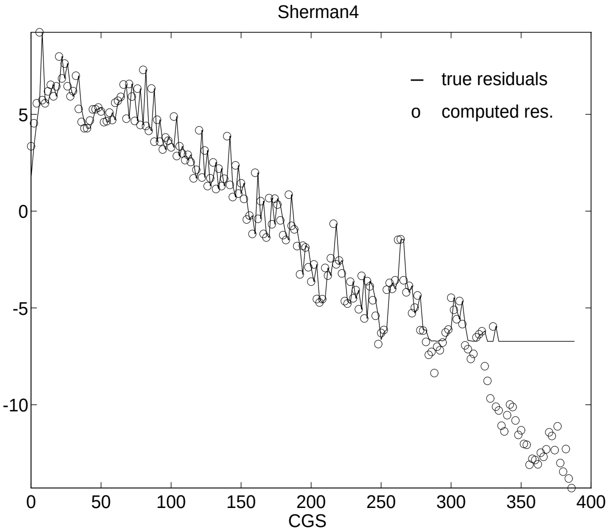
Use $\cancel{\mathbf{x}}$, $\mathbf{r}$, $\mathbf{u}$, ...
to update $\mathbf{p}$

$\mathbf{q} \leftarrow A\mathbf{p}$
⋮
$\mathbf{y} \leftarrow \mathbf{y} + \mathbf{p}$
$\mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}$

end while
$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}$

True and updated residuals in CGS

An example



Are the updated residuals $\mathbf{r}$ accurate:

is $\boxed{\dfrac{\|\,\mathbf{r}\,\|}{\|\,b - A\mathbf{x}\,\|} \approx 1}$?

Should we worry?

If $\mathbf{r}$ is used to monitor the progress (stop if $\|\,\mathbf{r}\,\| \leq tol$), then we may

- get a wrong answer,
- perform (many) useless iterations.

Cure: Use the true residual $b - A\mathbf{x}$.

If $b - A\mathbf{x}$ is used to monitor the progress, then we may

- not get an answer,
- need additional MVs.

Can inaccurate residuals be avoided?

∘ Restart if the updated residual is inaccurate:
  loss of acceleration &
  useless it. or ad. MVs have performed

∘ Replace the updated residual by $b - A\mathbf{x}$:
  loss of speed of convergence &
  add. MVs

- Use more accurate methods

• Use reliable update strategies for the computed approximations and residuals.

Update strategies
— should be (relatively) cheap
— should maintain the speed of convergence
— to be incorporated easily in existing code.

Restart

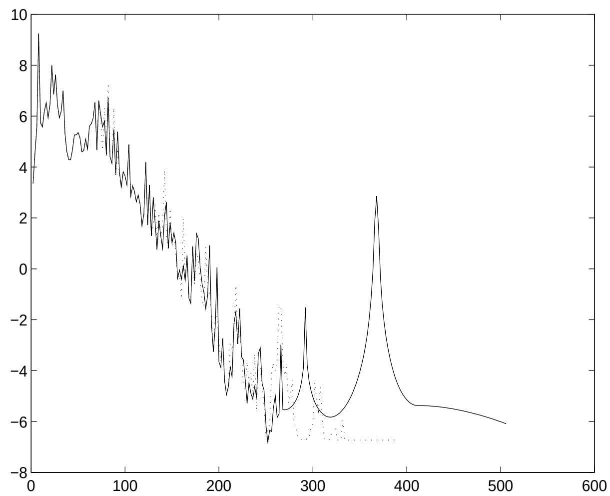
If $\mathbf{r}$ is not accurate, then restart with $\mathbf{x}_0 = \mathbf{x}$.

Example: Restarted CGS

Using true residuals in the recursions

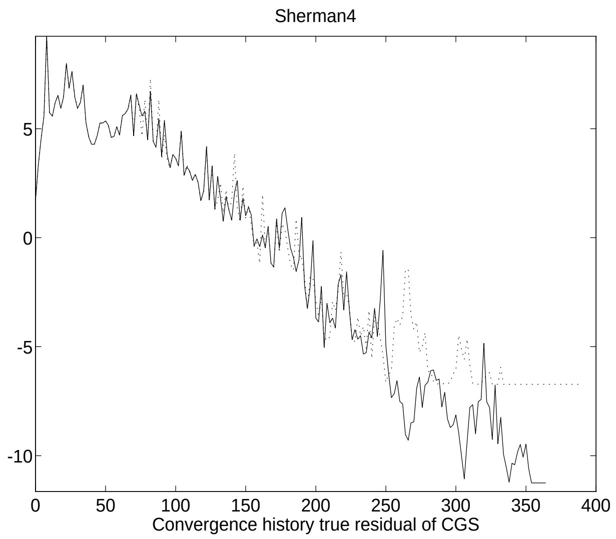
Replace $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}$ by $\mathbf{r} \leftarrow b - A\,\mathbf{x}$.

Example: Very cheap in, for instance, CGS.



Using true residuals in the recursions & strategically updated approximations

Example: Neumaier's approach for CGS



$$\mathbf{r} = \mathbf{b} - (A\mathbf{x})^\star = \mathbf{b} - A\mathbf{x} + \delta$$

δ such that

$$\begin{aligned} \|\,\delta\,\| &\leq \bar{\xi}\, n_A \,\big\|\,|A|\,\big\|\, \|\mathbf{x}\| \\ &\leq \bar{\xi}\, \underbrace{n_A \,\big\|\,|A|\,\big\|\, \big\|A^{-1}\big\|}\, \|A\mathbf{x}\| \\ &\leq \bar{\xi} \qquad \Gamma \quad (\|\mathbf{b}\| + \|\mathbf{r}\|) \end{aligned}$$

Accuracy: $\|\,\mathbf{r}\,\|/\|\,b - A\mathbf{x}\,\| \approx 1$?

