## Efficient methods

CG — Conjugate Gradients
if $\mathbb{A}^* = \mathbb{A} > 0 \;\Rightarrow$ MINRES if $\mathbb{A}^* = \mathbb{A}$, SYMMLQ

$\Downarrow$ *non-symmetric problems*

Bi-CG — Bi-Conjugate Gradients $\Rightarrow$ Quasi Minimal Res. simple QMR
*smoothing convergence*
*avoiding breakdown*

$\Downarrow$ *transpose free*
*faster convergence*

CGS — Conjugate Gradients Squared

$\Downarrow$ *more accuracy*
*faster convergence*

Bi-CGSTAB — Bi-CG STABilized

$\Downarrow$ *avoiding breakdown*
*faster convergence*

BiCGstab($\ell$)



## "New" aspects

- Breakdown
- Effects of finite precision arithmetic

— Loss of speed of convergence

due to ◦ **near breakdown**
◦ **loss of orthogonality**

— Loss of accuracy

i.e., for computed $\mathbf{r}_k$ and computed $\mathbf{x}_k$

$$\frac{\|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|}{\|\mathbf{r}_k\|} \not\approx 1$$

due to ◦ **non-monotone convergence**



## Orthogonal residual methods Bi-CG and related methods

**Program**

- Bi-CG
- Miscellaneous

  Smoothing the convergence (fake)
  irregular convergence
  maintaining accuracy

- Hybrid Bi-CG

  CGS, GCGS, Bi-CGSTAB, BiCGstab($\ell$)



## Conjugate Gradients $\quad \mathbf{K} = \mathbf{I}$

$$\boxed{\begin{aligned}\mathbf{u}_k &= \mathbf{r}_k - \beta_k\,\mathbf{u}_{k-1}\\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k\end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{u}_k \in \mathcal{K}_{k+1}(\mathbf{A};\mathbf{r}_0)$$

**Theorem**. If $\mathbf{A}^* = \mathbf{A}$ and

$\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \mathbf{r}_{k-1}$ then $\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \mathcal{K}_k(\mathbf{A};\mathbf{r}_0)$

*Proof.*

$\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_{k-1}\,\mathbf{A}\,\mathbf{u}_{k-1} \perp \mathbf{r}_{k-1}$ by construction

$\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_{k-1}\,\mathbf{A}\,\mathbf{u}_{k-1} \perp \mathcal{K}_{k-1}(\mathbf{A};\mathbf{r}_0)$ by induction

$\mathbf{A}\,\mathbf{u}_k = \mathbf{A}\,\mathbf{r}_k - \beta_k\,\mathbf{A}\,\mathbf{u}_{k-1} \perp \mathbf{r}_k$ by construction

$\mathbf{A}\,\mathbf{u}_k = \mathbf{A}\,\mathbf{r}_k - \beta_k\,\mathbf{A}\,\mathbf{u}_{k-1} \perp \mathcal{K}_{k-1}(\mathbf{A};\mathbf{r}_0)$ by induction:

$\mathbf{A}\mathbf{r}_k \perp \mathcal{K}_{k-1}(\mathbf{A};\mathbf{r}_0) \quad\Leftrightarrow\quad \mathbf{r}_k \perp \mathbf{A}\mathcal{K}_{k-1}(\mathbf{A};\mathbf{r}_0) \subset \mathcal{K}_k(\mathbf{A};\mathbf{r}_0)$



## Conjugate Gradients $\quad \mathbf{K} = \mathbf{I}$

$$\boxed{\begin{aligned}\mathbf{u}_k &= \mathbf{r}_k - \beta_k\,\mathbf{u}_{k-1}\\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k\end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{u}_k \in \mathcal{K}_{k+1}(\mathbf{A};\mathbf{r}_0)$$
$$\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \mathbf{r}_{k-1}$$

---

With $\quad \rho_k := (\mathbf{r}_k, \mathbf{r}_k) \quad \& \quad \sigma_k := (\mathbf{A}\mathbf{u}_k, \mathbf{r}_k)$

and, since $\quad \mathbf{r}_k + \alpha_{k-1}\,\mathbf{A}\,\mathbf{r}_{k-1} \in \mathcal{K}_k(\mathbf{A};\mathbf{r}_0)$,

we have $\quad \alpha_k = \dfrac{(\mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{A}\mathbf{u}_k, \mathbf{r}_k)} = \dfrac{\rho_k}{\sigma_k}$ and

$$\beta_k = \frac{(\mathbf{A}\mathbf{r}_k, \mathbf{r}_{k-1})}{(\mathbf{A}\mathbf{u}_{k-1}, \mathbf{r}_{k-1})} = \frac{(\mathbf{r}_k, \mathbf{A}\mathbf{r}_{k-1})}{\sigma_{k-1}} = \frac{-\rho_k}{\alpha_{k-1}\sigma_{k-1}}$$



## Conjugate Gradients

$$\boxed{\begin{aligned}\mathbf{u}_k &= \mathbf{K}^{-1}\mathbf{r}_k - \beta_k\,\mathbf{u}_{k-1}\\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k\end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{K}\mathbf{u}_k \in \mathcal{K}_{k+1}(\mathbb{A};\mathbf{r}_0)$$

**Theorem**. If $\mathbf{A}^* = \mathbf{A}$ & $\mathbf{K}^* = \mathbf{K}$ and

$\mathbf{r}_k, \mathbf{A}\mathbf{u}_k \perp \mathbf{K}^{-1}\mathbf{r}_{k-1}$ then $\mathbf{r}_k, \mathbf{A}\mathbf{u}_k \perp \mathbf{K}^{-1}\mathcal{K}_k(\mathbb{A};\mathbf{r}_0)$



[Hestenes Stiefel '52]

## CG $\quad \mathbf{A} > 0,\ \mathbf{K} > 0$

Choose $\mathbf{x}_0$
$k = 0$
$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
$\mathbf{u}_{-1} = 0,\ \rho_{-1} = 1$
while $\|\mathbf{r}_k\| > tol$ do
  solve $\mathbf{w}$ from $\mathbf{K}\mathbf{w} = \mathbf{r}_k$
  $\rho_k = (\mathbf{r}_k, \mathbf{w}),\ \beta_k = -\rho_k/\rho_{k-1}$
  $\mathbf{u}_k = \mathbf{w} - \beta_k\,\mathbf{u}_{k-1}$
  $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$
  $\sigma_k = (\mathbf{c}_k, \mathbf{u}_k),\ \alpha_k = \rho_k/\sigma_k$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\,\mathbf{u}_k$
  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\,\mathbf{c}_k$
  $k \leftarrow k + 1$
end while



[Hestenes Stiefel '52]

## CG $\quad \mathbf{A} > 0,\ \mathbf{K} > 0$

Choose $\mathbf{x}_0$
$\mathbf{x} = \mathbf{x}_0,\ \mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$
$\mathbf{u} = 0,\ \rho = 1$
while $\|\mathbf{r}\| > tol$ do
  $\sigma \leftarrow -\rho$
  solve $\mathbf{w}$ from $\mathbf{K}\mathbf{w} = \mathbf{r}$
  $\rho \leftarrow (\mathbf{r}, \mathbf{w}),\ \beta \leftarrow \rho/\sigma$
  $\mathbf{u} \leftarrow \mathbf{w} - \beta\,\mathbf{u}$
  $\mathbf{c} \leftarrow \mathbf{A}\mathbf{u}$
  $\sigma \leftarrow (\mathbf{c}, \mathbf{u}),\ \alpha \leftarrow \rho/\sigma$
  $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}$
  $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,\mathbf{c}$
end while

## Properties CG

**Advantages**

- 1 MV, 2 DOT, 3 AXPY per step
to increase dimension Krylov subspace by one.
- Storage: 5 large vectors (incl. $\mathbf{b}$) (6 if $\mathbf{K} \neq \mathbf{I}$).
- Minimal residual method if $\mathbf{A}$, $\mathbf{K}$ pos.def.: $\|\mathbf{r}_k\|_{\mathbf{A}^{-1}}$ is minimal.
- Orth. residual method if $\mathbf{A}^* = \mathbf{A}$, $\mathbf{K}^* = \mathbf{K}$: $\mathbf{r}_k \perp \mathbf{K}^{-1}\mathcal{K}_k(\mathbb{A}; \mathbf{r}_0)$.
- No knowledge on properties of $\mathbb{A}$ is needed.
- ◦ Robust: CG (for pos.def.) always converges.

**Drawback**

- Does not work if $\mathbf{A} \neq \mathbf{A}^*$.
- ◦ CG is sensitive to evaluation errors (often loss of super-linear convergence).



## Bi-Conjugate Gradients $\quad \mathbf{K} = \mathbf{I}$

$$\boxed{\begin{aligned} \mathbf{u}_k &= \mathbf{r}_k - \beta_k\, \mathbf{u}_{k-1} \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\, \mathbf{A}\, \mathbf{u}_k \end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{u}_k \in \mathcal{K}_{k+1}(\mathbf{A}; \mathbf{r}_0)$$

**Theorem**. $\tilde{\mathbf{r}}_0, \dots, \tilde{\mathbf{r}}_{k-1}$ Krylov basis $\mathcal{K}_k(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$

$\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \tilde{\mathbf{r}}_{k-1}$ then $\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \mathcal{K}_k(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$

*Proof.*

$\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_{k-1}\, \mathbf{A}\, \mathbf{u}_{k-1} \perp \tilde{\mathbf{r}}_{k-1}$ by construction

$\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_{k-1}\, \mathbf{A}\, \mathbf{u}_{k-1} \perp \mathcal{K}_{k-1}(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$ by induction

$\mathbf{A}\, \mathbf{u}_k = \mathbf{A}\, \mathbf{r}_k - \beta_k\, \mathbf{A}\, \mathbf{u}_{k-1} \perp \tilde{\mathbf{r}}_{k-1}$ by construction

$\mathbf{A}\, \mathbf{u}_k = \mathbf{A}\, \mathbf{r}_k - \beta_k\, \mathbf{A}\, \mathbf{u}_{k-1} \perp \mathcal{K}_{k-1}(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$ by induction:

$$\mathbf{A}\, \mathbf{r}_k \perp \mathcal{K}_{k-1}(\mathbf{A}^*; \tilde{\mathbf{r}}_0) \ \Leftrightarrow\ \mathbf{r}_k \perp \mathbf{A}^* \mathcal{K}_{k-1}(\mathbf{A}^*; \tilde{\mathbf{r}}_0) \subset \mathcal{K}_k(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$$



## Bi-Conjugate Gradients $\quad \mathbf{K} = \mathbf{I}$

$$\boxed{\begin{aligned} \mathbf{u}_k &= \mathbf{r}_k - \beta_k\, \mathbf{u}_{k-1} \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\, \mathbf{A}\, \mathbf{u}_k \end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \tilde{\mathbf{r}}_{k-1}$$

with $\tilde{\mathbf{r}}_0, \dots, \tilde{\mathbf{r}}_{k-1}$ Krylov basis $\mathcal{K}_k(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$

---

With $\quad \rho_k := (\mathbf{r}_k, \tilde{\mathbf{r}}_k) \quad \& \quad \sigma_k := (\mathbf{A}\mathbf{u}_k, \tilde{\mathbf{r}}_k)$

and $\theta_k$ such that $\tilde{\mathbf{r}}_k - \bar{\theta}_k\, \mathbf{A}^*\, \tilde{\mathbf{r}}_{k-1} \in \mathcal{K}_k(\mathbf{A}^*; \tilde{\mathbf{r}}_0)$,

we have $\quad \alpha_k = \dfrac{(\mathbf{r}_k, \tilde{\mathbf{r}}_k)}{(\mathbf{A}\mathbf{u}_k, \tilde{\mathbf{r}}_k)} = \dfrac{\rho_k}{\sigma_k} \quad$ and

$$\beta_k = \frac{(\mathbf{A}\mathbf{r}_k, \tilde{\mathbf{r}}_{k-1})}{(\mathbf{A}\mathbf{u}_{k-1}, \tilde{\mathbf{r}}_{k-1})} = \frac{(\mathbf{r}_k, \mathbf{A}^*\, \tilde{\mathbf{r}}_{k-1})}{\sigma_{k-1}} = \frac{\rho_k}{\theta_k\, \sigma_{k-1}}$$

Fletcher '76: $\tilde{\mathbf{u}}_k = \tilde{\mathbf{r}}_k - \bar{\beta}_k\, \tilde{\mathbf{u}}_{k-1}$ & $\tilde{\mathbf{r}}_{k+1} = \tilde{\mathbf{r}}_k - \bar{\alpha}_k\, \mathbf{A}^*\, \tilde{\mathbf{u}}_k$



## Bi-Conjugate Gradients $\quad \mathbf{K} = \mathbf{I}$

$$\boxed{\begin{aligned} \mathbf{u}_k &= \mathbf{r}_k - \beta_k\, \mathbf{u}_{k-1} \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\, \mathbf{A}\, \mathbf{u}_k \end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \bar{q}_{k-1}(\mathbf{A}^*)\tilde{\mathbf{r}}_0$$

with $q_{k-1}$ a polynomial of exact degree $k-1$.

