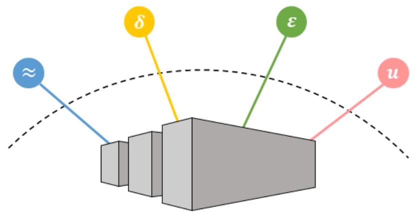


Forward and backward error bounds for a mixed precision preconditioned conjugate gradient algorithm

Thomas Bake[1], Erin Carson[1], Yuxin Ma[1].

HPCSE 2026 - Hotel Solaň

May 19th, 2026

[1]Charles University

Contributions

1. General framework for preconditioned conjugate gradient (PCG).

Contributions

1. General framework for preconditioned conjugate gradient (PCG).

2. Forward and backward error bounds for PCG.

Contributions

1. General framework for preconditioned conjugate gradient (PCG).
2. Forward and backward error bounds for PCG.
3. Extension of bounds to incorporate mixed precision.

The conjugate gradient method

CG method

- $Ax = b$, $A \in \mathbb{R}^{n\times n}$ symmetric positive definite (SPD), large and sparse, $b \in \mathbb{R}^n$.

Algorithm 1: The CG algorithm

$r_0 = b - Ax_0, \quad p_0 = r_0$

for $k = 0, 1, \dots, n_{\max}$

  $\alpha_k = \frac{r_k^\top r_k}{p_k^\top A p_k}$

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

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

  Stop when stopping criterion is satisfied

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

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

end for

CG method

- $Ax = b, \quad A \in \mathbb{R}^{n\times n}$ symmetric positive definite (SPD), large and sparse, $b \in \mathbb{R}^n$.

- Mathematically, the *residual* vectors

$$r_0, r_1, \dots, r_k$$

and the *direction* vectors

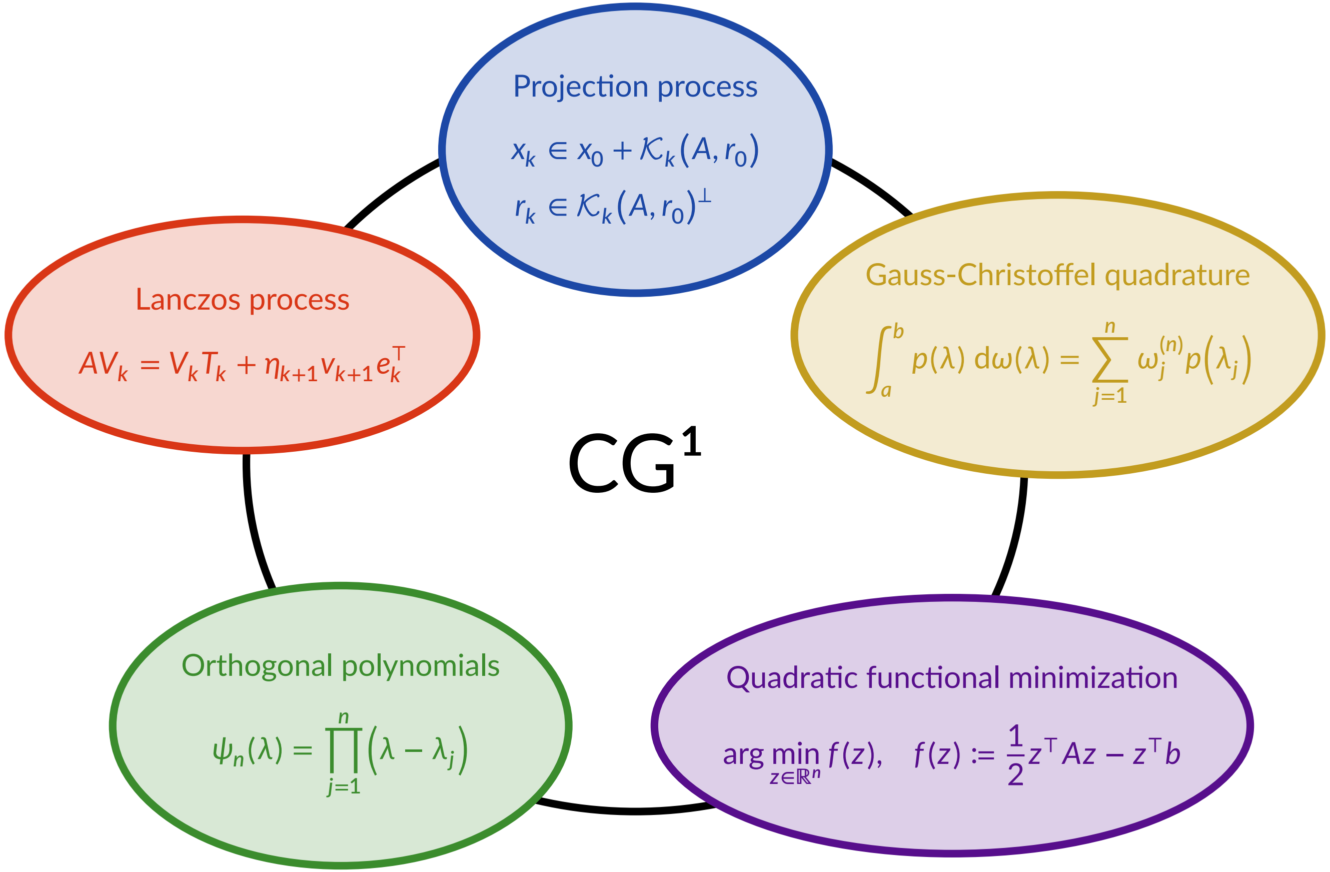
$$p_0, p_1, \dots, p_k$$

form orthogonal bases of $\mathcal{K}_k(A, r_0)$ with respect to the Euclidean and A-inner products, respectively.