Exact arithmetic:

$$\begin{aligned} \mathbf{x}_{j+1} &= \mathbf{x}_j + \mathbf{p}_j \\ \mathbf{r}_{j+1} &= \mathbf{r}_j - A\mathbf{p}_j \end{aligned}$$

Finite precision:

$$\begin{aligned} \mathbf{x}_{j+1} &= \mathbf{x}_j + \mathbf{p}_j^\star \\ \mathbf{r}_{j+1} &= \mathbf{r}_j - (A\mathbf{p}_j^\star)^\star = \mathbf{r}_j - A\mathbf{p}_j^\star - \delta_j \\ & \qquad \delta_j = \delta_j(\mathbf{x}_j) \end{aligned}$$

$$\begin{aligned} b - A\mathbf{x}_{j+1} &= b - A\mathbf{x}_j - A\mathbf{p}_j^\star \\ \mathbf{r}_{j+1} &= \mathbf{r}_j \quad - A\mathbf{p}_j^\star - \delta_j \end{aligned}$$

$$\|\,\mathbf{r}_k - (b - A\mathbf{x}_k)\,\| \leq \sum_{j<k} \|\,\delta_j\,\|$$

Theorem

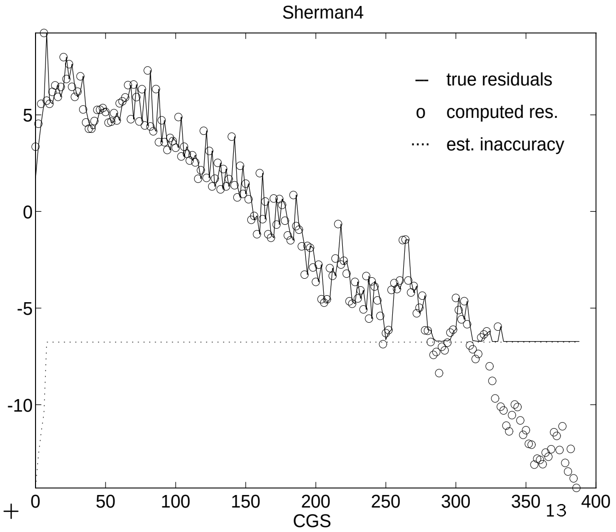
$$\boxed{\Big|\,\|\,\mathbf{r}_k\| - \|\,b - A\mathbf{x}_k\|\,\Big| \leq 2\,\bar{\xi}\,\Gamma \sum_{j\leq k} \|\,\mathbf{r}_j\| \qquad \text{where} \quad \Gamma := n_A \,\big\|\,|A|\,\big\|\, \big\|\,A^{-1}\big\|}$$

Accuracy is determined by
the size of maximal <u>intermediate</u> residuals

(also by the residuals that are not shown in convergence plots)

Stopping criterion

$\vdots$
$\tau = \kappa\,\overline{\xi}\,\|\,\mathbf{r}\,\|$
while $\|\,\mathbf{r}\,\| > tol$ and $\|\,\mathbf{r}\,\| \geq \tau$ do
  $\vdots$
  $\tau \leftarrow \tau + \kappa\,\overline{\xi}\,\|\,\mathbf{r}\,\|$
end while



$\kappa\,\overline{\xi}\,\|\,b\,\|$ maximal accuracy we may hope for:
$b - (Ax)^\star = \delta$ and $\|\,b - (Ax)^\star\,\| \leq \overline{\xi}\,\Gamma\,\|\,b\,\|$

$$\Big|\,\|\,\mathbf{r}_k\| - \|\,b - A\,\mathbf{x}_k\|\,\Big| \leq \sum_{j<k} \|\,\delta_j(\mathbf{x}_j)\|$$

For accuracy,
we should have local errors $\delta(\mathbf{x})$ that are
globally absolutely small,
say $\|\,\delta(\mathbf{x})\,\| \leq \kappa\,\overline{\xi}\,\|\,b\,\|$.

Convergence

$\mathbf{p} = \mathbf{p}_{\text{exact}} + \delta(\mathbf{p})$ where $\mathbf{p}_{\text{exact}}$, $\mathbf{r}_{\text{exact}}$
$\mathbf{r} = \mathbf{r}_{\text{exact}} + \delta(\mathbf{r})$ are the search direction,
and the residual if (only)
last sweep would
have been exact.

$\delta(\mathbf{p})$, $\delta(\mathbf{r})$ **local** recursion errors.

If the algorithm is insensitive
then locally relatively small errors

i.e. $\|\,\delta(\mathbf{p})\,\| \leq \kappa\,\overline{\xi}\,\|\,\mathbf{p}\,\|$, $\|\,\delta(\mathbf{r})\,\| \leq \kappa\,\overline{\xi}\,\|\,\mathbf{r}\,\|$,

do not slow down speed of convergence

Bi-CG (and related methods) is based on
global bi-orthogonality but
local bi-orth. seems to work equally well.

For maintaining convergence,
we should have local errors $\delta(\mathbf{r})$ that are
locally relatively small,
say $\|\,\delta(\mathbf{r})\,\| \leq \kappa\,\overline{\xi}\,\|\,\mathbf{r}\,\|$.

Are local rec. errors δ_j locally rel. small?