---

With

$\rho_k := (\mathbf{r}_k, \bar{q}_k(\mathbf{A}^*)\tilde{\mathbf{r}}_0) \quad \& \quad \sigma_k := (\mathbf{A}\mathbf{u}_k, \bar{q}_k(\mathbf{A}^*)\tilde{\mathbf{r}}_0)$

and $\theta_k$ such that

$$q_k(\zeta) - \theta_k\, \zeta\, q_{k-1}(\zeta) \in \mathcal{P}_{k-1}$$

we have

$$\alpha_k = \frac{\rho_k}{\sigma_k} \quad \& \quad \beta_k = \frac{\rho_k}{\theta_k\, \sigma_{k-1}}$$



[Sonneveld '89]

## Generalized Bi-CG

Choose $\mathbf{x}_0$ and $\tilde{\mathbf{r}}_0$
$k = 0$
$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
$\mathbf{u}_{-1} = 0,\ \theta_0 = 1,\ \sigma_{-1} = 1$
while $\|\mathbf{r}_k\| > tol$ do
  solve $\mathbf{w}$ from $\mathbf{K}\mathbf{w} = \mathbf{r}_k$
  $\rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_k),\ \beta_k = \rho_k/(\theta_k\, \sigma_{k-1})$
  $\mathbf{u}_k = \mathbf{w} - \beta_k\, \mathbf{u}_{k-1}$
  $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$
  $\sigma_k = (\mathbf{c}_k, \tilde{\mathbf{r}}_k),\ \alpha_k = \rho_k/\sigma_k$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\, \mathbf{u}_k$
  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\, \mathbf{c}_k$
  $k \leftarrow k+1$
    select a $\tilde{\mathbf{r}}_k \in \mathcal{K}_{k+1}(\mathbb{A}^*; \tilde{\mathbf{r}}_0)$
    such that, for some $\theta_k \neq 0$,
    $\tilde{\mathbf{r}}_k - \bar{\theta}_k\, \mathbb{A}^* \tilde{\mathbf{r}}_{k-1} \in \mathcal{K}_k(\mathbb{A}^*; \tilde{\mathbf{r}}_0)$
end while



[Fletcher '76]

## Bi-CG

Choose $\mathbf{x}_0$ and $\tilde{\mathbf{r}}_0$
$k = 0$
$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
$\mathbf{u}_{-1} = 0,\ \rho_{-1} = 1,\quad \tilde{\mathbf{c}}_{-1} = 0$
while $\|\mathbf{r}_k\| > tol$ do
  solve $\mathbf{w}$ from $\mathbf{K}\mathbf{w} = \mathbf{r}_k$
  $\rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_k),\ \beta_k = -\rho_k/\rho_{k-1}$
  $\mathbf{u}_k = \mathbf{w} - \beta_k\, \mathbf{u}_{k-1}$
  $\mathbf{c}_k = \mathbf{A}\mathbf{u}_k$
  $\sigma_k = (\mathbf{c}_k, \tilde{\mathbf{r}}_k),\ \alpha_k = \rho_k/\sigma_k$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\, \mathbf{u}_k$
  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\, \mathbf{c}_k$
    $\tilde{\mathbf{c}}_k = \mathbb{A}^* \tilde{\mathbf{r}}_k - \bar{\beta}_k\, \tilde{\mathbf{c}}_{k-1}$
    $\tilde{\mathbf{r}}_{k+1} = \tilde{\mathbf{r}}_k - \bar{\alpha}_k\, \tilde{\mathbf{c}}_k$
  $k \leftarrow k+1$
end while



[Fletcher '76]

## Bi-CG

Choose $\mathbf{x}_0$ and $\tilde{\mathbf{r}}$
$\mathbf{x} = \mathbf{x}_0,\ \mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$
$\mathbf{u} = 0,\ \rho = 1,\quad \tilde{\mathbf{c}} = 0$
while $\|\mathbf{r}\| > tol$ do
  $\sigma \leftarrow -\rho$
  solve $\mathbf{w}$ from $\mathbf{K}\mathbf{w} = \mathbf{r}$
  $\rho \leftarrow (\mathbf{r}, \tilde{\mathbf{r}}),\ \beta \leftarrow \rho/\sigma$
  $\mathbf{u} \leftarrow \mathbf{w} - \beta\, \mathbf{u}$
  $\mathbf{c} \leftarrow \mathbf{A}\mathbf{u}$
  $\sigma \leftarrow (\mathbf{c}, \tilde{\mathbf{r}}),\ \alpha \leftarrow \rho/\sigma$
  $\mathbf{x} \leftarrow \mathbf{x} + \alpha\, \mathbf{u}$
  $\mathbf{r} \leftarrow \mathbf{r} - \alpha\, \mathbf{c}$
    $\tilde{\mathbf{c}} \leftarrow \mathbb{A}^* \tilde{\mathbf{r}} - \bar{\beta}\, \tilde{\mathbf{c}}$
    $\tilde{\mathbf{r}} \leftarrow \tilde{\mathbf{r}} - \bar{\alpha}\, \tilde{\mathbf{c}}$
end while



## Bi-CG = CG

Instead of solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ we may as well solve

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \tilde{\mathbf{x}} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \tilde{\mathbf{b}} \end{bmatrix} \qquad (\star)$$

$$2\left[\begin{bmatrix} \mathbf{x} \\ \tilde{\mathbf{x}} \end{bmatrix}, \begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}} \end{bmatrix}\right] := \begin{bmatrix} \mathbf{y} \\ \tilde{\mathbf{y}} \end{bmatrix}^* \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \tilde{\mathbf{x}} \end{bmatrix}$$
$$= \mathbf{y}^*\tilde{\mathbf{x}} + \tilde{\mathbf{y}}^*\mathbf{x} = (\mathbf{y}, \tilde{\mathbf{x}}) + (\mathbf{x}, \tilde{\mathbf{y}})$$

defines an inner product except for the positivity requirement (we now do not have $[\vec{\mathbf{x}}, \vec{\mathbf{x}}] \geq 0$ nor $[\vec{\mathbf{x}}, \vec{\mathbf{x}}] = 0 \Leftrightarrow \vec{\mathbf{x}} = 0$).

$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^* \end{bmatrix}$ is symmetric with respect to this "quarter inner product".

Bi-CG = CG applied to the **augmented problem** $(\star)$ with respect to this "inner product".

## Properties Bi-CG

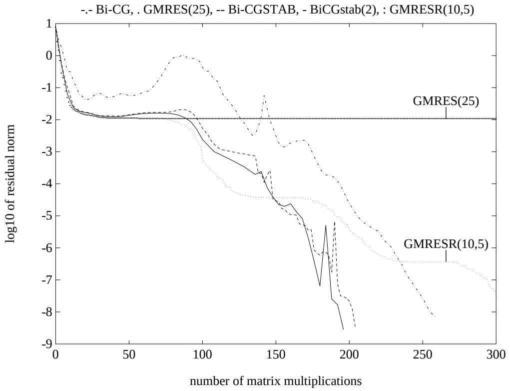
**Advantages**

- Usually selects good approximations from the search subspaces (Krylov subspaces).
- 2 DOT, 5 AXPY per step.
- Storage: 8 large vectors.
- No knowledge on properties of $\mathbb{A}$ is needed.

**Drawbacks**

- ∘ Non-optimal Krylov subspace method.
- Not robust: Bi-CG may break down.
- Bi-CG is sensitive to evaluation errors (often loss of super-linear convergence).
- ∘ Convergence depends on shadow residual $\tilde{\mathbf{r}}_0$.
- 2 MV needed to expand search subspace.
- 1 MV is by $\mathbf{A}^*$.

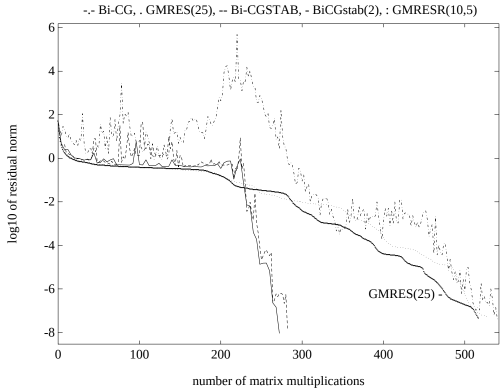





$-(a\,u_x)_x - (a\,u_y)_y = 1$ on $[0,1]\times[0,1]$.
$a = 1000$ for $0.1 \le x, y \le 0.9$
and $a = 1$ elsewhere.

Dirichlet BC on $y = 0$,
Neumann BC on other parts of Boundary.
$82 \times 82$ volumes. ILU Decomp.






$-u_{xx} - u_{yy} + a\,u_x + 100\,u_y = f$ on $[0,1]\times[0,1]$.
$a = 100$ if $x \in [0,\frac{1}{4}] \cup [\frac{1}{2},\frac{3}{4}]$, $a = -100$ elsewhere

$f$ such that $u(x,y) = sin(\pi x)\sin(\pi y)$.
Dirichlet BC. $81 \times 81$ volumes. No precond..



## Expanding a Krylov basis

Suppose $\mathbf{w}_1, \mathbf{w}_2, \ldots$ are linearly independent.

Expand $\mathbf{V}_k = [\mathbf{v}_1|\ldots|\mathbf{v}_k]$ (recursively) with $\mathbf{A}\mathbf{v}_k$ orthogonalized w.r.t. $\mathbf{W}_k = [\mathbf{w}_1|\ldots|\mathbf{w}_k]$:

$\tilde{\mathbf{v}} = \mathbf{A}\mathbf{v}_k$, $M_k = \mathbf{W}_k^*\mathbf{V}_k$
$\vec{h}_k = M_k^{-1}(\mathbf{W}_k^*\tilde{\mathbf{v}})$, $\tilde{\mathbf{v}} \leftarrow \tilde{\mathbf{v}} - \mathbf{V}_k\vec{h}_k$
Select a $h_{k+1\,k} \neq 0$, $\mathbf{v}_{k+1} = \tilde{\mathbf{v}}/h_{k+1\,k}$

**Theorem**. Then $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$ is a Krylov basis and $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\underline{H}_k$, $\underline{H}_k = \begin{bmatrix} \underline{H}_{k-1} & \vec{h}_k \\ 0\ldots 0 & h_{k+1\,k}\end{bmatrix}$

**Examples**. In all examples, $\forall k$

- $\mathbf{w}_k = \mathbf{v}_k$ [Arnoldi '51]
- $\mathbf{w}_1, \ldots, \mathbf{w}_k$ Krylov basis $\mathcal{K}_k(\mathbf{A}^*;\mathbf{w}_1)$ [Lanczos '52]
- $\pi$ "selects" from $\{1,\ldots,n\}$: $\mathbf{w}_k = e_{\pi(k)}$ [Sadok '96]



$\rho_k := \min\{\|\mathbf{r}_0 - \mathbf{A}\mathbf{y}\| \mid \mathbf{y} \in \mathcal{K}_k(\mathbf{A};\mathbf{r}_0)\}$
$\mathbf{r}_0/\rho = \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}$ Krylov basis $\mathcal{K}_{k+1}(\mathbf{A};\mathbf{r}_0)$
$\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\underline{H}_k$

---

Then $\rho_k = \min_{\vec{\gamma}} \|\mathbf{r}_0 - \mathbf{A}\mathbf{V}_k\vec{\gamma}\|$
$= \min_{\vec{\gamma}} \|\mathbf{V}_{k+1}(\rho e_1 - \underline{H}_k\vec{\gamma})\|$

### Quasi minimalization

Solve $\tilde{\rho}_k := \min_{\vec{\gamma}} \|\rho e_1 - \underline{H}_k\vec{\gamma}\|$
With solution $\vec{\gamma}_k$, take $\mathbf{x}_k = \mathbf{x}_0 + \mathbf{V}_k\vec{\gamma}_k$

**Theorem**. $\sigma_1 \le \ldots \le \sigma_k$ singular values $\mathbf{V}_k$
$\Rightarrow$ $\sigma_1\,\tilde{\rho}_k \le \rho_k \le \sigma_k\,\tilde{\rho}_k$,
$\rho_k \le \|\mathbf{r}_k\| = \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\| \le \sigma_k\,\tilde{\rho}_k \le \mathcal{C}(\mathbf{V}_k)\,\rho_k$.

$\sigma_k \le \sqrt{k}\,\max_j \|\mathbf{v}_j\|$



$\mathbf{r}_0/\rho = \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}$ Krylov basis $\mathcal{K}_{k+1}(\mathbf{A};\mathbf{r}_0)$
$\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\underline{H}_k$

---

### Orthogonalization

Solve $H_k\vec{\gamma} = \rho\,e_1$. With solution $\vec{\gamma}_k$, take $\mathbf{y}_k = \mathbf{V}_k\,\vec{\gamma}_k$ and $\mathbf{x}_k = \mathbf{x}_0 + \mathbf{y}_k$

need not exist!

**Theorem**. $\mathbf{y}_k \in \mathcal{K}_k(\mathbf{A};\mathbf{r}_0)$ and
$\mathbf{r}_k = \mathbf{r}_0 - \mathbf{A}\mathbf{y}_k = \mathbf{v}_{k+1}(h_{k+1\,k}\,e_k^*\vec{\gamma}_k) =: \tau_k\,\mathbf{v}_{k+1}$
If $\mathbf{W}_k^*\mathbf{v}_{k+1} = 0$ then $\mathbf{r}_k \perp \mathrm{span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$

*Proof.* $\mathbf{r}_0 - \mathbf{A}\,\mathbf{y}_k = \mathbf{V}_{k+1}\,(\rho\,e_1 - \underline{H}_k\,\vec{\gamma}_k)$

**Theorem**. $|\tau_k| = \tilde{\rho}_k/\sqrt{1 - \tilde{\rho}_k^2/\tilde{\rho}_{k-1}^2}$
with $\tilde{\rho}_k := \min_{\vec{\gamma}} \|\rho\,e_1 - \underline{H}_k\,\vec{\gamma}\|$



$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k+1}$ Krylov basis, $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\underline{H}_k$

---

**Theorem**. $H_k$ is tri-diagonal $\forall k$
if, for some $\mathbf{w}_1$,
$\left.\begin{array}{l}\mathbf{v}_{k+1} \text{ is orthogonal with respect to the} \\ \textbf{shadow Krylov subspace } \mathcal{K}_k(\mathbf{A}^*;\mathbf{w}_1)\end{array}\right\}\forall k\ (\star)$

$\mathbf{w}_1, \ldots, \mathbf{w}_{k+1}$ Krylov basis $\mathcal{K}_{k+1}(\mathbf{A}^*;\mathbf{w}_1)$,
$\mathbf{A}^*\,\mathbf{W}_k = \mathbf{W}_{k+1}\,\underline{F}_k$.
Then, with $M_k := \mathbf{W}_k^*\mathbf{V}_k$:
$(\star)$ holds iff $M_k$ non-singular lower triangular;
$(\star)$ implies $H_k = M_k^{-1}\,((\underline{F}_k)^*\,\underline{M}_k)$
where $\underline{M}_k := \mathbf{W}_{k+1}^*\mathbf{V}_k$ is lower triangular.