Algorithm 1: The CG algorithm

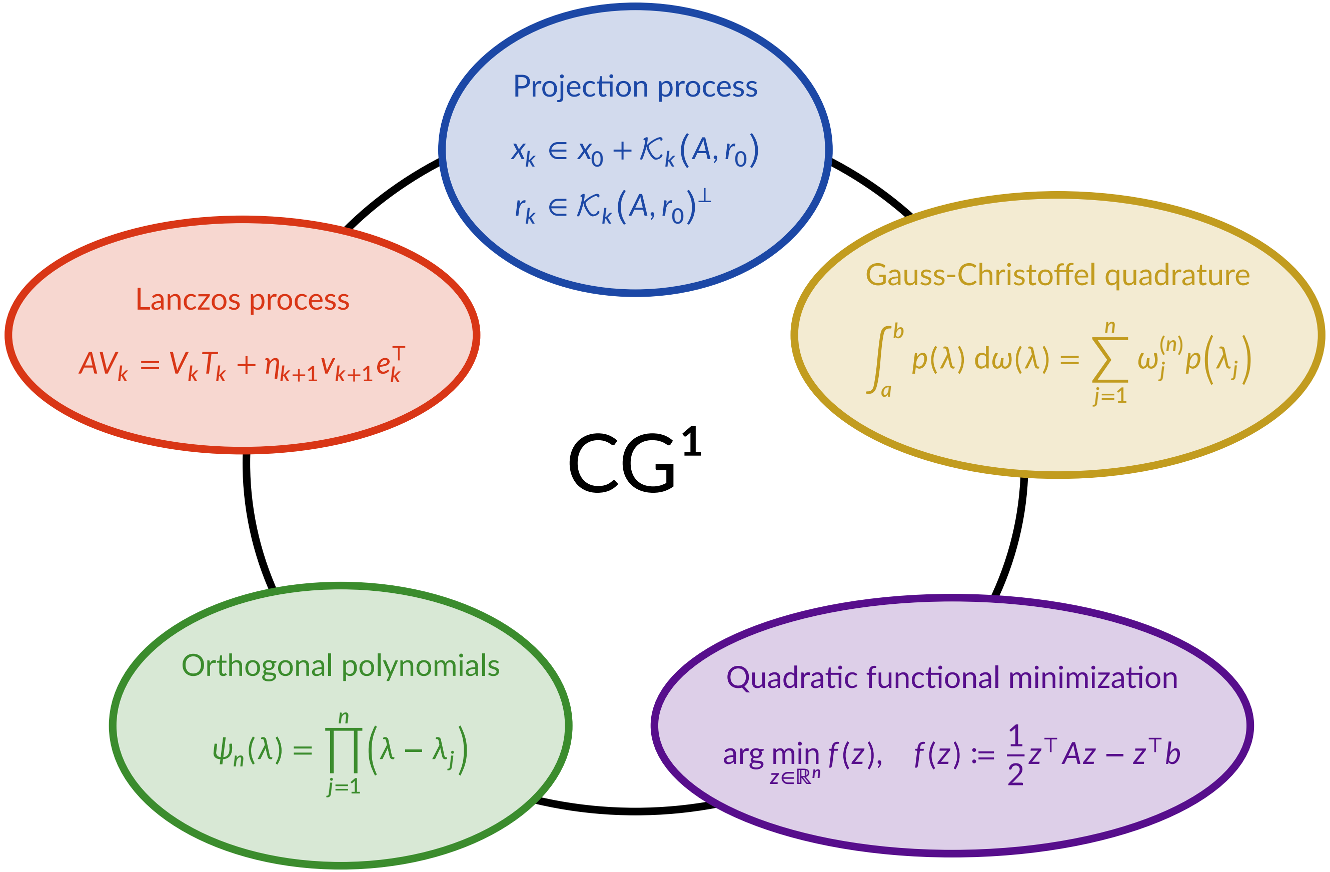
$r_0 = b - Ax_0, \quad p_0 = r_0$
for $k = 0, 1, \dots, n_{\max}$
  $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$
  $x_{k+1} = x_k + \alpha_k p_k$
  $r_{k+1} = r_k - \alpha_k A p_k$
  Stop when stopping criterion is satisfied
  $\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$
  $p_{k+1} = r_{k+1} + \beta_{k+1} p_k$
end for

CG method



[1]Idea of graphic taken from Petr Tichý.

CG method



[1]Idea of graphic taken from Petr Tichý.

Preconditioned conjugate gradient

Preconditioning and PCG

Algorithm 1: The CG algorithm

$r_0 = b - Ax_0, \quad p_0 = r_0$

for $k = 0, 1, \dots, n_{\max}$

  $\alpha_k = \frac{r_k^\top r_k}{p_k^\top A p_k}$

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

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

  Stop when stopping criterion is satisfied

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

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

end for

$$\xrightarrow[M = M_L \cdot M_R \in \mathbb{R}^{n\times n}]{\text{SPD preconditioner}}$$

Preconditioning and PCG

Algorithm 1: The CG algorithm

$r_0 = b - Ax_0, \quad p_0 = r_0$

for $k = 0, 1, \ldots, n_{\max}$

$\quad \alpha_k = \frac{r_k^\top r_k}{p_k^\top A p_k}$

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

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

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}$

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

end for

SPD preconditioner

$\longrightarrow$

$M = M_L \cdot M_R \in \mathbb{R}^{n \times n}$

Algorithm 2: The PCG algorithm

$r_0 = b - Ax_0, \quad s_0 = M_L^{-1} r_0, \quad q_0 = M_R^{-1} s_0,$
$z_0 = M_R^{-\top} r_0, \quad p_0 = q_0$

for $k = 0, 1, \ldots, n_{\max}$

$\quad \alpha_k = \frac{z_k^\top s_k}{p_k^\top A p_k}$

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

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

$\quad s_{k+1} = M_L^{-1} r_{k+1}, \quad q_{k+1} = M_R^{-1} s_{k+1}, \quad z_{k+1} = M_R^{-\top} r_{k+1}$

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{z_{k+1}^\top s_{k+1}}{z_k^\top s_k}$

$\quad p_{k+1} = q_{k+1} + \beta_{k+1} p_k$

end for

Preconditioning and PCG

Algorithm 1: The CG algorithm

$r_0 = b - Ax_0, \quad p_0 = r_0$

for $k = 0, 1, \ldots, n_{\max}$

$\quad \alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$

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

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

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

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

end for

$$\xrightarrow[M = M_L \cdot M_R \in \mathbb{R}^{n\times n}]{\text{SPD preconditioner}}$$

Algorithm 2: The PCG algorithm

$r_0 = b - Ax_0, \quad s_0 = M_L^{-1} r_0, \quad q_0 = M_R^{-1} s_0,$
$z_0 = M_R^{-T} r_0, \quad p_0 = q_0$

for $k = 0, 1, \ldots, n_{\max}$

$\quad \alpha_k = \frac{z_k^T s_k}{p_k^T A p_k}$

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

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

$\quad s_{k+1} = M_L^{-1} r_{k+1}, \quad q_{k+1} = M_R^{-1} s_{k+1}, \quad z_{k+1} = M_R^{-T} r_{k+1}$

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{z_{k+1}^T s_{k+1}}{z_k^T s_k}$

$\quad p_{k+1} = q_{k+1} + \beta_{k+1} p_k$

end for

$$\boxed{M_L^{-1} A M_R^{-1} y = M_L^{-1} b, \quad x = M_R^{-1} y}$$

- $M_L^{-1} A M_R^{-1}$ must be SPD with respect to an inner product.