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{p}_j^\star$$
$$\mathbf{r}_{j+1} = \mathbf{r}_j - (A\mathbf{p}_j^\star)^\star = \mathbf{r}_j - A\mathbf{p}_j^\star - \delta_j$$

$$\|\,\delta_j\,\| \leq \overline{\xi}\,\Gamma\,\Big(\|\mathbf{r}_j\| + \|\mathbf{r}_{j+1}\|\Big)$$

Also if $\|\,\mathbf{r}_{j+1}\| \ll \|\,\mathbf{r}_j\|$ *?*

Using $\mathbf{r} \leftarrow b - A\,\mathbf{x}$ in each step

Updated residual:

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \mathbf{p}_j$$
$$\mathbf{r}_{j+1} = \mathbf{r}_j - (A\mathbf{p}_j)^\star$$

True residual:

$$\begin{aligned}\mathbf{r}_{j+1} &= b - (A\mathbf{x}_{j+1})^\star = b - A\mathbf{x}_{j+1} + \delta \\ &= b - A\mathbf{x}_j - A\mathbf{p}_j + \delta \\ &= b - (A\mathbf{x}_j)^\star - (A\mathbf{p}_j)^\star + \delta + \delta' + \delta''\end{aligned}$$

$$\|\delta^+\| \leq 2\,\overline{\xi}\,\Gamma\,\left(\|\,b\,\| + \|\,\mathbf{r}_j\,\| + \|\,\mathbf{r}_{j+1}\,\|\right)$$

Disadvantages:

- disturbs convergence if $\|\,\mathbf{r}_{j+1}\,\| \ll \|\,b\,\|$,
- (may) cost (many) additional MVs.

Using $\mathbf{r} \leftarrow b - A\,\mathbf{x}$ occasionally

$$\mathbf{r}_{j_1} \leftarrow b - (A\mathbf{x}_{j_1})^\star \qquad \& \qquad \mathbf{r}_{j_2} \leftarrow b - (A\mathbf{x}_{j_2})^\star$$

$$\|\delta^+\| \leq 2\,\overline{\xi}\,\Gamma\,\left(\|\,b\,\| + \sum_{j_1 \leq j \leq j_2} \|\,\mathbf{r}_j\,\|\right)$$

Disadvantages:

- disturbs convergence if $\|\,\mathbf{r}_{j_2}\,\| \ll \|\,b\,\|$,
- disturbs convergence if $\|\,\mathbf{r}_{j_2}\,\| \ll \|\,\mathbf{r}_j\,\|$, for some j, $j_1 \leq j \leq j_2$
- (may) cost (a few) additional MVs.

Advantage: More accurate.

For $k > j_*$

$$\begin{aligned}\mathbf{r}_k &= b - (A\mathbf{x}_{j_*})^\star - (A\mathbf{p}_{j_*})^\star - \ldots - (A\mathbf{p}_{k-1})^\star \\ &= b - A\mathbf{x}_{j_*} - A\mathbf{p}_{j_*} - \ldots - A\mathbf{p}_{k-1} + \textstyle\sum \delta\end{aligned}$$

$$\Big|\,\|\,\mathbf{r}_k\,\| - \|\,b - A\mathbf{x}_k\,\|\,\Big| \leq \overline{\xi}\,\Gamma\,\left(\|\,b\,\| + 2\sum_{j_* \leq j \leq k} \|\,\mathbf{r}_j\,\|\right)$$

of interest if $\|\,\mathbf{r}_j\,\| < \|b\|$ $\quad (j_* \leq j)$.

Cure: $\mathbf{r}_{j_2} \leftarrow b - (A\mathbf{x}_{j_2})^\star$ as soon as

$\|\,\mathbf{r}_{j_2}\,\| < 10^{-2}\|\,\mathbf{r}_j\,\|$ for some j, $j_1 \leq j \leq j_2$
and
$\|\,\mathbf{r}_{j_2}\,\| > 10^{-2}\|\,b\,\|$

Using $\mathbf{r} \leftarrow b - A\mathbf{x}$ without disturbing convergence

- only if $\|\mathbf{r}\| > 10^{-2}\|b\|$

or

- if you like but restart if $\|\mathbf{r}\| < 10^{-2}\|\mathbf{b}\|$

(• [Neumaier] $\mathbf{r} \leftarrow b - A\mathbf{x}$ in each step & restart if $\|\mathbf{r}\| < \|\mathbf{b}\|$.)

Flying restart

Choose $\mathbf{x}$, ...
Compute $\mathbf{b} = b - A\,\mathbf{x}$.
$\mathbf{y} = 0$, $\mathbf{r} = \mathbf{b}$.
$\mathbf{u} = 0$, ...
while $\|\mathbf{r}\| > tol$ do

Use $\cancel{\mathbf{x}}$, $\mathbf{r}$, $\mathbf{u}$, ...
to update $\mathbf{p}$

$\mathbf{q} \leftarrow A\,\mathbf{p}$
⋮
$\mathbf{y} \leftarrow \mathbf{y} + \mathbf{p}$
$\mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}$

if *flying_restart* is *true*
$\mathbf{r} \leftarrow \mathbf{b} - A\,\mathbf{y}$
$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}$, $\mathbf{y} \leftarrow 0$, $\mathbf{b} \leftarrow \mathbf{r}$
end if

end while
$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}$

Using $\mathbf{r} \leftarrow b - A\mathbf{x}$ without disturbing convergence

- only if $\|\mathbf{r}\| > 10^{-2}\|b\|$

or

- if you like but flying restart if $\|\mathbf{r}\| < 10^{-2}\|\mathbf{b}\|$

(• [Neumaier] $\mathbf{r} \leftarrow b - A\mathbf{x}$ in each step & flying restart if $\|\mathbf{r}\| < \|\mathbf{b}\|$.)