*Proof.* $\mathbf{W}_k^*\mathbf{A}\mathbf{V}_k = (\mathbf{W}_k^*\mathbf{V}_{k+1})\underline{H}_k = M_k\,H_k$
$(\underline{F}_k)^*\underline{M}_k = (\underline{F}_k)^*\mathbf{W}_{k+1}^*\mathbf{V}_k = \mathbf{W}_k^*\mathbf{A}\mathbf{V}_k$



[Lanczos '50]

## Bi–Lanczos

$\mathbf{v}_1, \ldots, \mathbf{v}_k$ Krylov basis $\mathcal{K}_k(\mathbf{A};\mathbf{r}_0)$.
$\mathbf{w}_1, \ldots, \mathbf{w}_k$ Krylov basis $\mathcal{K}_k(\mathbf{A}^*;\tilde{\mathbf{r}}_0)$.

**Bi-Lanczos** expands recursively

- $\mathbf{V}_k = [\mathbf{v}_1 \mid \ldots \mid \mathbf{v}_k]$ by $\mathbf{A}\mathbf{v}_k$ orthogonalized w.r.t. $\mathbf{w}_1, \ldots, \mathbf{w}_k$,

& • $\mathbf{W}_k = [\mathbf{w}_1 \mid \ldots \mid \mathbf{w}_k]$ by $\mathbf{A}^*\mathbf{w}_k$ orthogonalized w.r.t. $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

Bi-Lanczos does not **break down** iff $D_{k+1} := \mathbf{W}_{k+1}^*\mathbf{V}_{k+1}$ non-singular diagonal.
$\underline{H}_k$ and $\underline{F}_k$ Hessenberg such that
$\mathbf{A}\,\mathbf{V}_k = \mathbf{V}_{k+1}\,\underline{H}_k$ & $\mathbf{A}^*\,\mathbf{W}_k = \mathbf{W}_{k+1}\,\underline{F}_k$

**Theorem**. $H_k = D_k^{-1}\,F_k^*\,D_k$ is tridiagonal, if bi-Lanczos does not break down.

*Proof.* $F_k^*D_k = \mathbf{W}_k^*\,\mathbf{A}\,\mathbf{V}_k = D_k H_k$

[Lanczos '50]

## Bi-Lanczos

Select a $\mathbf{r}_0$, and a $\tilde{\mathbf{r}}_0$

$k=1$

$\mathbf{v}_1=\mathbf{r}_0/\|\mathbf{r}_0\|$, $\mathbf{v}_0=\mathbf{0}$, $\mathbf{w}_1=\tilde{\mathbf{r}}_0/\|\tilde{\mathbf{r}}_0\|$, $\mathbf{w}_0=\mathbf{0}$

$\gamma_0=0$, $\delta_0=1$, $\tilde{\gamma}_0=0$, $\tilde{\delta}_0=1$

while ... do

- $\delta_k=\mathbf{w}_k^*\mathbf{v}_k$,
- $\tilde{\mathbf{v}}=\mathbf{A}\,\mathbf{v}_k$, $\tilde{\mathbf{w}}=\mathbf{A}^*\,\mathbf{w}_k$
- $\beta_k=\overline{\tilde{\gamma}}_{k-1}\delta_k/\delta_{k-1}$, $\tilde{\beta}_k=\overline{\gamma}_{k-1}\overline{\delta}_k/\overline{\delta}_{k-1}$
- $\tilde{\mathbf{v}}\leftarrow\tilde{\mathbf{v}}-\beta_k\mathbf{v}_{k-1}$, $\tilde{\mathbf{w}}\leftarrow\tilde{\mathbf{w}}-\tilde{\beta}_k\mathbf{w}_{k-1}$
- $\alpha_k=\mathbf{w}_k^*\tilde{\mathbf{v}}/\delta_k$, $\tilde{\alpha}_k=\overline{\alpha}_k$
- $\tilde{\mathbf{v}}\leftarrow\tilde{\mathbf{v}}-\alpha_k\,\mathbf{v}_k$, $\tilde{\mathbf{w}}\leftarrow\tilde{\mathbf{w}}-\tilde{\alpha}_k\,\mathbf{w}_k$
- Select a $\gamma_k\neq0$ and a $\tilde{\gamma}_k\neq0$
- $\mathbf{v}_{k+1}=\tilde{\mathbf{v}}/\gamma_k$, $\mathbf{w}_{k+1}=\tilde{\mathbf{w}}/\tilde{\gamma}_k$,
- $k\leftarrow k+1$
- $\mathbf{V}_k=[\mathbf{V}_{k-1}\,|\,\mathbf{v}_k]$, $\mathbf{W}_k=[\mathbf{W}_{k-1}\,|\,\mathbf{w}_k]$

end while



$\mathbf{v}_1,\ldots,\mathbf{v}_k$ Krylov basis $\mathcal{K}_k(\mathbf{A};\mathbf{r}_0)$.
$\mathbf{w}_1,\ldots,\mathbf{w}_k$ Krylov basis $\mathcal{K}_k(\mathbf{A}^*;\tilde{\mathbf{r}}_0)$.

**Bi-Lanczos** expands recursively

- $\mathbf{V}_k=[\mathbf{v}_1\,|\,\ldots\,|\,\mathbf{v}_k]$ by $\mathbf{A}\mathbf{v}_k$ orthogonalized w.r.t. $\mathbf{w}_1,\ldots,\mathbf{w}_k$,

&
- $\mathbf{W}_k=[\mathbf{w}_1\,|\,\ldots\,|\,\mathbf{w}_k]$ by $\mathbf{A}^*\mathbf{w}_k$ orthogonalized w.r.t. $\mathbf{v}_1,\ldots,\mathbf{v}_k$.

---

## Lanczos breakdowns

Breakdown iff $\mathbf{w}_{k+1}^*\mathbf{v}_{k+1}=0$

- **Lucky** breakdown if $\mathbf{v}_{k+1}=0$.
Then $\mathbf{A}\,\mathbf{V}_k=\mathbf{V}_k T_k$.
- **Serious** or **Lanczos** breakdown:
$\mathbf{v}_{k+1}\neq0$, $\mathbf{w}_{k+1}\neq0$, $\mathbf{w}_{k+1}^*\mathbf{v}_{k+1}=0$

**Cure**: **Look ahead**
i.e. form a non-singular block diagonal $D_{k+\ell}$
$D_{k+\ell}=\mathbf{W}_{k+\ell}^*\mathbf{V}_{k+\ell}$:

Orth. $\mathbf{A}\,\mathbf{v}_{k+1}$ w.r.t. $\mathbf{W}_k$ and $\mathbf{A}^*\,\mathbf{w}_{k+1}$ w.r.t. $\mathbf{V}_k$.

Check non-singularity $[\mathbf{w}_{k+1}|\mathbf{w}_{k+2}]^*[\mathbf{v}_{k+1}|\mathbf{v}_{k+2}]$

...



$\mathbf{r}_0=\rho\,\mathbf{v}_1$, $\mathbf{W}_k^*\mathbf{V}_k=D_k$, $\mathbf{A}\mathbf{V}_k=\mathbf{V}_{k+1}\underline{T}_k$

with $D_k$ non-sing. diag., $\underline{T}_k$ tri-diagonal

Solve $T_k\vec{\gamma}=\rho\,e_1$. With solution $\vec{\gamma}_k$,
take $\mathbf{y}_k=\mathbf{V}_k\,\vec{\gamma}_k$ and $\mathbf{x}_k=\mathbf{x}_0+\mathbf{y}_k$

---

## Bi-Conjugate Gradients

$$T_k=L_k\,U_k \quad \text{need not exist!}$$
$$\mathbf{y}_k=(\mathbf{V}_k\,U_k^{-1})(L_k^{-1}(\rho\,e_1))$$

Expand $\mathbf{V}_k$, $\mathbf{W}_k$, $T_k$
Expand $L_k$ and $U_k$
Expand $\mathbf{U}_k:=\mathbf{V}_k\,U_k^{-1}$
Expand $(\zeta_0,\ldots,\zeta_{k-1})^T:=L_k^{-1}(\rho\,e_1)$
Update $\mathbf{x}_k$ $(=\mathbf{x}_{k-1}+\zeta_{k-1}\,\mathbf{U}_k\,e_k)$

Bi-CG = efficient (in computational costs and storage) implementation of this, using
$$\mathbf{A}\,\mathbf{U}_k=\mathbf{A}\,\mathbf{V}_k\,U_k^{-1}=\mathbf{V}_{k+1}\,\underline{L}_k$$
for expanding $\mathbf{V}_k$.
$(\mathbf{V}_k=[\mathbf{r}_0|\ldots|\mathbf{r}_{k-1}],\ \mathbf{U}_k=[\mathbf{u}_0|\ldots|\mathbf{u}_{k-1}],\ \mathbf{W}_k=[\tilde{\mathbf{r}}_0|\ldots|\tilde{\mathbf{r}}_{k-1}])$



[Freund Nachtigal '91]

$\mathbf{r}_0=\rho\,\mathbf{v}_1$, $\mathbf{W}_k^*\mathbf{V}_k=D_k$, $\mathbf{A}\mathbf{V}_k=\mathbf{V}_{k+1}\underline{T}_k$

with $D_k$ non-sing. diag., $\underline{T}_k$ tri-diagonal

Solve $\min_{\vec{\gamma}}\|\rho e_1-\underline{T}_k\vec{\gamma}\|$. With sol. $\vec{\gamma}_k$,
take $\mathbf{y}_k=\mathbf{V}_k\,\vec{\gamma}_k$ and $\mathbf{x}_k=\mathbf{x}_0+\mathbf{y}_k$.

---

## Quasi Minimal Residual

$$\underline{T}_k=Q_{k+1}\,\underline{R}_k$$

with $Q_{k+1}$ product Givens rotations
$\underline{R}_k$ $(k+1)\times k$ upper triangular

$$\mathbf{y}_k=(\mathbf{V}_k\,R_k^{-1})\vec{\zeta}_k \text{ with } \begin{bmatrix}\vec{\zeta}_k\\ \tilde{\rho}_k\end{bmatrix}:=Q_{k+1}^*(\rho\,e_1)$$

QMR = efficient (in computational costs and storage) implementation of this.



## QMR versus Bi-CG

– Both methods may suffer from
**Lanczos breakdown** $(\rho_k=0)$.

+ Look ahead strategies can easily be incorporated in QMR (much harder in Bi-CG).
*In fact, look ahead forms part of QMR.*

+ In contrast to QMR, Bi-CG may suffer from
**LU-decomposition breakdown**
or **pivot breakdown** $(\sigma_k=0)$.
$(T_k=L_k\,U_k$ need not exist: pivot $\sigma_k$ can be 0)

Cure: composite step Bi-CG [Bank Chan '94]
(uses $2\times2$ pivots if $\sigma_k=0$ or $\sigma_k$ is small).

+? Convergence history QMR tends to be smoother. No faster convergence!!

– QMR is a little less stable
Coupled 2-term recurrences (Bi-CG) appear to be more stable than 3-term recurrence (QMR).

– QMR algorithm is a little more complicated.



## Near Breakdowns

**Bi-CG breaks down if**:

- $\rho_k=0$ – $\mathbf{r}_k=0$ **lucky breakdown**
  – $\tilde{\mathbf{r}}_k=0$
  – $\mathbf{r}_k\neq0$, $\tilde{\mathbf{r}}_k\neq0$ and $\mathbf{r}_k\perp\tilde{\mathbf{r}}_k$
  **Lanczos breakdown**
- $\sigma_k=0$ **pivot breakdown**

*Breakdown hardly ever occurs in practice.*
Cure: restart with another $\tilde{\mathbf{r}}_0$.

However, **near-breakdown** may lead to

- inaccurate coefficients ($\rho_k$ or $\sigma_k$)
- irregular convergence

**Cure**

- $\rho_k\approx0$: look-ahead Lanczos
- $\sigma_k\approx0$: composite step Bi-CG, QMR



## Smoothing the convergence: cosmetic approaches I

$\mathbf{x}_0,\ldots,\mathbf{x}_k$ approximations with residuals
$\mathbf{r}_0,\ldots,\mathbf{r}_k$ produced by some method.

[Neumaier '94]

- To "**smooth**" the convergence,

take $\mathbf{r}_k^{\text{best}}=\mathbf{r}_j$ with $\|\mathbf{r}_j\|=\min_{i\leq k}\|\mathbf{r}_i\|$
$\mathbf{x}_k^{\text{best}}=\mathbf{x}_j$.

∘ *additional storage required*

[Weiss '90]

- To "**smooth**" the convergence,

take $\mathbf{r}_k^{\text{MR}}=(1-\gamma_k)\,\mathbf{r}_{k-1}^{\text{MR}}+\gamma_k\,\mathbf{r}_k$
$\mathbf{x}_k^{\text{MR}}=(1-\gamma_k)\,\mathbf{x}_{k-1}^{\text{MR}}+\gamma_k\,\mathbf{x}_k$

with $\gamma_k$ minimizing $\|(1-\gamma)\,\mathbf{r}_{k-1}^{\text{MR}}+\gamma\,\mathbf{r}_k\|$.