Preconditioning and PCG

Algorithm 1: The CG algorithm

$r_0 = b - Ax_0, \quad p_0 = r_0$

for $k = 0, 1, \ldots, n_{\max}$

$\quad \alpha_k = \frac{r_k^\top r_k}{p_k^\top A p_k}$

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

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

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}$

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

end for

$$\xrightarrow{\text{SPD preconditioner}}$$

$$M = M_L \cdot M_R \in \mathbb{R}^{n \times n}$$

Algorithm 2: The PCG algorithm

$r_0 = b - Ax_0, \quad s_0 = M_L^{-1} r_0, \quad q_0 = M_R^{-1} s_0,$
$z_0 = M_R^{-\top} r_0, \quad p_0 = q_0$

for $k = 0, 1, \ldots, n_{\max}$

$\quad \alpha_k = \frac{z_k^\top s_k}{p_k^\top A p_k}$

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

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

$\quad s_{k+1} = M_L^{-1} r_{k+1}, \quad q_{k+1} = M_R^{-1} s_{k+1}, \quad z_{k+1} = M_R^{-\top} r_{k+1}$

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{z_{k+1}^\top s_{k+1}}{z_k^\top s_k}$

$\quad p_{k+1} = q_{k+1} + \beta_{k+1} p_k$

end for

$$M_L^{-1} A M_R^{-1} y = M_L^{-1} b, \quad x = M_R^{-1} y$$

- $M_L^{-1} A M_R^{-1}$ must be SPD with respect to an inner product.
- Example: Let $M = LL^\top$ be the Cholesky factorization of M.

Scheme		
Left	$M_L = LL^\top$	$M_R = I_n$
Right	$M_L = I_n$	$M_R = LL^\top$
Split	$M_L = L$	$M_R = L^\top$

Preconditioning and PCG

Algorithm 1: The CG algorithm

$r_0 = b - Ax_0, \quad p_0 = r_0$

for $k = 0, 1, \dots, n_{\max}$

$\quad \alpha_k = \frac{r_k^\top r_k}{p_k^\top A p_k}$

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

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

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}$

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

end for

$$\xrightarrow{\text{SPD preconditioner}}$$

$$M = M_L \cdot M_R \in \mathbb{R}^{n \times n}$$

Algorithm 2: The PCG algorithm

$r_0 = b - Ax_0, \quad s_0 = M_L^{-1} r_0, \quad q_0 = M_R^{-1} s_0,$
$z_0 = M_R^{-\top} r_0, \quad p_0 = q_0$

for $k = 0, 1, \dots, n_{\max}$

$\quad \alpha_k = \frac{z_k^\top s_k}{p_k^\top A p_k}$

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

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

$\quad s_{k+1} = M_L^{-1} r_{k+1}, \quad q_{k+1} = M_R^{-1} s_{k+1}, \quad z_{k+1} = M_R^{-\top} r_{k+1}$

$\quad$ Stop when stopping criterion is satisfied

$\quad \beta_{k+1} = \frac{z_{k+1}^\top s_{k+1}}{z_k^\top s_k}$

$\quad p_{k+1} = q_{k+1} + \beta_{k+1} p_k$

end for

$$\boxed{M_L^{-1} A M_R^{-1} y = M_L^{-1} b, \quad x = M_R^{-1} y}$$

- $M_L^{-1} A M_R^{-1}$ must be SPD with respect to an inner product.
- Example: Let $M = LL^\top$ be the Cholesky factorization of M.

Scheme

Left	$M_L = LL^\top$	$M_R = I_n$
Right	$M_L = I_n$	$M_R = LL^\top$
Split	$M_L = L$	$M_R = L^\top$

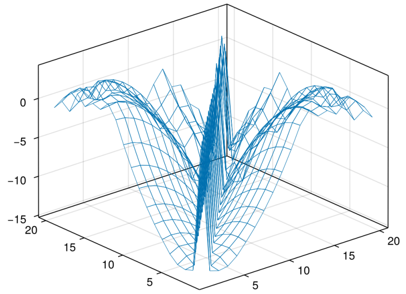
What can we say about this algorithm in *finite precision*?

Forward and backward error bounds

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
x \arrow[r, "f"] \arrow[d, "\|\Delta x\|_{\mathcal{X}}"'] & y = f(x) \arrow[d, "\|y - \hat{y}\|_{\mathcal{Y}}"] \\
x + \Delta x \arrow[r, "f"'] & \hat{y} = f(x + \Delta x)
\end{tikzcd}
$$

Previous work on finite precision Lanczos / CG

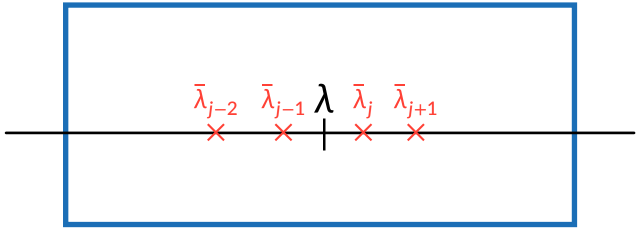
(Paige 1976, 1980)



(Greenbaum 1997)

$$\|b - A\hat{x}_k\| \le \|b - A\hat{x}_k - \hat{r}_k\| + \|\hat{r}_k\| \xrightarrow{\|\hat{r}_k\| \to 0} \|b - A\hat{x}_k - \hat{r}_k\|$$

$$\|b - A\hat{x}_k - \hat{r}_k\| \le O(nk\boldsymbol{u})\,\|A\| \max_{j\le k}\left(\|\hat{x}_j\|, \|x\|\right)$$



(Greenbaum 1989)

$\|\hat{r}_k\| \le ?$

Preliminaries

- Let the symbol $\hat{\cdot}$ denote quantities computed in finite precision. Assume $n\boldsymbol{u} < 1$, where $\boldsymbol{u}$ denotes the unit roundoff.