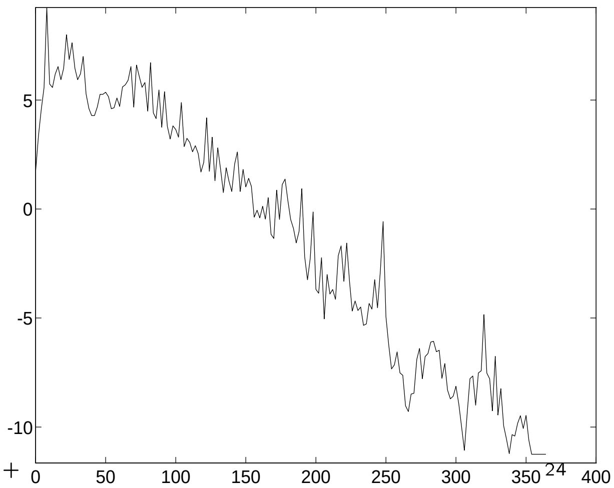
$\mathbf{x}_{\pi(j)} = \mathbf{x}_{\pi(j-1)} + \mathbf{y}_j$
$\mathbf{b}_j := \mathbf{r}_{\pi(j)} = \mathbf{r}_{\pi(j-1)} - A\mathbf{y}_j$

$$\Big|\|\mathbf{r}_{\pi(j)}\| - \|b - A\mathbf{x}_{\pi(j)}\|\Big| \leq 2\,\overline{\xi}\,\Gamma \sum_{i \leq j} \|\mathbf{r}_{\pi(i)}\|$$

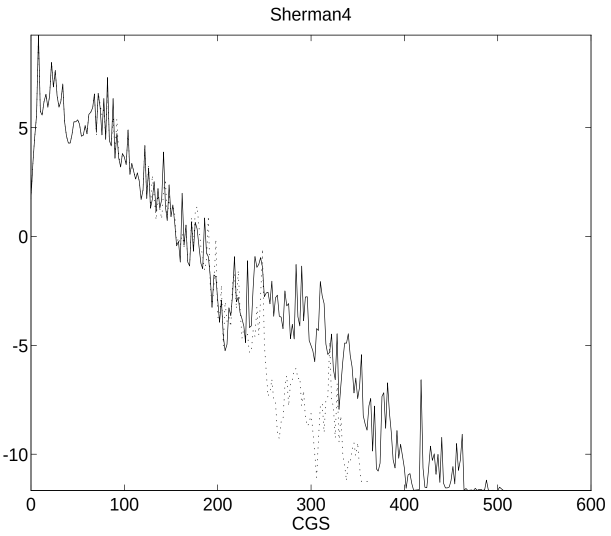
For accuracy: $\|\mathbf{b}_j\| = \|\mathbf{r}_{\pi(j)}\| < \|b\|$.

When to perform flying restarts?

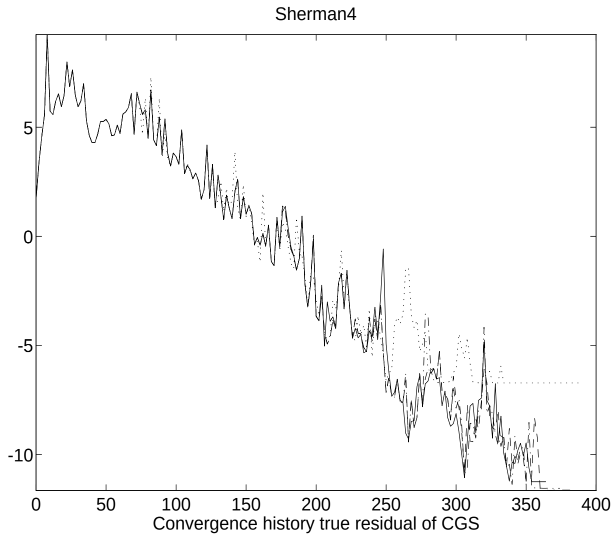
Sherman4



Convergence history true residual of CGS [Neumaier]



one flying restart, no other true residuals



compute_res is *true*		in each step	
flying_restart is *true*		three strategies	
restart if			
$\|\mathbf{r}\| \leq \|\mathbf{b}\|$:	*true*	$29\times$	—
$\|\mathbf{r}\| \leq 10^{-2}\|\mathbf{b}\|$:	*true*	$5\times$	-.-.
$\|\mathbf{r}\| \leq 10^{-2}\|b\|$ & ...:	*true*	$1\times$	- -

Choose $\mathbf{x}$, ...

Compute $\mathbf{b} = b - A\mathbf{x}$.

$\mathbf{y} = 0$, $\mathbf{r} = \mathbf{b}$.

$\mathbf{u} = 0$, ...

while $\|\mathbf{r}\| > tol$ do

> *Use* $\cancel{\mathbf{x}}$, $\mathbf{r}$, $\mathbf{u}$, ...
> *to update* $\mathbf{p}$