∘ *additional storage & DOTs required*



[Zhou Walker '94]

## Smoothing the convergence cosmetic approaches II

$\mathbf{x}_0,\ldots,\mathbf{x}_k$ approximations with residuals
$\mathbf{r}_0,\ldots,\mathbf{r}_k$ produced by some method.

- To "**smooth**" the convergence,
"**quasi-minimize**" $\|\sum\gamma_j\,\mathbf{r}_j\|$, that is:

$$\mathbf{r}_k^{\text{QMR}}=\sum_{j=0}^{k}\gamma_j\,\mathbf{r}_j \quad\text{with}\quad \sum_{j=0}^{k}\gamma_j^2=1$$

minimizing $\sum_{j=0}^{k}\gamma_j^2\,\|\mathbf{r}_k\|^2$ $(=\tau_k^2)$

Inductively: $\mathbf{r}_k^{\text{QMR}}=(1-\gamma_k)\,\mathbf{r}_{k-1}^{\text{QMR}}+\gamma_k\,\mathbf{r}_k$
$\mathbf{x}_k^{\text{QMR}}=(1-\gamma_k)\,\mathbf{x}_{k-1}^{\text{QMR}}+\gamma_k\,\mathbf{x}_k$

with
$\gamma_k^{-1}=1+\tau_{k-1}^{-2}\|\mathbf{r}_k\|^2$ & $\tau_k^{-2}=\tau_{k-1}^{-2}+\|\mathbf{r}_k\|^{-2}$

∘ *no additional DOTs.* Stable?

**Theorem**. $\|\mathbf{r}_k\|=\tau_k/\sqrt{1-\tau_k^2/\tau_{k-1}^2}$.

[Zhou Walker '94]

## QMR as smoothed Bi-CG

In exact arithmetic, if no breakdown occurs:

**Theorem**. QMR = producing $\mathbf{x}_k^{\mathrm{QMR}}$ and $\mathbf{r}_k^{\mathrm{QMR}}$ from the Bi-CG approximates and residuals by quasi-minimization.

If QMR reduces well, i.e. $\left\|\mathbf{r}_k^{\mathrm{QMR}}\right\| \ll \left\|\mathbf{r}_{k-1}^{\mathrm{QMR}}\right\|$,

then ($\tau_k \ll \tau_{k-1}$ and) $\left\|\mathbf{r}_k^{\mathrm{QMR}}\right\| \approx \left\|\mathbf{r}_k^{\text{Bi-CG}}\right\|$

– *No real improvement on Bi-CG*
*(in exact arithmetic, if no breakdown occurs).*

- *QMR allows look-ahead strategies.*



## Inaccurate coefficients

- Loss of speed of convergence

## Irregular convergence

- Loss of speed of convergence
- Loss of accuracy



$$\mathbf{u}_k = \mathbf{r}_k - \beta_k\,\mathbf{u}_{k-1}$$
$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k$$

## Rounded arithmetic

In rounded arithmetic $\cdot^\star$ computed quantity

$$\mathbf{r}_{k+1}^\star = \mathbf{r}_k^\star - \alpha_k\,\mathbf{A}\,\mathbf{u}_k^\star + \mathbf{f}_k$$

$\mathbf{f}_k$ is the **local** rounding error

**Theorem**. assuming MV main source of error

$$\boxed{\|\mathbf{f}_k\| \lesssim \overline{\xi}\,\Gamma\left(\|\mathbf{r}_{k+1}^\star\| + \|\mathbf{r}_k^\star\|\right)}$$

where $\overline{\xi}$ is the **relative machine precision**

and $\Gamma := n_A\,\left\||\mathbf{A}|\right\|\,\left\|\mathbf{A}^{-1}\right\|$

with $n_A :=$ max # non-zeros per row of $\mathbf{A}$

*Proof.* $\|\mathbf{f}_k\| \lesssim \overline{\xi}\,n_A\,\||\mathbf{A}|\|\,\|\alpha_k\mathbf{u}^\star\| \le \overline{\xi}\,\Gamma\,\|\mathbf{A}(\alpha_k\mathbf{u}_k^\star)\|$



$$\mathbf{u}_k = \mathbf{r}_k - \beta_k\,\mathbf{u}_{k-1}$$
$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k$$

In rounded arithmetic

$$\mathbf{r}_{k+1}^\star = \mathbf{r}_k^\star - \alpha_k\,\mathbf{A}\,\mathbf{u}_k^\star + \mathbf{f}_k$$
$$\|\mathbf{f}_k\| \lesssim \overline{\xi}\,\Gamma\left(\|\mathbf{r}_{k+1}^\star\| + \|\mathbf{r}_k^\star\|\right)$$

## Perturbed convergence

**Experimental result**

Convergence depends very sensitively on $\mathbf{f}_k$.

**Speed** of convergence does **not** depend heavily on $\mathbf{f}_k$ if $\|\mathbf{f}_k\| \ll \|\mathbf{r}_{k+1}^\star\|$

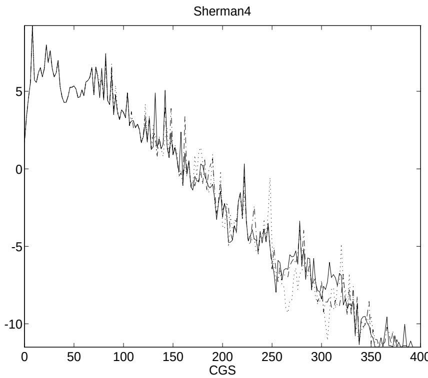
Speed of convergence may be disturbed seriously if $\|\mathbf{f}_k\| \not\ll \|\mathbf{r}_{k+1}^\star\|$

Therefore,
irregularities harmless if $\overline{\xi}\,\Gamma\,\|\mathbf{r}_k^\star\| \ll \|\mathbf{r}_{k+1}^\star\|$



$$\mathbf{r}_{k+1}^\star = \mathbf{r}_k^\star - \alpha_k\,\mathbf{A}\,\mathbf{u}_k^\star + \mathbf{f}_k, \quad \|\mathbf{f}_k\| \ll \|\mathbf{r}_{k+1}^\star\|$$

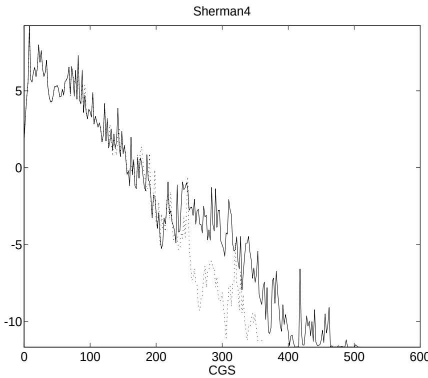
## Small pert. in (hybrid) Bi-CG






$$\mathbf{r}_{k+1}^\star = \mathbf{r}_k^\star - \alpha_k\,\mathbf{A}\,\mathbf{u}_k^\star + \mathbf{f}_k, \quad \|\mathbf{f}_k\| \not\ll \|\mathbf{r}_{k+1}^\star\|$$

## Large pert. in (hybrid) Bi-CG






[S., van der Vorst, Fokkema '94, Greenbaum '94 '96]

$$\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k$$
$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\,\mathbf{u}_k$$

In rounded arithmetic

$$\mathbf{r}_{k+1}^\star = \mathbf{r}_k^\star - \alpha_k\,\mathbf{A}\,\mathbf{u}_k^\star + \mathbf{f}_k$$
$$\|\mathbf{f}_k\| \lesssim \overline{\xi}\,\Gamma\left(\|\mathbf{r}_{k+1}^\star\| + \|\mathbf{r}_k^\star\|\right)$$

## Accuracy

*True residual*

$$\mathbf{b} - \mathbf{A}\mathbf{x}_{j+1}^\star = \mathbf{b} - \mathbf{A}\mathbf{x}_j^\star - \mathbf{A}(\alpha_j\,\mathbf{u}_j^\star)$$

*and recursively updated residual*

$$\mathbf{r}_{j+1}^\star \quad = \quad \mathbf{r}_j^\star \quad - \alpha_j\,\mathbf{A}\mathbf{u}_j^\star + \mathbf{f}_j$$

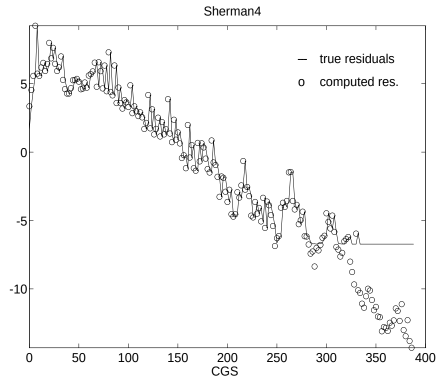
*drift apart.*

**Theorem**.

$$\boxed{\begin{array}{l}\|(\mathbf{b} - \mathbf{A}\mathbf{x}_k^\star) - \mathbf{r}_k^\star\| \le 2\,\overline{\xi}\,\Gamma\displaystyle\sum_{j\le k}\|\mathbf{r}_j^\star\| \\ \quad\text{where} \quad \Gamma := n_A\,\left\||\mathbf{A}|\right\|\,\left\|\mathbf{A}^{-1}\right\|\end{array}}$$



## Large residuals in (hybrid) Bi-CG



[S., van der Vorst '96]

## Avoiding inaccurate solutions

Use $\quad \mathbf{r}_{k+1} \leftarrow \mathbf{b} - \mathbf{A}\,\mathbf{x}_{k+1} \quad$ in each step?

$$\begin{aligned} \mathbf{r}^\star_{k+1} &= \mathbf{b} - (\mathbf{A}\mathbf{x}^\star_{k+1})^\star = \mathbf{b} - \mathbf{A}\mathbf{x}^\star_{k+1} + \mathbf{f}'_k \\ &= \mathbf{b} - \mathbf{A}\mathbf{x}^\star_k - \mathbf{A}(\alpha_k \mathbf{u}^\star_k) + \mathbf{f}'_k \\ &= \mathbf{b} - (\mathbf{A}\mathbf{x}^\star_k)^\star - \mathbf{A}(\alpha_k \mathbf{u}^\star_k) + \mathbf{f}''_k + \mathbf{f}'_k \end{aligned}$$

Additional recursion error:

$$\|\mathbf{f}_k\| \leq \bar{\xi}\,\Gamma\,\left(2\,\|\,\mathbf{b}\,\| + \|\,\mathbf{r}^\star_k\,\| + \|\,\mathbf{r}^\star_{k+1}\,\|\right)$$

**Disadvantages:**

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

Use $\quad \mathbf{r}_{k+1} \leftarrow \mathbf{b} - \mathbf{A}\,\mathbf{x}_{k+1} \quad$ occasionally!

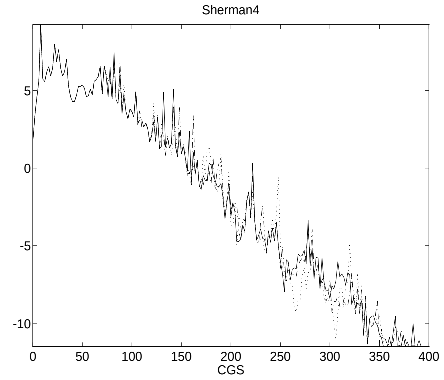
Additional recursion error:

$$\|\mathbf{f}\| \leq \bar{\xi}\,\Gamma\,\left(2\,\|\,\mathbf{b}\,\| + \sum \|\,\mathbf{r}^\star_j\,\|\right),$$

summing over all res. since previous use of true res.



Use $\quad \mathbf{r}_{k+1} \leftarrow \mathbf{b} - \mathbf{A}\,\mathbf{x}_{k+1} \quad$ only if the residual-norm decreases significantly and is not (much) smaller than $\|\,\mathbf{b}\,\|$.




7 true residual computations



## Bi-Conjugate Gradients $\quad \mathbf{K} = \mathbf{I}$

$$\boxed{\begin{aligned} \mathbf{u}_k &= \mathbf{r}_k - \beta_k\,\mathbf{u}_{k-1} \\ \mathbf{r}_{k+1} &= \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\mathbf{u}_k \end{aligned}}$$

$$\mathbf{r}_k,\ \mathbf{A}\mathbf{u}_k \perp \bar{q}_{k-1}(\mathbf{A}^*)\tilde{\mathbf{r}}_0$$

with $q_{k-1}$ a polynomial of exact degree $k-1$.

---

With

$\rho_k := (\mathbf{r}_k, \bar{q}_k(\mathbf{A}^*)\tilde{\mathbf{r}}_0) \quad \& \quad \sigma_k := (\mathbf{A}\mathbf{u}_k, \bar{q}_k(\mathbf{A}^*)\tilde{\mathbf{r}}_0)$

and $\theta_k$ such that

$$q_k(\zeta) - \theta_k\,\zeta\,q_{k-1}(\zeta) \in \mathcal{P}_{k-1}$$

we have

$$\alpha_k = \frac{\rho_k}{\sigma_k} \qquad \& \qquad \beta_k = \frac{\rho_k}{\theta_k\,\sigma_{k-1}}$$



[Sonneveld '89]

## Transpose-free Bi-CG

$\rho_k = (\mathbf{r}_k, \bar{q}_k(\mathbf{A}^*)\,\tilde{\mathbf{r}}_0) = (q_k(\mathbf{A})\mathbf{r}_k, \tilde{\mathbf{r}}_0)$,

$\sigma_k = (\mathbf{A}\mathbf{u}_k, \bar{q}_k(\mathbf{A}^*)\,\tilde{\mathbf{r}}_0) = (\mathbf{A}q_k(\mathbf{A})\mathbf{u}_k, \tilde{\mathbf{r}}_0)$

$\mathbf{Q}_k := q_k(\mathbf{A})$

(Bi-CG) $\left\{ \begin{array}{ll} \rho_k, & \mathbf{Q}_k\mathbf{u}_k = \mathbf{Q}_k\mathbf{r}_k - \beta_k\,\mathbf{Q}_k\mathbf{u}_{k-1}, \\ \sigma_k, & \mathbf{Q}_k\mathbf{r}_{k+1} = \mathbf{Q}_k\mathbf{r}_k - \alpha_k\,\mathbf{A}\mathbf{Q}_k\mathbf{u}_k, \end{array} \right.$

(Pol) Compute $q_{k+1}$ of degree $k+1$ such that $q_{k+1}(0) = q_k(0) = 1$.