Preliminaries

- Let the symbol $\hat{\cdot}$ denote quantities computed in finite precision. Assume $n\boldsymbol{u} < 1$, where $\boldsymbol{u}$ denotes the unit roundoff.

$$x_{k+1} = x_k + \alpha_k p_k \dashrightarrow \begin{aligned} \hat{x}_{k+1} = \hat{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta\bar{x}_{k+1} + \Delta x_{k+1}, \quad &\|\Delta\bar{x}_{k+1}\| \le \boldsymbol{u}\|\hat{x}_k\|, \\ &\|\Delta x_{k+1}\| \le O(\boldsymbol{u})|\hat{\alpha}_k|\|\hat{p}_k\|. \end{aligned}$$

Preliminaries

- Let the symbol $\hat{\cdot}$ denote quantities computed in finite precision. Assume $n\boldsymbol{u} < 1$, where $\boldsymbol{u}$ denotes the unit roundoff.

$$x_{k+1} = x_k + \alpha_k p_k \dashrightarrow \hat{x}_{k+1} = \hat{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta\bar{x}_{k+1} + \Delta x_{k+1}, \quad \begin{aligned} &\|\Delta\bar{x}_{k+1}\| \le \boldsymbol{u}\|\hat{x}_k\|, \\ &\|\Delta x_{k+1}\| \le O(\boldsymbol{u})|\hat{\alpha}_k|\|\hat{p}_k\|. \end{aligned}$$

- Define the recurrence $\bar{x}_0 = x_0, \quad \bar{x}_{k+1} = \bar{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta x_{k+1}.$

Preliminaries

- Let the symbol $\hat{\cdot}$ denote quantities computed in finite precision. Assume $n\boldsymbol{u} < 1$, where $\boldsymbol{u}$ denotes the unit roundoff.

$$x_{k+1} = x_k + \alpha_k p_k \dashrightarrow \hat{x}_{k+1} = \hat{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta\bar{x}_{k+1} + \Delta x_{k+1}, \quad \begin{aligned} &\|\Delta\bar{x}_{k+1}\| \le \boldsymbol{u}\|\hat{x}_k\|, \\ &\|\Delta x_{k+1}\| \le O(\boldsymbol{u})|\hat{\alpha}_k|\|\hat{p}_k\|. \end{aligned}$$

- Define the recurrence $\bar{x}_0 = x_0, \quad \bar{x}_{k+1} = \bar{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta x_{k+1}.$

- Recall that CG minimizes $f(z) = \frac{1}{2} z^\top A z - z^\top b$. Then

Preliminaries

- Let the symbol $\hat{\cdot}$ denote quantities computed in finite precision. Assume $n\boldsymbol{u} < 1$, where $\boldsymbol{u}$ denotes the unit roundoff.

$$x_{k+1} = x_k + \alpha_k p_k \quad \dashrightarrow \quad \hat{x}_{k+1} = \hat{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta\bar{x}_{k+1} + \Delta x_{k+1}, \quad \begin{array}{l} \|\Delta\bar{x}_{k+1}\| \le \boldsymbol{u}\|\hat{x}_k\|, \\ \|\Delta x_{k+1}\| \le O(\boldsymbol{u})|\hat{\alpha}_k|\|\hat{p}_k\|. \end{array}$$

- Define the recurrence $\bar{x}_0 = x_0, \quad \bar{x}_{k+1} = \bar{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta x_{k+1}$.

- Recall that CG minimizes $f(z) = \frac{1}{2} z^\top A z - z^\top b$. Then

$$f(\bar{x}_k) - f(\bar{x}_{k+1}) = -\frac{1}{2}\hat{\alpha}_k(1 + \delta\alpha_k)\|\hat{r}_k\|^2_{M^{-1}} + \delta f_k,$$

where $\delta\alpha_k$ and δf_k are bounded error terms.

Preliminaries

- Let the symbol $\hat{\cdot}$ denote quantities computed in finite precision. Assume $n\boldsymbol{u} < 1$, where $\boldsymbol{u}$ denotes the unit roundoff.

$$x_{k+1} = x_k + \alpha_k p_k \quad \dashrightarrow \quad \begin{aligned} \hat{x}_{k+1} = \hat{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta\bar{x}_{k+1} + \Delta x_{k+1}, \quad & \|\Delta\bar{x}_{k+1}\| \le \boldsymbol{u}\|\hat{x}_k\|, \\ & \|\Delta x_{k+1}\| \le O(\boldsymbol{u})|\hat{\alpha}_k|\|\hat{p}_k\|. \end{aligned}$$

- Define the recurrence $\bar{x}_0 = x_0, \quad \bar{x}_{k+1} = \bar{x}_k + \hat{\alpha}_k \hat{p}_k + \Delta x_{k+1}.$

- Recall that CG minimizes $f(z) = \frac{1}{2} z^\top A z - z^\top b$. Then

$$f(\bar{x}_k) - f(\bar{x}_{k+1}) = -\frac{1}{2}\hat{\alpha}_k(1 + \delta\alpha_k)\|\hat{r}_k\|^2_{M^{-1}} + \delta f_k,$$

where $\delta\alpha_k$ and δf_k are bounded error terms.