$\mathbf{q} \leftarrow A\,\mathbf{p}$

$\vdots$

$\mathbf{y} \leftarrow \mathbf{y} + \mathbf{p}$

$\mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}$

> if *compute_res* is *true*
> $\quad\mathbf{r} \leftarrow \mathbf{b} - A\mathbf{y}$
> end if
> if *flying_restart* is *true*
> $\quad\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}, \quad \mathbf{y} \leftarrow 0, \quad \mathbf{b} \leftarrow \mathbf{r}$
> end if

end while

$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}$

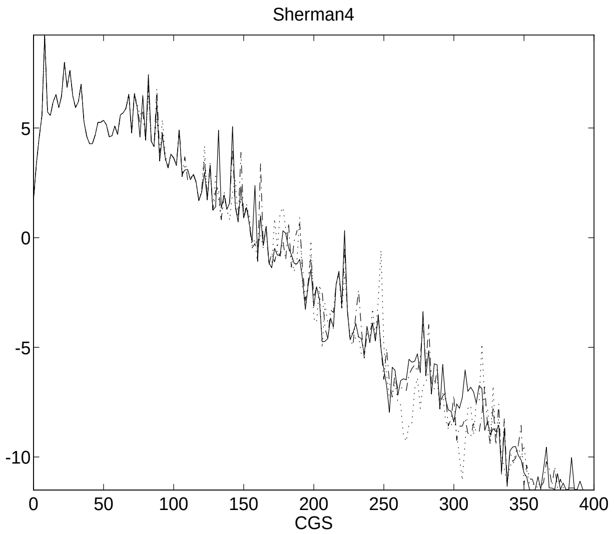
Saving MVs

If

$$\boxed{\begin{array}{l}\vdots\\ \mathbf{q} \leftarrow A\,\mathbf{p}\\ \vdots\\ \mathbf{p} \leftarrow \alpha\,\mathbf{p}, \quad \mathbf{q} \leftarrow \alpha\,\mathbf{q}\\ \mathbf{x} \leftarrow \mathbf{x} + \mathbf{p}\\ \mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}\end{array}} = \boxed{\begin{array}{l}\vdots\\ \mathbf{p} \leftarrow \alpha\,\mathbf{p}\\ \mathbf{x} \leftarrow \mathbf{x} + \mathbf{p}\\ \mathbf{q} \leftarrow A\,\mathbf{p}\\ \mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}\end{array}}$$

then we can replace this by

> $\vdots$
> $\mathbf{p} \leftarrow \alpha\,\mathbf{p}$
> $\mathbf{y} \leftarrow \mathbf{y} + \mathbf{p}$
> if *compute_res* is *true*
> $\quad\hat{\mathbf{r}} = \mathbf{b} - A\mathbf{y}, \quad \mathbf{q} = \mathbf{r} - \hat{\mathbf{r}}, \quad \mathbf{r} \leftarrow \hat{\mathbf{r}}$
> else
> $\quad\mathbf{q} \leftarrow A\,\mathbf{p}, \quad \mathbf{r} \leftarrow \mathbf{r} - \mathbf{q}$
> end if
> if *flying_restart* is *true*
> $\quad\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}, \quad \mathbf{y} \leftarrow 0, \quad \mathbf{b} \leftarrow \mathbf{r}$
> end if



special updates as few as possible

MV saving st. — & simple st. -.-.-.

Resulting in
7 true residual computations and
1 flying restart

...Neumaier

CGS

Choose $\mathbf{x}_0$ and $\tilde{r}_0$.
Compute $\mathbf{r}_0 = b - A\,\mathbf{x}_0$.
$\mathbf{x} = \mathbf{x}_0, \quad \mathbf{r} = \mathbf{r}_0$.
$\mathbf{u} = \mathbf{w} = 0, \quad \rho = 1$.
while $\|\mathbf{r}\| > \textit{tol}$ do

$$\sigma = -\rho, \quad \rho = (\mathbf{r}, \tilde{r}_0), \quad \beta = \frac{\rho}{\sigma}$$
$$\mathbf{v} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$$
$$\mathbf{w} \leftarrow \mathbf{v} - \beta\,(\mathbf{u} - \beta\,\mathbf{w}),$$
$$\mathbf{t} \leftarrow A\,\mathbf{w}$$
$$\sigma = (\mathbf{t}, \tilde{r}_0), \quad \alpha = \frac{\rho}{\sigma}$$
$$\mathbf{u} \leftarrow \mathbf{v} - \alpha\,\mathbf{t}$$
$$\mathbf{x} \leftarrow \mathbf{x} + \alpha\,(\mathbf{v} + \mathbf{u})$$
$$\mathbf{r} \leftarrow \mathbf{r} - \alpha\,A\,(\mathbf{v} + \mathbf{u})$$

end while

Reliably and cheaply updated CGS

Choose $\mathbf{x}$ and $\tilde{r}_0$.
Compute $\mathbf{b} = b - A\,\mathbf{x}$.
$\mathbf{y} = 0, \quad \mathbf{r} = \mathbf{b}$.
$\mathbf{u} = \mathbf{w} = 0, \quad \rho = 1$.
while $\|\mathbf{r}\| > \textit{tol}$ do

⋮

$$\mathbf{y} \leftarrow \mathbf{y} + \alpha\,(\mathbf{v} + \mathbf{u})$$

if *compute_res* is *true*
$\quad \mathbf{r} \leftarrow \mathbf{b} - A\,\mathbf{y}$
else
$\quad \mathbf{r} \leftarrow \mathbf{r} - \alpha\,A\,(\mathbf{v} + \mathbf{u})$
end if
if *flying_restart* is *true*
$\quad \mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}, \quad \mathbf{y} \leftarrow 0, \quad \mathbf{b} \leftarrow \mathbf{r}$
end if

end while
$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}$

Neumaier's CGS

Choose $\mathbf{x}$ and $\tilde{r}_0$.
Compute $\mathbf{b} = b - A\,\mathbf{x}$.
$\mathbf{y} = 0, \quad \mathbf{r} = \mathbf{b}$.
$\mathbf{u} = \mathbf{w} = 0, \quad \rho = 1$.
while $\|\mathbf{r}\| > \textit{tol}$ do

$$\sigma = -\rho, \quad \rho = (\mathbf{r}, \tilde{r}_0), \quad \beta = \frac{\rho}{\sigma}$$
$$\mathbf{v} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$$
$$\mathbf{w} \leftarrow \mathbf{v} - \beta\,(\mathbf{u} - \beta\,\mathbf{w})$$
$$\mathbf{t} \leftarrow A\mathbf{w}$$
$$\sigma = (\mathbf{t}, \tilde{r}_0), \quad \alpha = \frac{\rho}{\sigma}$$
$$\mathbf{u} \leftarrow \mathbf{v} - \alpha\,\mathbf{t}$$
$$\mathbf{y} \leftarrow \mathbf{y} + \alpha\,(\mathbf{v} + \mathbf{u})$$