Compute

$$\hat{\hat{\mathbf{u}}}_k := \mathbf{Q}_{k+1}\mathbf{u}_k, \quad \hat{\hat{\mathbf{r}}}_{k+1} := \mathbf{Q}_{k+1}\mathbf{r}_{k+1}$$



## Transpose-free Bi-CG; Practice

Work with $\hat{\mathbf{u}}_k = \mathbf{Q}_k\mathbf{u}_k$ and $\hat{\mathbf{r}}_k = \mathbf{Q}_k\mathbf{r}_{k+1}$

Write $\mathbf{u}_k$, $\mathbf{r}_k$, $\mathbf{x}_k$ instead of $\hat{\hat{\mathbf{u}}}_k$, $\hat{\hat{\mathbf{r}}}_k$, $\hat{\hat{\mathbf{x}}}_k$, resp.

$\rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_0)$,

$\sigma_k = (\mathbf{A}\,\hat{\mathbf{u}}_k, \tilde{\mathbf{r}}_0)$

(Bi-CG) $\left\{ \begin{array}{ll} \rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_0), & \hat{\mathbf{u}}_k = \mathbf{r}_k - \beta_k\mathbf{u}_{k-1}, \\ \sigma_k = (\mathbf{A}\,\hat{\mathbf{u}}_k, \tilde{\mathbf{r}}_0), & \hat{\mathbf{r}}_k = \mathbf{r}_k - \alpha_k\,\mathbf{A}\,\hat{\mathbf{u}}_k, \\ & \hat{\mathbf{x}} = \mathbf{x}_k + \alpha_k\hat{\mathbf{u}}_k \end{array} \right.$

(Pol) Compute updating coefficients for $q_{k+1}$.

Compute

$$\mathbf{u}_k, \qquad \mathbf{r}_{k+1}, \qquad \mathbf{x}_{k+1}$$



## Hybrid Bi-CG

$$\boxed{\mathbf{r}_k = q_k(\mathbf{A})\,\phi_k^{\text{Bi-CG}}(\mathbf{A})\,\mathbf{r}_0}$$

$r_k := \mathbf{r}_k^{\text{Bi-CG}} := \phi_k^{\text{Bi-CG}}(\mathbf{A})\mathbf{r}_0$:

$\phi_k^{\text{Bi-CG}}$ is the "Bi-CG polynomial" of degree $k$

$q_k$??

$q_k$ for <u>efficient steps</u> & <u>fast convergence</u>.

Fast convergence by

- reducing the residual
- stabilizing the Bi-CG part
- other — when used as linear solver for the Jacobian system in a Newton scheme for non-linear equations, by reducing the number of Newton steps



## Properties hybrid Bi-CG

**Advantages**

- Converges often twice as fast as Bi-CG w.r.t. # MVs:

each MV expands the search subspace

Bi-CG: $\quad \mathbf{x}_k - \mathbf{x}_0 \in \mathcal{K}_k(\mathbf{A}; \mathbf{r}_0)$ á 2k MV.

Hybrid Bi-CG: $\quad \mathbf{x}_k - \mathbf{x}_0 \in \mathcal{K}_{2k}(\mathbf{A}; \mathbf{r}_0)$ á 2k MV.

- Work/MV and storage similar to Bi-CG.
- Transpose free.
- ◦ Explicit computation of Bi-CG scalars.

**Drawbacks**

- Non-optimal Krylov subspace method.
- Peaks in the convergence history.
- Large intermediate residuals.
- ◦ Breakdown possibilities.



[Sonneveld '89]

## CGS: Conjugate Gradients Squared

$$q_k = \phi_k^{\text{Bi-CG}}, \quad \mathbf{r}_k = \phi_k^{\text{Bi-CG}}(\mathbf{A})\,\phi_k^{\text{Bi-CG}}(\mathbf{A})\,\mathbf{r}_0$$

**CGS** exploits recurrence relations for the Bi-CG polynomials to design a very efficient algorithm.

**Properties**

+ Hybrid Bi-CG.
+ A very efficient algorithm:
  1 DOT/MV, 3.25 AXPY/MV;
  storage: 7 large vectors.
− Often high peaks in its convergence history
− Often large intermediate residuals
+ Seems to do well as linear solver in a Newton scheme

## CGS

Put $\quad r_k := \mathbf{r}_k^{\text{Bi-CG}}, \quad u_k := \mathbf{u}_k^{\text{Bi-CG}}$

$$\boxed{r_k = \phi_k(\mathbf{A})\mathbf{r}_0 = \Phi_k \mathbf{r}_0, \; u_k = \psi_k(\mathbf{A})\mathbf{r}_0 = \Psi_k \mathbf{r}_0}$$

$$\text{(Bi–CG)} \quad \left\{ \begin{array}{rcl} u_k &=& r_k - \beta_k u_{k-1} \ \text{(U)} \\ r_{k+1} &=& r_k - \alpha_k \mathbf{A} u_k \ \text{(R)} \end{array} \right.$$

$$\text{(Pol)} \quad \left\{ \begin{array}{rcl} \Psi_k &=& \Phi_k - \beta_k \Psi_{k-1} \ \text{(U)} \\ \Phi_{k+1} &=& \Phi_k - \alpha_k \mathbf{A} \Psi_k \ \text{(R)} \end{array} \right.$$

$$\theta_k = -\alpha_{k-1} \quad \Rightarrow \quad \beta_k = -\frac{\rho_k}{\rho_{k-1}}$$

$\psi_k - \phi_k$ degree $< k$

$$\Rightarrow \left\{ \begin{array}{l} \rho_k = (\Phi_k r_k, \tilde{\mathbf{r}}_0) = (\Psi_k r_k, \tilde{\mathbf{r}}_0) \\ \sigma_k = (\mathbf{A}\Phi_k u_k, \tilde{\mathbf{r}}_0) = (\mathbf{A}\Psi_k u_k, \tilde{\mathbf{r}}_0) \end{array} \right.$$



| | | |
|---|---|---|
| Bi-CG(1) | $\Phi_k u_k = \Phi_k r_k - \beta_k \Phi_k u_{k-1}$ | $\Phi_k *$ (U) |
| Bi-CG(2) | $\Psi_k u_k = \Psi_k r_k - \beta_k \Psi_k u_{k-1}$ | $\Psi_k *$ (U) |
| Bi-CG(3) | $\Psi_k r_{k+1} = \Psi_k r_k - \alpha_k \mathbf{A} \Psi_k u_k$ | $\Psi_k *$ (R) |
| Bi-CG(4) | $\Phi_k r_{k+1} = \Phi_k r_k - \alpha_k \mathbf{A} \Phi_k u_k$ | $\Phi_k *$ (R) |
| Pol(5) | $\Phi_{k+1} r_{k+1} = \Phi_k r_{k+1} - \alpha_k \mathbf{A} \Psi_k r_{k+1}$ | (R) $* r_{k+1}$ |
| Pol(6) | $\Phi_{k+1} u_k = \Phi_k u_k - \alpha_k \mathbf{A} \Psi_k u_k$ | (R) $* u_k$ |
| Pol(7) | $\Psi_{k+1} u_k = \Phi_{k+1} u_k - \beta_{k+1} \Psi_k u_k$ | (U) $* u_k$ |
| Pol(8) | $\Psi_{k+1} r_{k+1} = \Phi_{k+1} r_{k+1} - \beta_{k+1} \Psi_k r_{k+1}$ | (U) $* r_{k+1}$ |

Since $\quad \Psi_k r_k = \Psi_k \Phi_k \mathbf{r}_0 = \Phi_k u_k, \quad \Psi_k r_{k+1} = \Phi_{k+1} u_k$

we have $\quad (6) = (3) \quad$ and $\quad (8) = (1) \quad$ next sweep.

Combine (4) & (5) to save one MV.

Skip (6) & (8). Start at (7). Put

$\mathbf{r}_k := \Phi_k r_k, \quad \mathbf{u}_k := \Phi_{k+1} u_k = \Psi_k r_{k+1}, \quad \mathbf{w}_k := \Psi_k u_k$



[Sonneveld '89]

## CGS

Choose $\mathbf{x}_0$ and $\tilde{\mathbf{r}}_0$

$k = 0$

$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$

$\mathbf{u}_{-1} = \mathbf{w}_{-1} = 0, \; \rho_{-1} = 1$

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

$\quad \rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_0), \quad \beta_k = -\rho_k/\rho_{k-1}$

$\quad \mathbf{w} = \mathbf{u}_{k-1} - \beta_k \mathbf{w}_{k-1}$

$\quad \tilde{\mathbf{u}} = \mathbf{r}_k - \beta_k \mathbf{u}_{k-1}$

$\quad \mathbf{w}_k = \tilde{\mathbf{u}} - \beta_k \mathbf{w}$

$\quad \mathbf{c} = \mathbf{A}\mathbf{w}_k$

$\quad \sigma_k = (\mathbf{c}, \tilde{\mathbf{r}}_0), \quad \alpha_k = \rho_k/\sigma_k$

$\quad \mathbf{u}_k = \tilde{\mathbf{u}} - \alpha_k \mathbf{c}$

$\quad \mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{A}(\tilde{\mathbf{u}} + \mathbf{u}_k)$

$\quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k(\tilde{\mathbf{u}} + \mathbf{u}_k)$

$\quad k \leftarrow k + 1$

end while



[Sonneveld '89]

## CGS

Choose $\mathbf{x}_0$ and $\tilde{\mathbf{r}}_0$.

$\mathbf{x} = \mathbf{x}_0, \; \mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$.

$\mathbf{u} = \mathbf{w} = 0, \; \rho = 1$.

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

$\quad \sigma \leftarrow -\rho$

$\quad \rho \leftarrow (\mathbf{r}, \tilde{\mathbf{r}}_0), \quad \beta \leftarrow \rho/\sigma$

$\quad \mathbf{w} \leftarrow \mathbf{u} - \beta\, \mathbf{w}$

$\quad \mathbf{v} \leftarrow \mathbf{r} - \beta\, \mathbf{u}$

$\quad \mathbf{w} \leftarrow \mathbf{v} - \beta\, \mathbf{w}$

$\quad \mathbf{c} \leftarrow \mathbf{A}\mathbf{w}$

$\quad \sigma \leftarrow (\mathbf{c}, \tilde{\mathbf{r}}_0), \quad \alpha \leftarrow \rho/\sigma$

$\quad \mathbf{u} \leftarrow \mathbf{v} - \alpha\, \mathbf{c}$

$\quad \mathbf{r} \leftarrow \mathbf{r} - \alpha\, \mathbf{A}(\mathbf{v} + \mathbf{u})$

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

end while



## Why high peaks and large res.?

**Model.** $\quad \mathbf{A}\mathbf{v}_m = \lambda_m \mathbf{v}_m, \qquad \mathbf{r}_0 = \sum \gamma_m \mathbf{v}_m$

in Bi-CG: $\quad r_k = \sum \gamma_m \phi_k(\lambda_m) \mathbf{v}_m$

$|\gamma_k \phi_k(\lambda_m)|$ *moderate or small*

in **CGS**: $\quad \mathbf{r}_k = \sum \gamma_m (\phi_k(\lambda_m))^2 \mathbf{v}_m$

*is* $|\gamma_k(\phi_k(\lambda_m))^2|$ *moderate or small?*

## Why nice in a Newton scheme?

For extremal $\lambda_m$ we may expect

$|\phi_k(\lambda_m)|$ small also for non-large $m$.

Hence $|\gamma_k(\phi_k(\lambda_m))^2|$ very small.

**In**exact Newton with **CGS**

$\approx$ exact Newton w.r.t. extremal components.



How

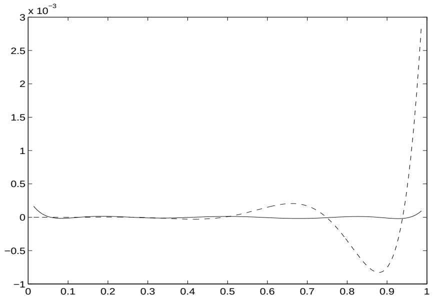
to preserve the reducing effect of squaring

and to get rid of the amplifying effects?

**Idea**. Take, for instance, $\quad q_k = \tilde{\phi}_k \quad$ with

$\tilde{\phi}_k \quad$ Bi-CG polynomial for $\mathbf{A}$,

but with another $\tilde{\mathbf{r}}_0$, say, $\quad \tilde{\mathbf{s}}_0 \neq \tilde{\mathbf{r}}_0$.






[Fokkema, S, van der Vorst '96]

## Generalized CGS

Choose $\mathbf{x}_0$, $\tilde{\mathbf{r}}_0$ and $\tilde{\mathbf{s}}_0$.

$\mathbf{x} = \mathbf{x}_0, \; \mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}_0$.

$\mathbf{v} = \mathbf{w} = \mathbf{u} = 0, \; \alpha = \tilde{\sigma} = \tilde{\alpha} = \sigma = 1$.