Key result:

$$O(n\boldsymbol{u})\kappa(A) + 2\epsilon_{\text{pre}}^{(s,z)} \le 1/2 \quad \Longrightarrow \quad \exists k^\star \ge 1 : f(\bar{x}_{k^\star}) - f(\bar{x}_{k^\star+1}) \le O\Big[n^2(k^\star)^2\boldsymbol{u}^2\Big]\hat{\alpha}_{k^\star}\|A\|^2\|M^{-1}\|\|x\|^2.$$

Main result[1]

Furthermore, if $k^\star$ satisfies

$$\boxed{\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1-O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\left[\epsilon_{\text{pre}}^{(s,q)} + \epsilon_{\text{pre}}^{(s,z)}\right] + O[(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\left[1 + k^\star\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}}\right] \le \frac{1}{2},}$$

then there exists an iteration step $i \le k^\star$ such that

[1](B., Carson, and Ma 2025, Theorem 2)

Main result[1]

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1-O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\Big[\epsilon_{\text{pre}}^{(s,q)} + \epsilon_{\text{pre}}^{(s,z)}\Big] + O[(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\Big[1 + k^\star\kappa\Big(M_L^{-1}AM_R^{-1}\Big)^{\frac{1}{2}}\Big] \le \frac{1}{2},$$

then there exists an iteration step $i \le k^\star$ such that

$$\frac{\|\hat{r}_i\|}{\|A\|\|x\|} \le O\Big[n(k^\star)^2\boldsymbol{u}\Big]\kappa(M)^{\frac{1}{2}} \max_{j\le k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{\|\hat{x}_i\|}{\|x\|}\cdot\left[\frac{\|b-A\hat{x}_i\|}{\|A\|\|\hat{x}_i\|+\|b\|}\right] \le \frac{\|b-A\hat{x}_i\|}{\|A\|\|x\|} \le O\Big[n(k^\star)^2\boldsymbol{u}\Big]\kappa(M)^{\frac{1}{2}} \max_{j\le k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{1}{\kappa(A)^{\frac{1}{2}}}\cdot\left[\frac{\|x-\hat{x}_i\|_A}{\|x\|_A}\right] \le \frac{\|x-\hat{x}_i\|_A}{\|A\|^{\frac{1}{2}}\|x\|} \le O\Big[n(k^\star)^2\boldsymbol{u}\Big]\kappa(M)^{\frac{1}{2}}\kappa(A)^{\frac{1}{2}} \max_{j\le k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right).$$

[1](B., Carson, and Ma 2025, Theorem 2)

Main result[1]

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1-O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\left[\epsilon_{\text{pre}}^{(s,q)} + \epsilon_{\text{pre}}^{(s,z)}\right] + O[(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\left[1 + k^\star\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}}\right] \leq \frac{1}{2},$$

then there exists an iteration step $i \leq k^\star$ such that

$$\frac{\|\hat{r}_i\|}{\|A\|\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{\|\hat{x}_i\|}{\|x\|} \cdot \left[\frac{\|b-A\hat{x}_i\|}{\|A\|\|\hat{x}_i\| + \|b\|}\right] \leq \frac{\|b-A\hat{x}_i\|}{\|A\|\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{1}{\kappa(A)^{\frac{1}{2}}} \cdot \left[\frac{\|x-\hat{x}_i\|_A}{\|x\|_A}\right] \leq \frac{\|x-\hat{x}_i\|_A}{\|A\|^{\frac{1}{2}}\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}}\kappa(A)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right).$$

Remarks:

[1](B., Carson, and Ma 2025, Theorem 2)

Main result[1]

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1-O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\left[\epsilon_{\text{pre}}^{(s,q)} + \epsilon_{\text{pre}}^{(s,z)}\right] + O[(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\left[1 + k^\star\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}}\right] \leq \frac{1}{2},$$

then there exists an iteration step $i \leq k^\star$ such that

$$\frac{\|\hat{r}_i\|}{\|A\|\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{\|\hat{x}_i\|}{\|x\|} \cdot \left[\frac{\|b - A\hat{x}_i\|}{\|A\|\|\hat{x}_i\| + \|b\|}\right] \leq \frac{\|b - A\hat{x}_i\|}{\|A\|\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{1}{\kappa(A)^{\frac{1}{2}}} \cdot \left[\frac{\|x - \hat{x}_i\|_A}{\|x\|_A}\right] \leq \frac{\|x - \hat{x}_i\|_A}{\|A\|^{\frac{1}{2}}\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}}\kappa(A)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right).$$

Remarks:

- Influence of problem size n and iteration steps k are usually *overestimates*.

[1](B., Carson, and Ma 2025, Theorem 2)

Main result[1]

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1-O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\left[\epsilon_{\text{pre}}^{(s,q)} + \epsilon_{\text{pre}}^{(s,z)}\right] + O[(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\left[1 + k^\star\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}}\right] \le \frac{1}{2},$$

then there exists an iteration step $i \le k^\star$ such that

$$\frac{\|\hat{r}_i\|}{\|A\|\|x\|} \le O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\le k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{\|\hat{x}_i\|}{\|x\|}\cdot\left[\frac{\|b-A\hat{x}_i\|}{\|A\|\|\hat{x}_i\|+\|b\|}\right] \le \frac{\|b-A\hat{x}_i\|}{\|A\|\|x\|} \le O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\le k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right),$$

$$\frac{1}{\kappa(A)^{\frac{1}{2}}}\cdot\left[\frac{\|x-\hat{x}_i\|_A}{\|x\|_A}\right] \le \frac{\|x-\hat{x}_i\|_A}{\|A\|^{\frac{1}{2}}\|x\|} \le O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}}\kappa(A)^{\frac{1}{2}} \max_{j\le k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right).$$

Remarks:

- Influence of problem size n and iteration steps k are usually *overestimates*.
- The iteration step $k^\star$ is *unknown*.

[1](B., Carson, and Ma 2025, Theorem 2)

Main result[1]

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1-O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\left[\epsilon_{\text{pre}}^{(s,q)} + \epsilon_{\text{pre}}^{(s,z)}\right] + O[(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\left[1 + k^\star\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}}\right] \leq \frac{1}{2},$$

then there exists an iteration step $i \leq k^\star$ such that

$$\frac{\|\hat{r}_i\|}{\|A\|\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right), \tag{1}$$

$$\frac{\|\hat{x}_i\|}{\|x\|} \cdot \left[\frac{\|b - A\hat{x}_i\|}{\|A\|\|\hat{x}_i\| + \|b\|}\right] \leq \frac{\|b - A\hat{x}_i\|}{\|A\|\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right), \tag{2}$$