$\mathbf{r} \leftarrow \mathbf{b} - A\,\mathbf{y}$
if *flying_restart* is *true*
$\quad \mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}, \quad \mathbf{y} \leftarrow 0, \quad \mathbf{b} \leftarrow \mathbf{r}$
end if

true if $\|\mathbf{r}\| \le \|\mathbf{b}\|$

end while
$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{y}$

BiCGSTAB

Choose $\mathbf{x}_0$ and $\tilde{r}_0$.
Compute $\mathbf{r}_0 = b - A\mathbf{x}_0$.
$\mathbf{x} = \mathbf{x}_0$, $\mathbf{r} = \mathbf{r}_0$.
$\mathbf{u} = 0$, $\omega = \sigma = 1$.
While $\|\mathbf{r}\| > tol$ do
    $\sigma = -\omega\sigma$, $\rho = (\mathbf{r}, \tilde{r}_0)$, $\beta = \frac{\rho}{\sigma}$
    $\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$
    $\mathbf{c} = A\mathbf{u}$
    $\sigma = (\mathbf{c}, \tilde{r}_0)$, $\alpha = \frac{\rho}{\sigma}$
    $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$
    $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,\mathbf{c}$
    $\mathbf{t} = A\mathbf{r}$
    $\omega = (\mathbf{r}, \mathbf{t})/(\mathbf{t}, \mathbf{t})$
    $\mathbf{u} \leftarrow \mathbf{u} - \omega\,\mathbf{c}$
    $\mathbf{x} \leftarrow \mathbf{x} + \omega\,\mathbf{r}$
    $\mathbf{r} \leftarrow \mathbf{r} - \omega\,\mathbf{t}$
end while

BiCGSTAB

Choose $\mathbf{x}_0$ and $\tilde{r}_0$.
Compute $\mathbf{r}_0 = b - A\mathbf{x}_0$.
$\mathbf{x} = \mathbf{x}_0$, $\mathbf{r} = \mathbf{r}_0$.
$\mathbf{u} = 0$, $\omega = \sigma = 1$.
While $\|\mathbf{r}\| > tol$ do
    $\sigma = -\omega\sigma$, $\rho = (\mathbf{r}, \tilde{r}_0)$, $\beta = \frac{\rho}{\sigma}$
    $\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$, $\mathbf{c} = A\mathbf{u}$
    $\sigma = (\mathbf{c}, \tilde{r}_0)$, $\alpha = \frac{\rho}{\sigma}$
    $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$
    $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,\mathbf{c}$, $\mathbf{t} = A\mathbf{r}$
    $\omega = (\mathbf{r}, \mathbf{t})/(\mathbf{t}, \mathbf{t})$
    $\mathbf{u} \leftarrow \mathbf{u} - \omega\,\mathbf{c}$
    $\mathbf{x} \leftarrow \mathbf{x} + \omega\,\mathbf{r}$
    $\mathbf{r} \leftarrow \mathbf{r} - \omega\,\mathbf{t}$
end while

Computing α (that is σ) without $A\mathbf{u}$:

$$\sigma = (A\mathbf{u}, \tilde{r}\ ) = (\mathbf{u}, A^T\tilde{r}\ ) = -\frac{1}{\alpha}(\mathbf{u}, \tilde{r}\ )$$

BiCGstab(ℓ) Orthogonal basis variant

Choose $\mathbf{x}$ and $\tilde{r}$.
Compute $\mathbf{r}\ = b - A\mathbf{x}$.
$\mathbf{x} = \mathbf{x}$, $\mathbf{r} = \mathbf{r}$.
$\mathbf{u} = 0$, $\gamma_\ell = \sigma = 1$.
While $\|\mathbf{r}\| > tol$ do
  $\sigma \leftarrow -\gamma_\ell\sigma$, $\mathbf{C} = [\ ]$
  for $j = 1$ to ℓ do
    $\rho = (\mathbf{r}, \tilde{r}_{j-}\ )$, $\beta = -\frac{\rho}{\sigma}$
    $\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$, $\mathbf{C} \leftarrow [\,\mathbf{C}, A\mathbf{u}\,]$
    $\sigma = (\mathbf{c}_j, \tilde{r}_{j-}\ )$, $\alpha = \frac{\rho}{\sigma}$
    $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$
    $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,\mathbf{c}_j$, $\mathbf{b} = A\mathbf{r}$
    if $\tilde{r}_j$ undefined: compute $\tilde{r}_j$ as in Bi-CG
  end do
  Apply ℓ steps of ORTHODIR to solve $A\mathbf{y} = \mathbf{r}$
  finding the approximate solution $\hat{\mathbf{y}}$, residual $\hat{\mathbf{r}}$
  and orthonormal matrix $\mathbf{R}$ spanning $\mathcal{K}_\ell(A; \mathbf{b})$.
  $\mathbf{x} \leftarrow \mathbf{x} + \hat{\mathbf{y}}$, $\mathbf{r} \leftarrow \hat{\mathbf{r}}$
  Update $\mathbf{u}$ from $\mathbf{u}$, $\mathbf{R}$ and $\mathbf{C}$. Compute γ_ℓ
end while