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

$\quad \sigma \leftarrow -\tilde{\alpha}\sigma, \quad \rho \leftarrow (\mathbf{r}, \tilde{\mathbf{r}}_0), \quad \beta \leftarrow \rho/\sigma$

$\quad \mathbf{w} \leftarrow \mathbf{u} - \beta\, \mathbf{w}$

$\quad \mathbf{v} \leftarrow \mathbf{r} - \beta\, \mathbf{v}$

$\quad \tilde{\sigma} \leftarrow -\alpha\tilde{\sigma}, \quad \tilde{\rho} \leftarrow (\mathbf{r}, \tilde{\mathbf{r}}_0), \quad \tilde{\beta} \leftarrow \tilde{\rho}/\tilde{\sigma}$

$\quad \mathbf{u} \leftarrow \mathbf{r} - \tilde{\beta}\, \mathbf{u}$

$\quad \mathbf{w} \leftarrow \mathbf{v} - \tilde{\beta}\mathbf{w}, \quad \mathbf{c} \leftarrow \mathbf{A}\mathbf{w}$

$\quad \sigma \leftarrow (\mathbf{c}, \tilde{\mathbf{r}}_0), \; \alpha \leftarrow \rho/\sigma, \; \tilde{\sigma} \leftarrow (\mathbf{c}, \tilde{\mathbf{s}}_0), \; \tilde{\alpha} \leftarrow \tilde{\rho}/\tilde{\sigma}$

$\quad \mathbf{v} \leftarrow \mathbf{v} - \tilde{\alpha}\, \mathbf{c}$

$\quad \mathbf{r} \leftarrow \mathbf{r} - \mathbf{A}(\alpha\, \mathbf{v} + \tilde{\alpha}\, \mathbf{u})$

$\quad \mathbf{x} \leftarrow \mathbf{x} + \alpha\, \mathbf{v} + \tilde{\alpha}\, \mathbf{u}$

$\quad \mathbf{u} \leftarrow \mathbf{u} - \alpha\, \mathbf{c}$

end while



[van der Vorst '92]

## Bi-CGSTAB: Bi-CG STABilized

Put $\quad r_k := \mathbf{r}_k^{\text{Bi-CG}}, \quad u_k := \mathbf{u}_k^{\text{Bi-CG}}$

$$q_{k+1}(t) = (1 - \omega_k t) q_k$$

$$\begin{array}{rcl} \mathbf{Q}_{k+1} r_{k+1} &=& (\mathbf{I} - \omega_k \mathbf{A})\, \mathbf{Q}_k r_{k+1}, \\ && \mathbf{x}_{k+1} = \hat{\mathbf{x}} + \omega_k \mathbf{Q}_k r_{k+1} \\ \mathbf{Q}_{k+1} u_k &=& (\mathbf{I} - \omega_k \mathbf{A})\, \mathbf{Q}_k u_k \end{array}$$

where $\omega_k$ minimizes $\quad \|(\mathbf{I} - \omega\mathbf{A})\, \underbrace{\mathbf{Q}_k r_{k+1}}_{\hat{\mathbf{r}}}\|$

$$\omega_k = \frac{(\hat{\mathbf{r}}, \mathbf{A}\hat{\mathbf{r}})}{(\mathbf{A}\hat{\mathbf{r}}, \mathbf{A}\hat{\mathbf{r}})} \quad \text{and} \quad \theta_k = -\omega_{k-1}$$

[van der Vorst '92]

## Bi-CGSTAB

Choose $\mathbf{x}_0$ and $\tilde{\mathbf{r}}_0$
$k = 0$
$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$
$\mathbf{u}_{-1} = 0, \quad \omega_{-1} = \sigma_{-1} = 1$
while $\|\mathbf{r}_k\| > tol$ do
  $\rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_0), \ \beta_k = -\rho_k/(\omega_{k-1}\sigma_{k-1})$
  $\hat{\mathbf{u}} \leftarrow \mathbf{r}_k - \beta_k\, \mathbf{u}_{k-1}, \qquad \hat{\mathbf{c}} \leftarrow \mathbf{A}\hat{\mathbf{u}}$
  $\sigma_k = (\hat{\mathbf{c}}, \tilde{\mathbf{r}}_0), \quad \alpha_k = \rho_k/\sigma_k$
  $\hat{\mathbf{x}} \leftarrow \mathbf{x}_k + \alpha_k\, \hat{\mathbf{u}}$
  $\hat{\mathbf{r}} \leftarrow \mathbf{r}_k - \alpha_k\, \hat{\mathbf{c}}, \qquad \hat{\mathbf{s}} \leftarrow \mathbf{A}\hat{\mathbf{r}}$
  $\omega_k = (\hat{\mathbf{r}}, \hat{\mathbf{s}})/(\hat{\mathbf{s}}, \hat{\mathbf{s}})$
  $\mathbf{u}_k = \hat{\mathbf{u}} - \omega_k\, \hat{\mathbf{c}}$
  $\mathbf{x}_{k+1} = \hat{\mathbf{x}} + \omega_k\, \hat{\mathbf{r}}$
  $\mathbf{r}_{k+1} = \hat{\mathbf{r}} - \omega_k\, \hat{\mathbf{s}}$
  $k \leftarrow k + 1$
end while



[van der Vorst '92]

## Bi-CGSTAB

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



## Properties Bi-CGSTAB

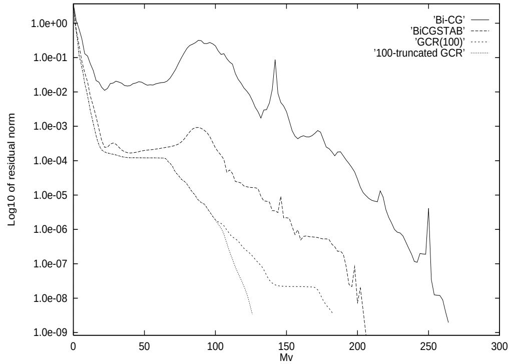
**Advantages**

- Hybrid Bi-CG.
- Converges faster (& smoother) than CGS.
- More accurate than CGS.
- 2 DOT/MV, 3 AXPY/MV.
- Storage: 6 large vectors.

**Drawbacks** Danger of

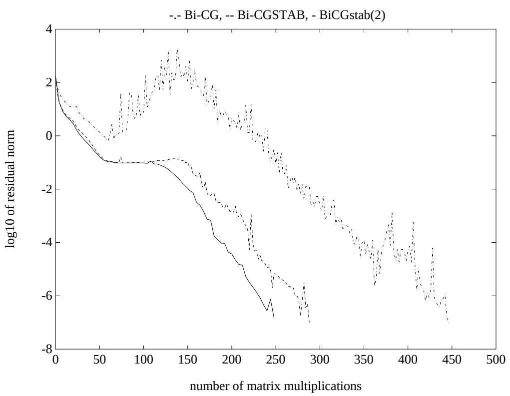
(A) Lanczos breakdown $(\rho_k = 0)$,
(B) pivot breakdown $(\sigma_k = 0)$,
(C) breakdown minimization $(\omega_k = 0)$.





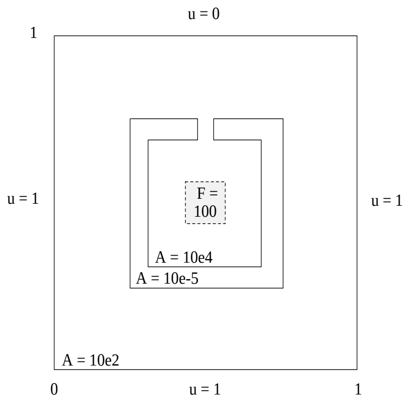

$-(a\, u_x)_x - (a\, u_y)_y = 1$ on $[0,1] \times [0,1]$.
$a = 1000$ for $0.1 \le x, y \le 0.9$
and $a = 1$ elsewhere.
Dirichlet BC on $y = 0$,
Neumann BC on other parts of Boundary.
$82 \times 82$ volumes. ILU Decomp.





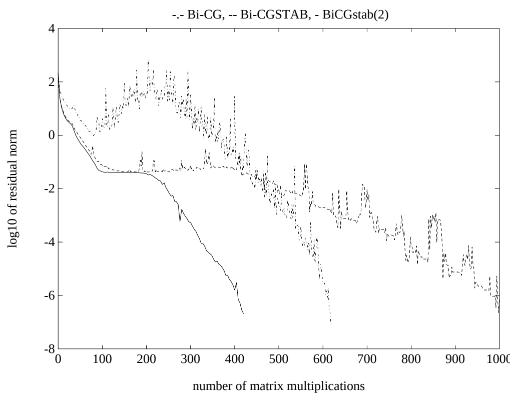

$-(a\, u_x)_x - (a\, u_y)_y + b u_x = f$ on $[0,1] \times [0,1]$.
$b(x,y) = 2\exp(2(x^2 + y^2))$, $a$ changes strongly
Dirichlet BC
$129 \times 129$ volumes. ILU Decomp.






$-(a\, u_x)_x - (a\, u_y)_y + b\, u_x = f$ on $[0,1] \times [0,1]$.
definition of $a = A$ and of $f = F$; $F = 0$ except ...






$-(a\, u_x)_x - (a\, u_y)_y + b u_x = f$ on $[0,1] \times [0,1]$.
$b(x,y) = 2\exp(2(x^2 + y^2))$, $a$ changes strongly
Dirichlet BC
$201 \times 201$ volumes. ILU Decomp.



## Breakdown of the minimization

**Exact arithmetic, $\omega_k = 0$:**

– No reduction of residual by

$$\mathrm{Q}_{k+1}\, r_{k+1} = (\mathbf{I} - \omega_k \mathbf{A}\,)\, \mathrm{Q}_k\, r_{k+1}. \qquad (\star)$$

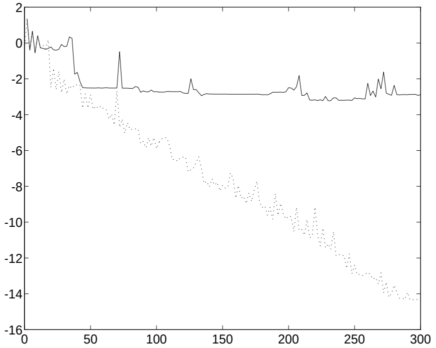
– $q_{k+1}$ is of degree $k$:
Bi-CG scalars can not be computed;
breakdown of incorporated Bi-CG.

**Finite precision arithmetic, $\omega_k \approx 0$:**

– Poor reduction of residual by $(\star)$

– Bi-CG scalars are seriously affected by evaluation errors: drop of speed of convergence.

$\omega_k \approx 0$ to be expected if
$\mathbf{A}$ has eigenvalues with rel. large imaginary part
$\omega_k$ is real!



$$u_{xx} + u_{yy} + u_{zz} + 1000\, u_x = f.$$

$f$ is defined by the solution
$u(x, y, z) = \exp(xyz)\sin(\pi x)\sin(\pi y)\sin(\pi z)$.

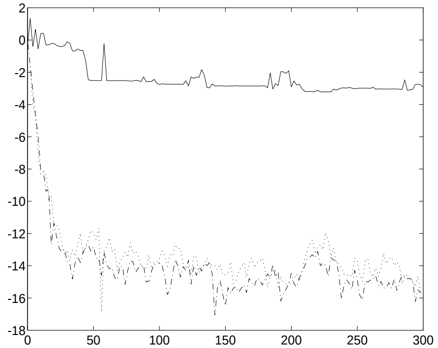
Discretized using $(10 \times 10 \times 10)$ volumes.
No preconditioning was used.



$$\rho_k = (\mathbf{r}_k, \tilde{\mathbf{r}}_0), \qquad \rho_k^\star = \rho_k(1 + \epsilon)$$

## Accurate Bi-CG coefficients

$$|\epsilon| \le n\,\overline{\xi}\,\frac{\|\mathbf{r}_k\|\,\|\tilde{\mathbf{r}}_0\|}{|(\mathbf{r}_k, \tilde{\mathbf{r}}_0)|} \quad = \frac{n\,\overline{\xi}}{\widehat{\rho}_k} \quad \text{where } \widehat{\rho}_k := \frac{|(\mathbf{r}_k, \tilde{\mathbf{r}}_0)|}{\|\mathbf{r}_k\|\,\|\tilde{\mathbf{r}}_0\|}$$





For optimal accurate Bi-CG coefficient:

**minimize**

$$\frac{\|\,q_k(\mathbf{A})\,r_k\|\,\|\tilde{\mathbf{r}}_0\|}{|(q_k(\mathbf{A})\,r_k, \tilde{\mathbf{r}}_0)|} = \frac{\|\,q_k(\mathbf{A})\,r_k\|}{|\theta_k|} \cdot \frac{\|\tilde{\mathbf{r}}_0\|}{|(r_k, (\mathbf{A}^*)^k\,\tilde{\mathbf{r}}_0)|}$$

$$\begin{aligned}(q_k(\mathbf{A})\,r_k, \tilde{\mathbf{r}}_0) &= (r_k, q_k(\mathbf{A}^*)\,\tilde{\mathbf{r}}_0) \\ &= (r_k, \theta_k \cdot (\mathbf{A}^*)^k\,\tilde{\mathbf{r}}_0) \\ & \qquad r_k \perp \mathcal{K}_k(\mathbf{A}^*; \tilde{\mathbf{r}}_0)\end{aligned}$$

**Solution**

$$\left.\begin{array}{l} q_k(\mathbf{A})\,r_k \perp \mathcal{K}_k(\mathbf{A}; r_k) \\ q_k(0) = 1 \end{array}\right\} \quad \begin{array}{l} q_k \text{ is the} \\ \text{FOM polynomial} \end{array}$$



For minimal residuals:

**minimize**

$$\|\,q_k(\mathbf{A})\,r_k\|$$

**Solution**

$$\left.\begin{array}{l} q_k(\mathbf{A})\,r_k \perp \mathbf{A}\mathcal{K}_k(\mathbf{A}; r_k) \\ q_k(0) = 1 \end{array}\right\} \quad \begin{array}{l} q_k \text{ is the} \\ \text{GMRES polynomial} \end{array}$$