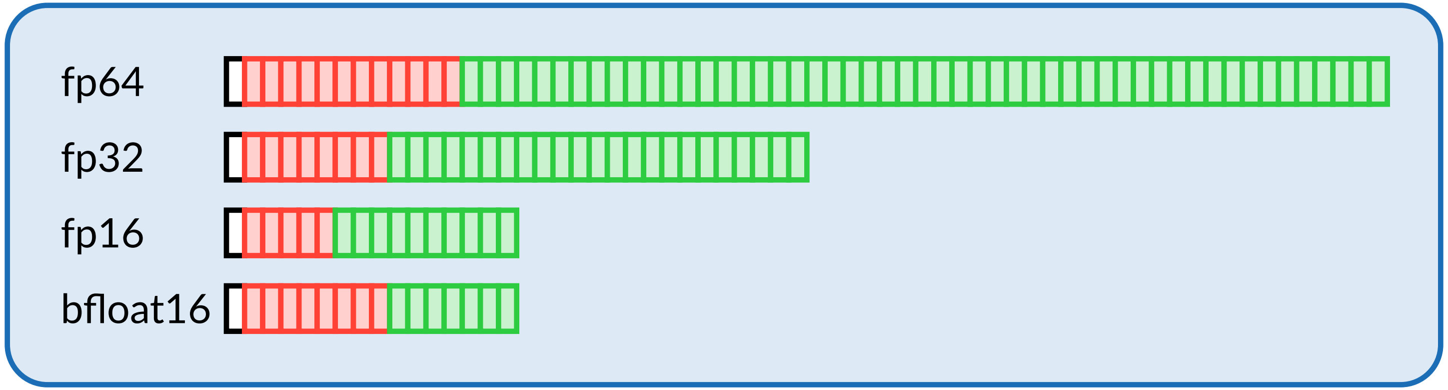
$$\frac{1}{\kappa(A)^{\frac{1}{2}}} \cdot \left[\frac{\|x - \hat{x}_i\|_A}{\|x\|_A}\right] \leq \frac{\|x - \hat{x}_i\|_A}{\|A\|^{\frac{1}{2}}\|x\|} \leq O\left[n(k^\star)^2\boldsymbol{u}\right]\kappa(M)^{\frac{1}{2}}\kappa(A)^{\frac{1}{2}} \max_{j\leq k^\star+1}\left(\frac{\|\hat{x}_j\|}{\|x\|}, 1\right). \tag{3}$$

Remarks:

- Influence of problem size n and iteration steps k are usually *overestimates*.
- The iteration step $k^\star$ is *unknown*.
- Forward and backward errors are small when $f(\bar{x}_k) - f(\bar{x}_{k+1})$ is sufficiently small for some $k \geq 1$.

[1](B., Carson, and Ma 2025, Theorem 2)

Extension to mixed precision



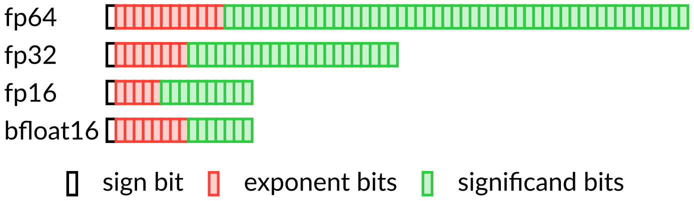
Extension to mixed precision

Main idea: Strategic combination of different computer arithmetics to achieve high accuracy at lower cost.

[1](Carson and Mary 2025)

Extension to mixed precision

Main idea: Strategic combination of different computer arithmetics to achieve high accuracy at lower cost.



Real world phenomena

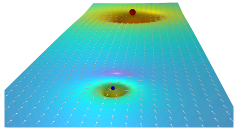


Figure 1: Source: https://www.comsol.com

$\downarrow \varepsilon_m$

Mathematical model — $\nabla^2 \cdot \boldsymbol{g} = -4\pi G\rho$

$\downarrow \varepsilon_d$

Discretization — $\mathcal{V} \to \mathcal{V}_n \subset \mathcal{V}$

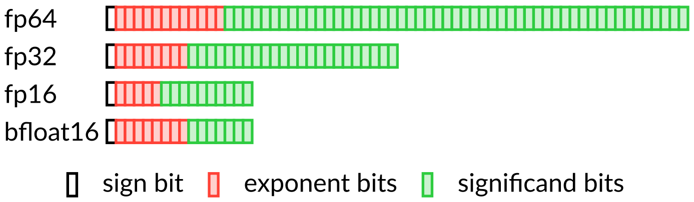
$\downarrow \varepsilon_a$

Computation — $Ax = b \to A\hat{x} = b$

[1](Carson and Mary 2025)

Extension to mixed precision

Main idea: Strategic combination of different computer arithmetics to achieve high accuracy at lower cost.



Approach: Perform rounding error analysis and identify precision requirements for different parts of computation.

Real world phenomena

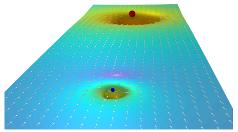


Figure 1: Source: https://www.comsol.com

$\downarrow \varepsilon_m$

Mathematical model — $\nabla^2 \cdot \boldsymbol{g} = -4\pi G\rho$

$\downarrow \varepsilon_d$

Discretization — $\mathcal{V} \rightarrow \mathcal{V}_n \subset \mathcal{V}$

$\downarrow \varepsilon_a$

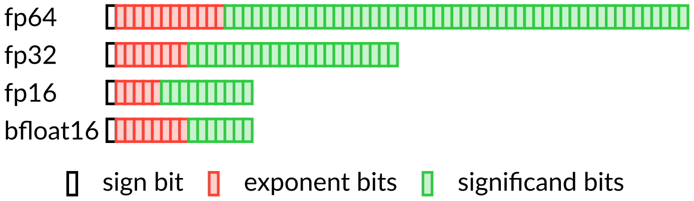
Computation — $Ax = b \rightarrow A\hat{x} = b$

---

[1](Carson and Mary 2025)

Extension to mixed precision

Main idea: Strategic combination of different computer arithmetics to achieve high accuracy at lower cost.



Approach: Perform rounding error analysis and identify precision requirements for different parts of computation.

- **Low precision:** Reduce computation and communication costs.

Real world phenomena

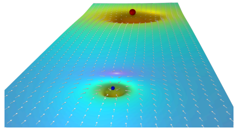


Figure 1: Source: https://www.comsol.com

ε_m

Mathematical model — $\nabla^2 \cdot \boldsymbol{g} = -4\pi G\rho$

ε_d

Discretization — $\mathcal{V} \to \mathcal{V}_n \subset \mathcal{V}$

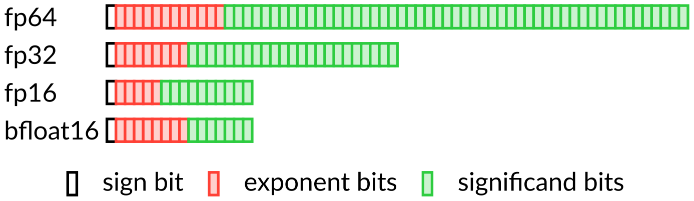
ε_a

Computation — $Ax = b \to A\hat{x} = b$

[1](Carson and Mary 2025)

Extension to mixed precision

Main idea: Strategic combination of different computer arithmetics to achieve high accuracy at lower cost.