$\mathbf{C} := [\mathbf{c}\ , \ldots, \mathbf{c}_j]$ and $\mathbf{R}$ are matrices with n rows

BiCGstab(ℓ) Power basis variant

Choose $\mathbf{x}$ and $\tilde{r}$.
Compute $\mathbf{r}\ = b - A\mathbf{x}$.
$\mathbf{x} = \mathbf{x}$, $\mathbf{r} = [\mathbf{r}\ ]$.
$\mathbf{u} = [0]$, $\gamma_\ell = \sigma = 1$.
While $\|\mathbf{r}\| > tol$ do
  $\sigma \leftarrow -\gamma_\ell\sigma$
  for $j = 1$ to ℓ do
    $\rho = (\mathbf{r}_j, \tilde{r}\ )$, $\beta = -\frac{\rho}{\sigma}$
    $\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$, $\mathbf{u} \leftarrow [\,\mathbf{u}, A\mathbf{u}_j\,]$
    $\sigma = (\mathbf{u}_j\ \ , \tilde{r}\ )$, $\alpha = \frac{\rho}{\sigma}$
    $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$
    $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,[\mathbf{u}\ , \ldots, \mathbf{u}_j\ \ ]$, $\mathbf{r} \leftarrow [\,\mathbf{r}, A\mathbf{r}_j\,]$
  end do
  $\mathbf{R} := [\mathbf{r}\ , \ldots, \mathbf{r}_\ell\ \ ]$
  solve $\vec{\gamma}$ from $(\mathbf{R}^T\mathbf{R})\vec{\gamma} = \mathbf{R}^T\mathbf{r}$
  $\mathbf{u} \leftarrow [\,\mathbf{u}\ \ - (\gamma\ \mathbf{u}\ \ + \ldots + \gamma_\ell\mathbf{u}_\ell\ \ )\,]$
  $\mathbf{x} \leftarrow \mathbf{x} + (\gamma\ \mathbf{r}\ \ + \ldots + \gamma_\ell\mathbf{r}_\ell)$
  $\mathbf{r} \leftarrow [\,\mathbf{r}\ \ - (\gamma\ \mathbf{r}\ \ + \ldots + \gamma_\ell\mathbf{r}_\ell\ \ )\,]$
end while

$\mathbf{r}$ and $\mathbf{u}$ are matrices with n rows, # columns varies:
$\mathbf{r} := [\mathbf{r}\ , \ldots, \mathbf{r}_j]$, $\mathbf{u} := [\mathbf{u}\ , \ldots, \mathbf{u}_j]$. $\vec{\gamma} := (\gamma\ , \ldots, \gamma_\ell)^T$

Power basis variant

$\rho = (\mathbf{r}_1, \tilde{r}_0), \quad \beta = \dfrac{\rho}{\sigma}$

$\mathbf{u}_1 \leftarrow \mathbf{r}_1 - \beta\,\mathbf{u}_1, \qquad \mathbf{u}_2 \leftarrow A\mathbf{u}_1$

$\sigma = (\mathbf{u}_2, \tilde{r}_0), \quad \alpha = \dfrac{\rho}{\sigma}$

$\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}_1$

(1) $\mathbf{r}_1 \leftarrow \mathbf{r}_1 - \alpha\,\mathbf{u}_2, \qquad \mathbf{r}_2 \leftarrow A\mathbf{r}_1$

$\rho = (\mathbf{r}_2, \tilde{r}_0), \quad \beta = \dfrac{\rho}{\sigma}$

(2) $\mathbf{u}_1 \leftarrow \mathbf{r}_1 - \beta\,\mathbf{u}_1, \quad \mathbf{u}_2 \leftarrow \mathbf{r}_2 - \beta\,\mathbf{u}_2, \qquad \mathbf{u}_3 \leftarrow A\mathbf{u}_2$

$\sigma = (\mathbf{u}_3, \tilde{r}_0), \quad \alpha = \dfrac{\rho}{\sigma}$

$\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}_1$

$\mathbf{r}_1 \leftarrow \mathbf{r}_1 - \alpha\,\mathbf{u}_2, \quad \mathbf{r}_2 \leftarrow \mathbf{r}_2 - \alpha\,\mathbf{u}_3, \qquad \mathbf{r}_3 \leftarrow A\mathbf{r}_2$

$\vdots$

Computing $\mathbf{u}_2$ in (1) from

$$\hat{\mathbf{r}} = \mathbf{b} - A\mathbf{x}, \quad \mathbf{u}_2 = (\mathbf{r}_1 - \hat{\mathbf{r}})/\alpha, \quad \mathbf{r}_1 \leftarrow \hat{\mathbf{r}}$$

may decouple $\mathbf{u}_1$ and $\mathbf{u}_2$ in (2):

$$\mathbf{u}_2 = A\mathbf{u}_1??$$

Orthogonal basis variant

$\rho = (\mathbf{r}_1, \tilde{r}_0), \quad \beta = \dfrac{\rho}{\sigma}$

$\mathbf{u}_1 \leftarrow \mathbf{r}_1 - \beta\,\mathbf{u}_1, \qquad \mathbf{u}_2 \leftarrow A\mathbf{u}_1$

$\sigma = (\mathbf{u}_2, \tilde{r}_0), \quad \alpha = \dfrac{\rho}{\sigma}$

$\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}_1$

(1) $\mathbf{r}_1 \leftarrow \mathbf{r}_1 - \alpha\,\mathbf{u}_2, \qquad \mathbf{r}_2 \leftarrow A\mathbf{r}_1$