[Sleijpen, van der Vorst '95]

$$q_{k+1}(t) = (1 - \omega_k t)\,q_k(t)$$

$$\hat{\mathbf{r}} = q_k(\mathbf{A})r_{k+1}, \qquad \widehat{\omega} := \frac{(\,\hat{\mathbf{r}}\,, \mathbf{A}\,\hat{\mathbf{r}}\,)}{\|\,\hat{\mathbf{r}}\,\|\,\|\,\mathbf{A}\,\hat{\mathbf{r}}\,\|}$$

## Bi-CGSTAB, I

For minimal residuals:

$$\|\mathbf{r}^{\mathrm{MR}}\| = \|(\mathbf{I} - \omega\mathbf{A})\,\hat{\mathbf{r}}\,\| \quad \text{minimal for} \quad \omega = \omega^{\mathrm{MR}} = \widehat{\omega}\,\frac{\|\,\hat{\mathbf{r}}\,\|}{\|\,\mathbf{A}\,\hat{\mathbf{r}}\,\|}$$

For optimal accurate Bi-CG coefficient:

$$\frac{\|\mathbf{r}^{\mathrm{OR}}\|}{|\,\omega\,|} = \frac{\|(\mathbf{I} - \omega\mathbf{A})\,\hat{\mathbf{r}}\,\|}{|\,\omega\,|} \quad \text{minimal for} \quad \omega = \omega^{\mathrm{OR}} = \frac{1}{\widehat{\omega}}\,\frac{\|\,\hat{\mathbf{r}}\,\|}{\|\,\mathbf{A}\,\hat{\mathbf{r}}\,\|}$$



[Sleijpen, van der Vorst '95]

$$q_{k+1}(t) = (1 - \omega_k t)\,q_k(t)$$

$$\hat{\mathbf{r}} = q_k(\mathbf{A})r_{k+1}, \qquad \widehat{\omega} := \frac{(\,\hat{\mathbf{r}}\,, \mathbf{A}\,\hat{\mathbf{r}}\,)}{\|\,\hat{\mathbf{r}}\,\|\,\|\,\mathbf{A}\,\hat{\mathbf{r}}\,\|}$$

## Bi-CGSTAB, II

$$\|\,\mathbf{r}^{\mathrm{OR}}\| = \frac{1}{\widehat{\omega}}\,\|\mathbf{r}^{\mathrm{MR}}\|, \qquad \frac{\|\,\mathbf{r}^{\mathrm{MR}}\|}{|\,\omega^{\mathrm{MR}}\,|} = \frac{1}{\widehat{\omega}}\,\frac{\|\,\mathbf{r}^{\mathrm{OR}}\|}{|\,\omega^{\mathrm{OR}}\,|}$$

Compromise:

take

$$\mathbf{r} = \mathbf{r}_{k+1} = (\mathbf{I} - \omega\mathbf{A})\,\hat{\mathbf{r}}$$

with

$$\omega = \mathrm{sign}(\widehat{\omega})\ \max(0.7, |\,\widehat{\omega}\,|)\ \frac{\|\,\hat{\mathbf{r}}\,\|}{\|\,\mathbf{A}\,\hat{\mathbf{r}}\,\|}$$



[Sleijpen, van der Vorst '95]

## Updating by pols. of dg. $\ell \ge 2$

Consider $k = m\ell$.

$$q_{m\ell+\ell} = p_m q_{m\ell} = p_m \cdot p_{m-1} \cdot \ldots \cdot p_0,$$

where $\gamma_\ell \ne 0$

- $p_m$ of exact degree $\ell$, $p_m(\zeta) = \gamma_\ell\,\zeta^\ell + \ldots$
- $p_m(0) = 1$ and, with $\hat{\mathbf{r}} := q_{m\ell}(\mathbf{A})r_{m\ell+\ell}$,

for minimal residuals: [S., Fokkema '93]

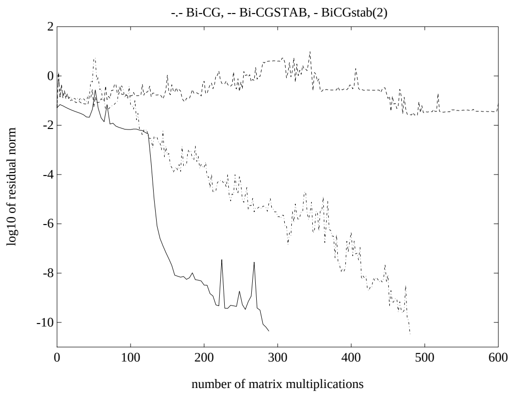
- $p_m$ minimizes $\|\,p_m(\mathbf{A})\,\hat{\mathbf{r}}\,\|$ BiCGstab($\ell$)

for accurate $\rho$'s: [S., van der Vorst '95]

- $p_m$ minimizes $\dfrac{\|\,p_m(\mathbf{A})\,\hat{\mathbf{r}}\,\|}{|\gamma_\ell|}$

Compromise can efficiently be computed enhanced
...






$$u_{xx} + u_{yy} + u_{zz} + 1000 u_x = f.$$

$f$ s.t. $u(x, y, z) = \exp(xyz)\sin(\pi x)\sin(\pi y)\sin(\pi z)$.
$(52 \times 52 \times 52)$ volumes. No preconditioning.

[Gutknecht '94]

## Updating by pols. dg. $\ell \geq 2$

Consider $k = m\ell$.

$$q_{m\ell+\ell} = p_m q_{m\ell} = p_m \cdot p_{m-1} \cdot \ldots \cdot p_0,$$

where $p_m$ of exact degree $\ell$, $p_m(0) = 1$ and

$p_m$ minimizes $\|p_m(\mathbf{A})\,\underbrace{q_{m\ell}(\mathbf{A})\,r_{m\ell+\ell}}_{\widehat{\mathbf{r}}}\|.$

---

**BiCGStab2**, $\ell = 2$.

$q_{m2}$

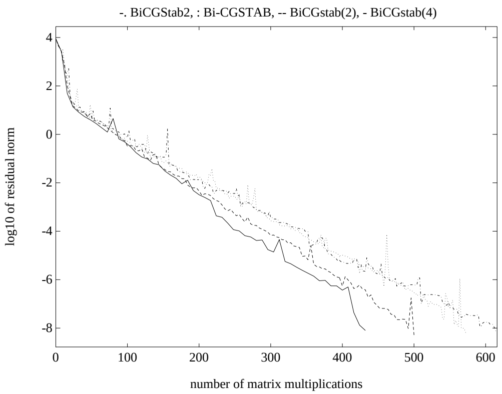
$q_{m2+1} = (1 - \omega_m t) q_{m2}$ min. w.r.t. $\omega_m$

$q_{m2+2} = p_m q_{m2} = (1 - \xi) q_{m2} + (\xi + \eta t) q_{m2+1}$
min. w.r.t. $\xi, \eta$

$p_m = 1 - \omega'_m t - \omega''_m t^2 = (1-\xi) + (\xi + \eta t)(1 - \omega_m t)$

Still breakdown dangers in the odd steps!






$-(a\,u_x)_x - (a\,u_y)_y = 1$ on $[0,1] \times [0,1]$.
$a = 1000$ for $0.1 \leq x, y \leq 0.9$
and $a = 1$ elsewhere.
Dirichlet BC on $y = 0$,
Neumann BC on other parts of Boundary.
$200 \times 200$ volumes. ILU Decomp.



[Sleijpen, Fokkema '93, Sleijpen, van der Vorst, Fokkema '94]

## Updating by pol. dg. $\ell \geq 2$

Consider $k = m\ell$.

$$q_{m\ell+\ell} = p_m q_{m\ell} = p_m \cdot p_{m-1} \cdot \ldots \cdot p_0,$$

where $p_m$ of exact degree $\ell$, $p_m(0) = 1$ and

$p_m$ minimizes $\|p_m(\mathbf{A})\,\underbrace{q_{m\ell}(\mathbf{A})\,r_{m\ell+\ell}}_{\widehat{\mathbf{r}}}\|.$

---

**BiCGstab($\ell$)** $\ell \geq 2$.

With respect to the $\gamma_i$, minimize

$$\|\widehat{\mathbf{r}} - \sum_{j=1}^{\ell} \gamma_j \mathbf{A}^j \widehat{\mathbf{r}}\|, \quad \left(p_m(t) = 1 - \sum_{j=1}^{\ell} \gamma_j t^j\right)$$

Compute $\mathbf{A}\widehat{\mathbf{r}}, \mathbf{A}^2\widehat{\mathbf{r}}, \ldots, \mathbf{A}^\ell\widehat{\mathbf{r}}$ explicitly.

With $\mathbf{V} := \left[\mathbf{A}\widehat{\mathbf{r}} \,|\, \ldots \,|\, \mathbf{A}^\ell\widehat{\mathbf{r}}\right]$, $\vec{\gamma} := (\gamma_1, \ldots, \gamma_\ell)^T$

we have [Normal Equations, use Choleski]

$$(\mathbf{V}^*\mathbf{V})\vec{\gamma} = \mathbf{V}^*\widehat{\mathbf{r}}$$



[Sleijpen, Fokkema '93]

$\widehat{\mathbf{r}} := \mathrm{Q}_{m\ell}\, r_{m\ell+\ell}$ and $\widehat{\mathbf{u}} := \mathrm{Q}_{m\ell}\, u_{m\ell+\ell-1}$

Suppose
$\widehat{\mathbf{r}}, \mathbf{A}\widehat{\mathbf{r}}, \ldots, \mathbf{A}^\ell\widehat{\mathbf{r}}$ and $\widehat{\mathbf{u}}, \mathbf{A}\widehat{\mathbf{u}}, \ldots, \mathbf{A}^\ell\widehat{\mathbf{u}}$
have explicitly been computed.

The $\gamma_i$ minimize

$$\|\widehat{\mathbf{r}} - \sum_{j=1}^{\ell} \gamma_j \mathbf{A}^j \widehat{\mathbf{r}}\|$$

---

## BiCGstab($\ell$) $\ell \geq 2$

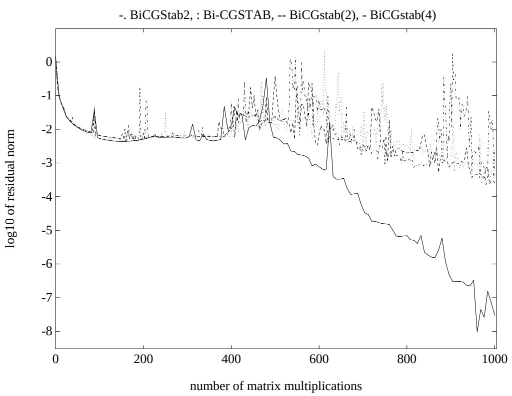
If $p_m(t) := 1 - \sum_{j=1}^{\ell} \gamma_j t^j$, then

$$\mathbf{r}_{m\ell+\ell} = p_m(\mathbf{A})\widehat{\mathbf{r}} = \widehat{\mathbf{r}} - \sum_{j=1}^{\ell} \gamma_j \mathbf{A}^j \widehat{\mathbf{r}}$$

$$\mathbf{x}_{m\ell+\ell} = \widehat{\mathbf{x}} = \widehat{\mathbf{x}} + \sum_{j=1}^{\ell} \gamma_j \mathbf{A}^{j-1} \widehat{\mathbf{r}}$$

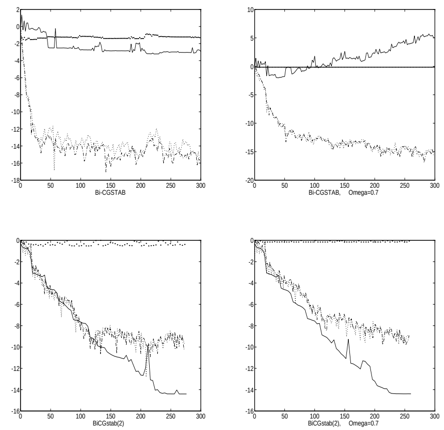
$$\mathbf{u}_{m\ell+\ell-1} = p_m(\mathbf{A})\widehat{\mathbf{u}} = \widehat{\mathbf{u}} - \sum_{j=1}^{\ell} \gamma_j \mathbf{A}^j \widehat{\mathbf{u}}$$





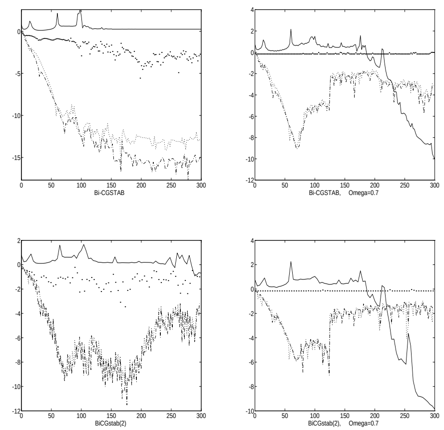

$$-\epsilon(u_{xx} + u_{yy}) + a(x,y)u_x + b(x,y)u_y = 0,$$
on $[0,1] \times [0,1]$, Dirichlet BC $\epsilon = 10^{-1}$
$$a(x,y) = 4x(x-1)(1-2y),$$
$$b(x,y) = 4y(1-y)(1-2x),$$
$u(x,y) = \sin(\pi x) + \sin(13\pi x) + \sin(\pi y) + \sin(13\pi y)$
$(201 \times 201)$ volumes, no preconditioning.






$u_{xx} + u_{yy} + u_{zz} + 1000\,u_x = f$, $(10 \times 10 \times 10)$ vol.






$$-u_{xx} - u_{yy} + \gamma(xu_x + yu_y) + \beta u = f$$

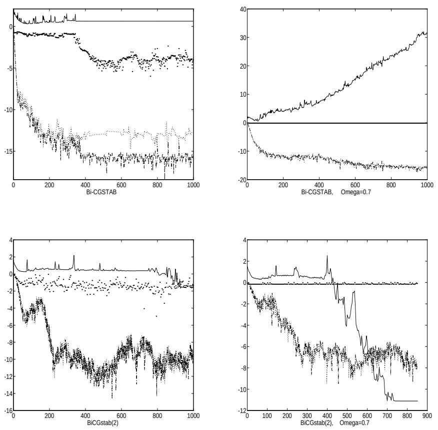
$\gamma = 100$ & $\beta = -200$, $(63 \times 63)$ vol.



$$-u_{xx} - u_{yy} + \gamma(xu_x + yu_y) + \beta u = f$$

on unit-square, Dirichlet BC,
with discretized solution $x = (1, 1, \ldots, 1)^T$,
$\gamma = 1000$ and $\beta = 10$.