Real world phenomena

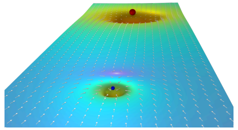


Figure 1: Source: https://www.comsol.com

ε_m

Mathematical model — $\nabla^2 \cdot \boldsymbol{g} = -4\pi G\rho$

ε_d

Discretization — $\mathcal{V} \to \mathcal{V}_n \subset \mathcal{V}$

ε_a

Computation — $Ax = b \to A\hat{x} = b$

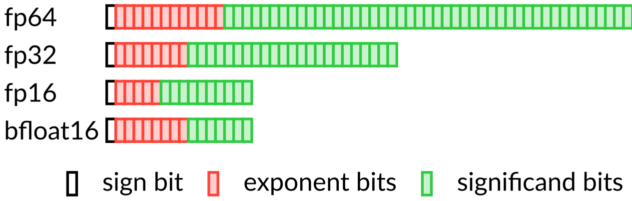
Approach: Perform rounding error analysis and identify precision requirements for different parts of computation.

- **Low precision:** Reduce computation and communication costs.
- **When:**[1] For instance, when the *finite precision error is small* compared to other sources of error.

[1](Carson and Mary 2025)

Extension to mixed precision

Main idea: Strategic combination of different computer arithmetics to achieve high accuracy at lower cost.



Approach: Perform rounding error analysis and identify precision requirements for different parts of computation.

- **Low precision:** Reduce computation and communication costs.
- **When:** [1] For instance, when the *finite precision error is small* compared to other sources of error.

Our context: Can we apply the preconditioners in *low* precision?

Real world phenomena

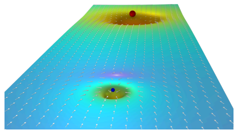


Figure 1: Source: https://www.comsol.com

$\downarrow \varepsilon_m$

Mathematical model — $\nabla^2 \cdot \boldsymbol{g} = -4\pi G\rho$

$\downarrow \varepsilon_d$

Discretization — $\mathcal{V} \to \mathcal{V}_n \subset \mathcal{V}$

$\downarrow \varepsilon_a$

Computation — $Ax = b \to A\hat{x} = b$

[1] (Carson and Mary 2025)

Mixed precision bounds

Corollary . Let $\boldsymbol{u}_s$, $\boldsymbol{u}_q$, and $\boldsymbol{u}_z$ denote the unit roundoff values associated with the precisions involved in the application of the preconditioners M_L, M_R, and M_R^T, respectively. If $O(n\boldsymbol{u})\kappa(A) + O(n\boldsymbol{u}_1)\kappa(M)^{m_1} \leq \frac{1}{2}$, where

$$\boldsymbol{u}_1 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \boldsymbol{u}_z & \text{for right PCG,} \\ (\boldsymbol{u}_s + \boldsymbol{u}_z) & \text{for split PCG,} \end{cases} \text{ and } \; m_1 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ \frac{1}{2} & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $k^\star$ such that $f(\bar{x}_{k^\star}) - f(\bar{x}_{k^\star+1}) \leq O\left[n^2(k^\star)^2\boldsymbol{u}^2\right]\hat{\alpha}_{k^\star}\|A\|^2\|M^{-1}\|\|x\|^2$ holds.

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1 - O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\boldsymbol{u}_2\kappa(M)^{m_2} + O[k^\star(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}} \leq \frac{1}{2},$$

where

$$\boldsymbol{u}_2 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \left(\boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for right PCG,} \\ \left(\boldsymbol{u}_s + \boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for split PCG,} \end{cases} \text{ and } \; m_2 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ 1 & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $i \leq k^\star$ such that bounds (1) - (3) hold.

Mixed precision bounds

Corollary . Let $\boldsymbol{u}_s$, $\boldsymbol{u}_q$, and $\boldsymbol{u}_z$ denote the unit roundoff values associated with the precisions involved in the application of the preconditioners M_L, M_R, and M_R^T, respectively. If $O(n\boldsymbol{u})\kappa(A) + O(n\boldsymbol{u}_1)\kappa(M)^{m_1} \leq \frac{1}{2}$, where

$$\boldsymbol{u}_1 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \boldsymbol{u}_z & \text{for right PCG,} \\ (\boldsymbol{u}_s + \boldsymbol{u}_z) & \text{for split PCG,} \end{cases} \quad \text{and} \quad m_1 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ \frac{1}{2} & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $k^\star$ such that $f(\bar{x}_{k^\star}) - f(\bar{x}_{k^\star+1}) \leq O\left[n^2(k^\star)^2\boldsymbol{u}^2\right]\hat{\alpha}_{k^\star}\|A\|^2\|M^{-1}\|\|x\|^2$ holds.

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1 - O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\boldsymbol{u}_2\kappa(M)^{m_2} + O[k^\star(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}} \leq \frac{1}{2},$$

where

$$\boldsymbol{u}_2 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \left(\boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for right PCG,} \\ \left(\boldsymbol{u}_s + \boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for split PCG,} \end{cases} \quad \text{and} \quad m_2 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ 1 & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $i \leq k^\star$ such that bounds (1) - (3) hold.

Main point: If some conditions are met, we can

Mixed precision bounds

Corollary . Let $\boldsymbol{u}_s$, $\boldsymbol{u}_q$, and $\boldsymbol{u}_z$ denote the unit roundoff values associated with the precisions involved in the application of the preconditioners M_L, M_R, and M_R^T, respectively. If $O(n\boldsymbol{u})\kappa(A) + O(n\boldsymbol{u}_1)\kappa(M)^{m_1} \leq \frac{1}{2}$, where

$$\boldsymbol{u}_1 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \boldsymbol{u}_z & \text{for right PCG,} \\ (\boldsymbol{u}_s + \boldsymbol{u}_z) & \text{for split PCG,} \end{cases} \text{ and } m_1 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ \frac{1}{2} & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $k^\star$ such that $f(\bar{x}_{k^\star}) - f(\bar{x}_{k^\star+1}) \leq O\left[n^2(k^\star)^2\boldsymbol{u}^2\right]\hat{\alpha}_{k^\star}\|A\|^2\|M^{-1}\|\|x\|^2$ holds.