$\rho = (\mathbf{r}_1, \tilde{r}_1), \quad \beta = \dfrac{\rho}{\sigma}$

(2) $\mathbf{u}_1 \leftarrow \mathbf{r}_1 - \beta\,\mathbf{u}_1, \qquad \mathbf{u}_2 \leftarrow A\mathbf{u}_1$

$\sigma = (\mathbf{u}_3, \tilde{r}_1), \quad \alpha = \dfrac{\rho}{\sigma}$

$\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}_1$

$\mathbf{r}_1 \leftarrow \mathbf{r}_1 - \alpha\,\mathbf{u}_2, \qquad \mathbf{r}_2 \leftarrow A\mathbf{r}_1$

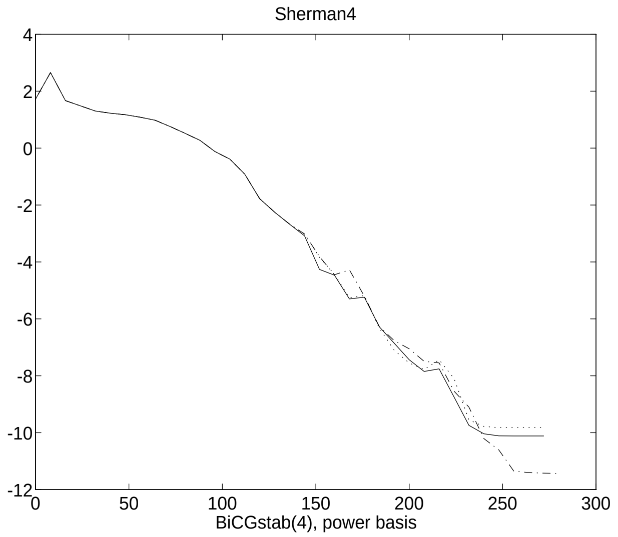
$\vdots$

Computing $\mathbf{u}_2$ in (1) from

$$\hat{\mathbf{r}} = \mathbf{b} - A\mathbf{x}, \quad \mathbf{u}_2 = (\mathbf{r}_1 - \hat{\mathbf{r}})/\alpha, \quad \mathbf{r}_1 \leftarrow \hat{\mathbf{r}}$$

does not decouple $\mathbf{u}_1$ and $\mathbf{u}_2$ in (2):

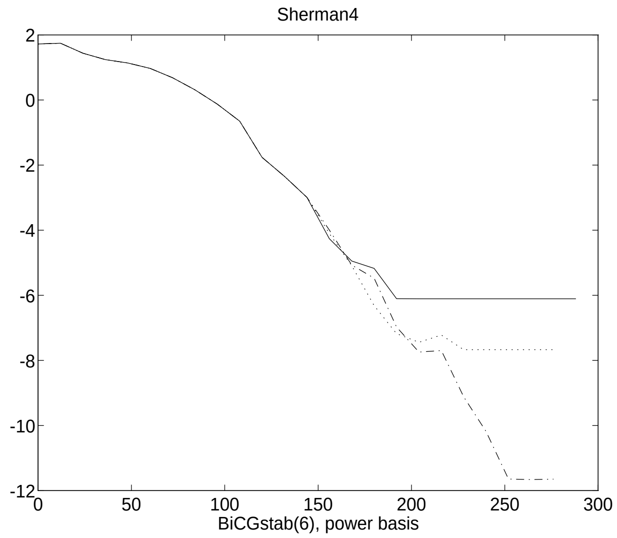
$$\mathbf{u}_2 = A\mathbf{u}_1!!$$



special updates as few as possible
MV saving strategy & simple strategy

In both cases, resulting in only
1 residual update, 1 flying restart

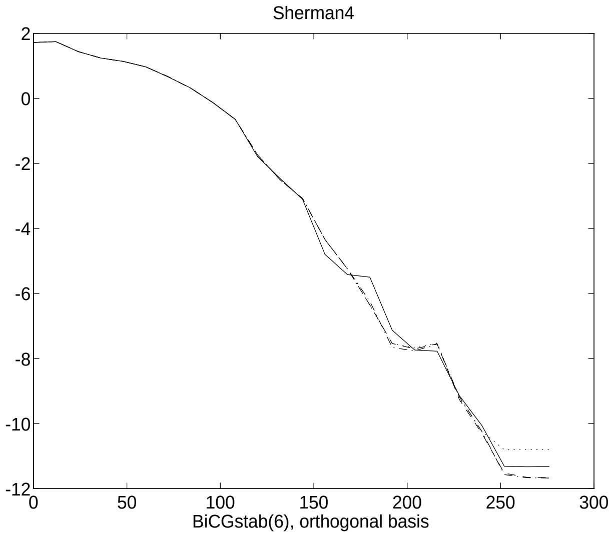
... standard; — MV saving; -.-.-. simple



special updates as few as possible
MV saving strategy & simple strategy

Resulting in only one special update!

... standard; — MV saving; -.-.-. simple



standard: . . .

Neumaier: —

as few as possible & MV saving: -.-.-.

as few as possible & simple strategy- - - -

Conclusions

To obtain accurate approximations
without loss of speed of convergence:

A. use true residuals whenever
the residual decreases significantly
and is not much smaller then $\| b \|$.

or

B. use true residuals
at least whenever the residual decreases
significantly and is not much smaller then $\| b \|$
and perform a flying restart if the residual
is (significantly) smaller than $\| \mathbf{b} \|$.