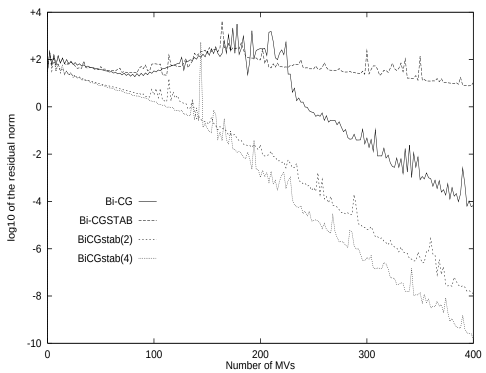
$(66 \times 66)$ volumes. No preconditioning.




$$-u_{xx} - u_{yy} + \gamma(xu_x + yu_y) + \beta u = f$$

$\gamma = 1000$ & $\beta = 10$, $(66 \times 66)$ vol.



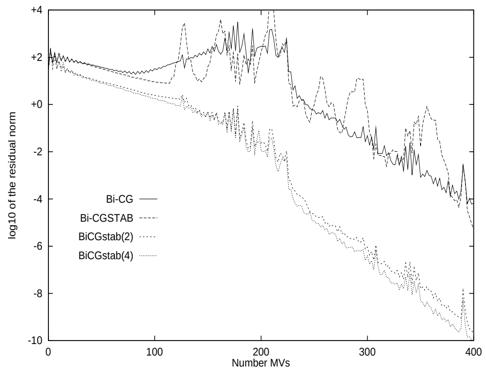



BiCGstab($\ell$)

$$-u_{xx} - u_{yy} + 1000(xu_x + yu_y) + 10u = f$$

on the unit cube, Dirichlet BC.
$(33 \times 33)$ volumes. No preconditioning.
$f$ such that $u(x,y) = \exp(xy)\sin(\pi x)\sin(\pi y)$






BiCGstab($\ell$), *using Bi-CG coefficients*

$$-u_{xx} - u_{yy} + 1000(xu_x + yu_y) + 10u = f$$

on the unit cube, Dirichlet BC
$(33 \times 33)$ volumes. No preconditioning.
$f$ such that $u(x,y) = \exp(xy)\sin(\pi x)\sin(\pi y)$



## Properties BiCGstab($\ell$), $\ell = 2, 4, 8$

**Advantages**

- Hybrid Bi-CG.
- Converges faster than Bi-CGSTAB.
- Stagnation can be avoided
- $1.75 + 0.25 \cdot \ell$ DOT/MV, $2.5 + 0.5 \cdot \ell$ AXPY/MV.
- Storage: $5 + 2 \cdot \ell$ large vectors.

**Drawbacks** Danger of

(A) Lanczos breakdown ($\rho_k = 0$),

(B) pivot breakdown ($\sigma_k = 0$),

*(C) breakdown minimization ($\omega_k = 0$) minimized.*



## Why using pol. fact. of dg. $\geq 2$?

1 sweep BiCGstab($\ell$) versus $\ell$ steps Bi-CGSTAB:

- GMRES($\ell$) will reduce residuals better than $\ell\times$ GMRES(1)
- GMRES($\ell$) contributes only once to an increase of $\hat{\rho}_k$

## Why not?

- Efficiency:
  $1.75 + 0.25 \cdot \ell$ DOT/MV, $2.5 + 0.5 \cdot \ell$ AXPY/MV
  Storage: $2\ell + 5$ large vector.
- Loss of accuracy:

$$\big| \|\mathbf{r}_k\| - \|\mathbf{b} - \mathbf{A}\mathbf{x}_k\| \big| \leq \ldots + c\,\bar{\xi}\,\max\left(|\gamma_i|\,\big\|\,|\mathbf{A}|\,|\mathbf{A}^{i-1}\hat{\mathbf{r}}|\,\big\|\right)$$



## The stability of BiCGstab($\ell$)

**Global stability**

How is the speed of convergence affected by perturbations on, for instance, the Bi-CG coefficients $\alpha_k$, $\beta_k$?

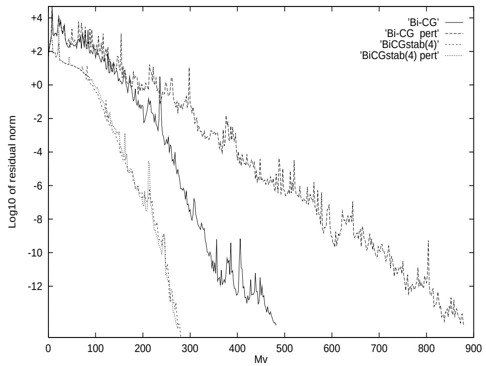
**Local stability**

The scaled normal equations are remarkably well-conditioned:
the polynomial coefficients are accurate.

$$\mathbf{V} = [\,\mathbf{A}\hat{\mathbf{r}} \,|\, \mathbf{A}^2\hat{\mathbf{r}} \,|\, \ldots \,|\, \mathbf{A}^\ell\hat{\mathbf{r}}\,]$$

$\mathbf{V}\Delta^{-1}$ is well-conditioned,
where $\Delta = \mathrm{diag}(\ldots, \|\mathbf{A}^j\hat{\mathbf{r}}\|, \ldots)$.






Bi-CG and BiCGstab(4), perturbed and unpert.

Sherman4



**Do we need polynomials that are 1 in 0 in the intermediate steps?**

For $\rho_k, \sigma_k$: $\mathrm{degr}(q_k) = k$. For $\mathbf{r}_k$: $q_k(0) = 1$.

**Cure**. Use $q'_k$ for $\rho_k, \sigma_k$ & $q''_k$ for $\mathbf{r}$.

(Bi-CG)

For $i = 0, \ldots, \ell - 1$, with $k = m\ell + i$,

$$\left\{\begin{array}{ll}
\rho_k = (\mathbf{A}^i\,\mathbf{Q}_{m\ell}\,r_k, \tilde{\mathbf{r}}_0) & q'_k(t) = t^i\,q_{m\ell}(t) \\
\text{for } j = 0, \ldots, i & \\
\quad \mathbf{A}^j\,\mathbf{Q}_{m\ell}\,u_k = \mathbf{A}^j\,\mathbf{Q}_{m\ell}\,r_k - \beta_k\,\mathbf{A}^j\,\mathbf{Q}_{m\ell}\,u_{k-1}, & \\
\text{Compute} \quad \mathbf{A}\,\mathbf{A}^i\,\mathbf{Q}_{m\ell}\,u_k & \\
& \\
\sigma_k = (\mathbf{A}\,\mathbf{A}^i\,\mathbf{Q}_{m\ell}\,u_k, \tilde{\mathbf{r}}_0) & \\
\text{for } j = 0, \ldots, i & \\
\quad \mathbf{A}^j\,\mathbf{Q}_{m\ell}\,r_{k+1} = \mathbf{A}^j\,\mathbf{Q}_{m\ell}\,r_k - \alpha_k\,\mathbf{A}^j\,\mathbf{A}\mathbf{Q}_{m\ell}\,u_k, & \\
\text{Compute} \quad \mathbf{A}\,\mathbf{A}^i\,\mathbf{Q}_{m\ell}\,r_{k+1} & \\
& \\
\hat{\mathbf{x}} = \hat{\mathbf{x}} + \alpha_k\,\mathbf{Q}_{m\ell}\,u_k & q''_k(t) = q_{m\ell}(t)
\end{array}\right.$$

$k = m2$

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Bi-CG | | $\mathbf{u}_{k-1}$, $Q_k u_{k-1}$ | $\mathbf{r}_k$, $\underline{Q_k r_k}$ | | | | |
| | $\rho_k$ | $Q_k u_k$ | $Q_k r_k$ | $\boxed{AQ_k u_k}$ | | | |
| | $\sigma_k$ | $Q_k u_k$ | $Q_k r_{k+1}$ | $AQ_k u_k$ | $\boxed{AQ_k r_{k+1}}$ | | |
| | $\rho_{k+1}$ | $Q_k u_{k+1}$ | $Q_k r_{k+1}$ | $AQ_k u_{k+1}$ | $AQ_k r_{k+1}$ | $\boxed{A^2Q_k u_{k+1}}$ | |
| | $\sigma_{k+1}$ | $Q_k u_{k+1}$ | $Q_k r_{k+2}$ | $AQ_k u_{k+1}$ | $AQ_k r_{k+2}$ | $A^2Q_k u_{k+1}$ | $\boxed{A^2Q_k r_{k+2}}$ |
| Pol | | $Q_k u_{k+1}$ | $Q_k r_{k+2}$ | $AQ_k u_{k+1}$ | $AQ_k r_{k+2}$ | $A^2Q_k u_{k+1}$ | $A^2Q_k r_{k+2}$ |
| | $\gamma_1, \gamma_2$ | $Q_{k+2} u_{k+1}$, $\mathbf{u}_{k+1}$ | $Q_{k+2} r_{k+2}$, $\mathbf{r}_{k+2}$ | | | | |

## The computational scheme for BiCGstab(2): Power basis variant



```
problem
% Accuracy (residual reduction), choice "ell".
epsilon=10^(-16); ell=4

% Choice initial approximation x
% and shadow residual rt
x=zeros(b); rt=rand(b);

% Initiation scalars and search direction
sigma=1; theta=1; u=zeros(b);
% Computation of initial residual
y=x; MV; y=b-y; r=y;

norm=r'*r;  epsilon=norm*epsilon^2; L=2:ell+1;

while norm > epsilon

   sigma=-theta*sigma; y=r;
   for j=1:ell
        rho=y'*rt;   beta=rho/sigma;
        u=r-beta*u;
        y=u(:,j); MV; u(:,j+1)=y;
        sigma=y'*rt; alpha=rho/sigma;
        x=x+alpha*u(:,1);
        r=r-alpha*u(:,2:j+1);
        y=r(:,j); MV; r(:,j+1)=y;
   end

   G=r'*r; gamma=G(L,L)\G(L,1); theta=gamma(ell);
   x=x+r*[gamma;0]; r=r*[1;-gamma]; u=u*[1;-gamma];
   norm=r'*r;

end
```

[Sleijpen, Fokkema '93]

## BiCGstab($\ell$) Power basis variant

Choose $\mathbf{x}_0$ and $\widetilde{\mathbf{r}}_0$.
$\mathbf{x} = \mathbf{x}_0$, $\mathbf{r} = [\mathbf{b} - \mathbf{A}\mathbf{x}]$.
$\mathbf{u} = [0]$, $\gamma_\ell = \sigma = 1$.
While $\|\mathbf{r}\| > tol$ do
  $\sigma \leftarrow -\gamma_\ell \sigma$
  for $j = 1$ to $\ell$ do
    $\rho \leftarrow (\mathbf{r}_j, \widetilde{\mathbf{r}}_0)$, $\beta \leftarrow \rho/\sigma$
    $\mathbf{u} \leftarrow \mathbf{r} - \beta\,\mathbf{u}$, $\mathbf{u} \leftarrow [\mathbf{u}, \mathbf{A}\mathbf{u}_j]$
    $\sigma \leftarrow (\mathbf{u}_{j+1}, \widetilde{\mathbf{r}}_0)$, $\alpha \leftarrow \rho/\sigma$
    $\mathbf{x} \leftarrow \mathbf{x} + \alpha\,\mathbf{u}_1$
    $\mathbf{r} \leftarrow \mathbf{r} - \alpha\,[\mathbf{u}_2, \ldots, \mathbf{u}_{j+1}]$, $\mathbf{r} \leftarrow [\mathbf{r}, \mathbf{A}\mathbf{r}_j]$
  end do
  $\mathbf{R} := [\mathbf{r}_2, \ldots, \mathbf{r}_{\ell+1}]$
  solve $\vec{\gamma}$ from $(\mathbf{R}^T\mathbf{R})\vec{\gamma} = \mathbf{R}^T\mathbf{r}_1$
  $\mathbf{u} \leftarrow [\mathbf{u}_1 - (\gamma_1\,\mathbf{u}_2 + \ldots + \gamma_\ell\,\mathbf{u}_{\ell+1})]$
  $\mathbf{x} \leftarrow \mathbf{x} + (\gamma_1\,\mathbf{r}_1 + \ldots + \gamma_\ell\,\mathbf{r}_\ell)$
  $\mathbf{r} \leftarrow [\mathbf{r}_1 - (\gamma_1\,\mathbf{r}_2 + \ldots + \gamma_\ell\,\mathbf{r}_{\ell+1})]$
end while

$\mathbf{r}$ and $\mathbf{u}$ are matrices with $n$ rows, # columns varies:
$\mathbf{r} := [\mathbf{r}_1, \ldots, \mathbf{r}_j]$, $\mathbf{u} := [\mathbf{u}_1, \ldots, \mathbf{u}_j]$. $\vec{\gamma} := (\gamma_1, \ldots, \gamma_\ell)^T$



## Conclusions

Hybrid Bi-CG meths are efficient linear solvers.

The auxiliary polynomial can be used

- to gain efficiency and
- to improve residual reduction

*but also for improving other properties*

- to obtain more accurate Bi-CG coefficients
- to improve inexact Newton schemes

Effects of erratic convergence can be reduced:

- Inaccurate results due to large intermediate residuals can inexpensively be avoided.
- Large residual peaks may deteriorate convergence, but can be avoided (BiCGstab($\ell$), GCGS).