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1 - O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\boldsymbol{u}_2\kappa(M)^{m_2} + O[k^\star(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}} \leq \frac{1}{2},$$

where

$$\boldsymbol{u}_2 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \left(\boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for right PCG,} \\ \left(\boldsymbol{u}_s + \boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for split PCG,} \end{cases} \text{ and } m_2 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ 1 & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $i \leq k^\star$ such that bounds (1) - (3) hold.

Main point: If some conditions are met, we can

- potentially apply the preconditioners in *different precisions*, and

Mixed precision bounds

Corollary . Let $\boldsymbol{u}_s$, $\boldsymbol{u}_q$, and $\boldsymbol{u}_z$ denote the unit roundoff values associated with the precisions involved in the application of the preconditioners M_L, M_R, and M_R^T, respectively. If $O(n\boldsymbol{u})\kappa(A) + O(n\boldsymbol{u}_1)\kappa(M)^{m_1} \leq \frac{1}{2}$, where

$$\boldsymbol{u}_1 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \boldsymbol{u}_z & \text{for right PCG,} \\ (\boldsymbol{u}_s + \boldsymbol{u}_z) & \text{for split PCG,} \end{cases} \text{ and } \; m_1 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ \frac{1}{2} & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $k^\star$ such that $f(\bar{x}_{k^\star}) - f(\bar{x}_{k^\star+1}) \leq O\left[n^2(k^\star)^2\boldsymbol{u}^2\right]\hat{\alpha}_{k^\star}\|A\|^2\|M^{-1}\|\|x\|^2$ holds.

Furthermore, if $k^\star$ satisfies

$$\frac{O[n(k^\star+1)\boldsymbol{u}]\kappa(A)}{1 - O(n\boldsymbol{u})\kappa(A)} + O(k^\star+1)\boldsymbol{u}_2\kappa(M)^{m_2} + O[k^\star(k^\star+1)\boldsymbol{u}]\kappa(M)^{\frac{1}{2}}\kappa\left(M_L^{-1}AM_R^{-1}\right)^{\frac{1}{2}} \leq \frac{1}{2},$$

where

$$\boldsymbol{u}_2 = \begin{cases} \boldsymbol{u}_s & \text{for left PCG,} \\ \left(\boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for right PCG,} \\ \left(\boldsymbol{u}_s + \boldsymbol{u}_q + \boldsymbol{u}_z\right) & \text{for split PCG,} \end{cases} \text{ and } \; m_2 = \begin{cases} \frac{3}{2} & \text{for left or right PCG,} \\ 1 & \text{for split PCG,} \end{cases}$$

then there exists an iteration step $i \leq k^\star$ such that bounds (1) - (3) hold.

Main point: If some conditions are met, we can

- potentially apply the preconditioners in *different precisions*, and
- our forward and backward error *bounds still hold*!

Numerical experiments

Numerical experiments

- Given $0 < \lambda_1 < \lambda_n$: $A = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$,

$$\boxed{\lambda_i = \lambda_1 + \frac{i-1}{n-1}(\lambda_n - \lambda_1)\rho^{(n-i)}, \quad i = 2, \dots, n-1, \quad \rho \in [0,1].}$$

[1] We follow the settings described in (Carson, Liesen, and Strakoš 2024), and (Vieublé 2022) for these constructions.

Numerical experiments

- Given $0 < \lambda_1 < \lambda_n$: $A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$,

$$\boxed{\lambda_i = \lambda_1 + \frac{i-1}{n-1}(\lambda_n - \lambda_1)\rho^{(n-i)}, \quad i = 2, \ldots, n-1, \quad \rho \in [0, 1].}$$

- For $j \in \{2, 3, \ldots, n\}$ we set $M = \mathrm{diag}(\lambda_1, \ldots, \lambda_{j-1}, \lambda_j, \ldots, \lambda_j)$, and let $M = LL^\top$[1].

[1] We follow the settings described in (Carson, Liesen, and Strakoš 2024), and (Vieublé 2022) for these constructions.

Numerical experiments

- Given $0 < \lambda_1 < \lambda_n$: $A = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$,

$$\boxed{\lambda_i = \lambda_1 + \frac{i-1}{n-1}(\lambda_n - \lambda_1)\rho^{(n-i)}, \quad i = 2, \dots, n-1, \quad \rho \in [0, 1].}$$

- For $j \in \{2, 3, \dots, n\}$ we set $M = \mathrm{diag}(\lambda_1, \dots, \lambda_{j-1}, \lambda_j, \dots, \lambda_j)$, and let $M = LL^\top$[1].

- In exact arithmetic, the matrices $M^{-1}A$, AM^{-1} and $L^{-1}AL^{-\top}$ are similar with the spectrum

$$S = \left\{1, \frac{\lambda_{j+1}}{\lambda_j}, \frac{\lambda_{j+2}}{\lambda_j}, \dots, \frac{\lambda_n}{\lambda_j}\right\},$$

 for some chosen value of j.

[1]We follow the settings described in (Carson, Liesen, and Strakoš 2024), and (Vieublé 2022) for these constructions.

Numerical experiments

- Given $0 < \lambda_1 < \lambda_n$: $A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$,

$$\boxed{\lambda_i = \lambda_1 + \frac{i-1}{n-1}(\lambda_n - \lambda_1)\rho^{(n-i)}, \quad i = 2, \ldots, n-1, \quad \rho \in [0,1].}$$

- For $j \in \{2, 3, \ldots, n\}$ we set $M = \mathrm{diag}(\lambda_1, \ldots, \lambda_{j-1}, \lambda_j, \ldots, \lambda_j)$, and let $M = LL^\top$[1].

- In exact arithmetic, the matrices $M^{-1}A$, AM^{-1} and $L^{-1}AL^{-\top}$ are similar with the spectrum

$$S = \left\{1, \frac{\lambda_{j+1}}{\lambda_j}, \frac{\lambda_{j+2}}{\lambda_j}, \ldots, \frac{\lambda_n}{\lambda_j}\right\},$$

for some chosen value of j.

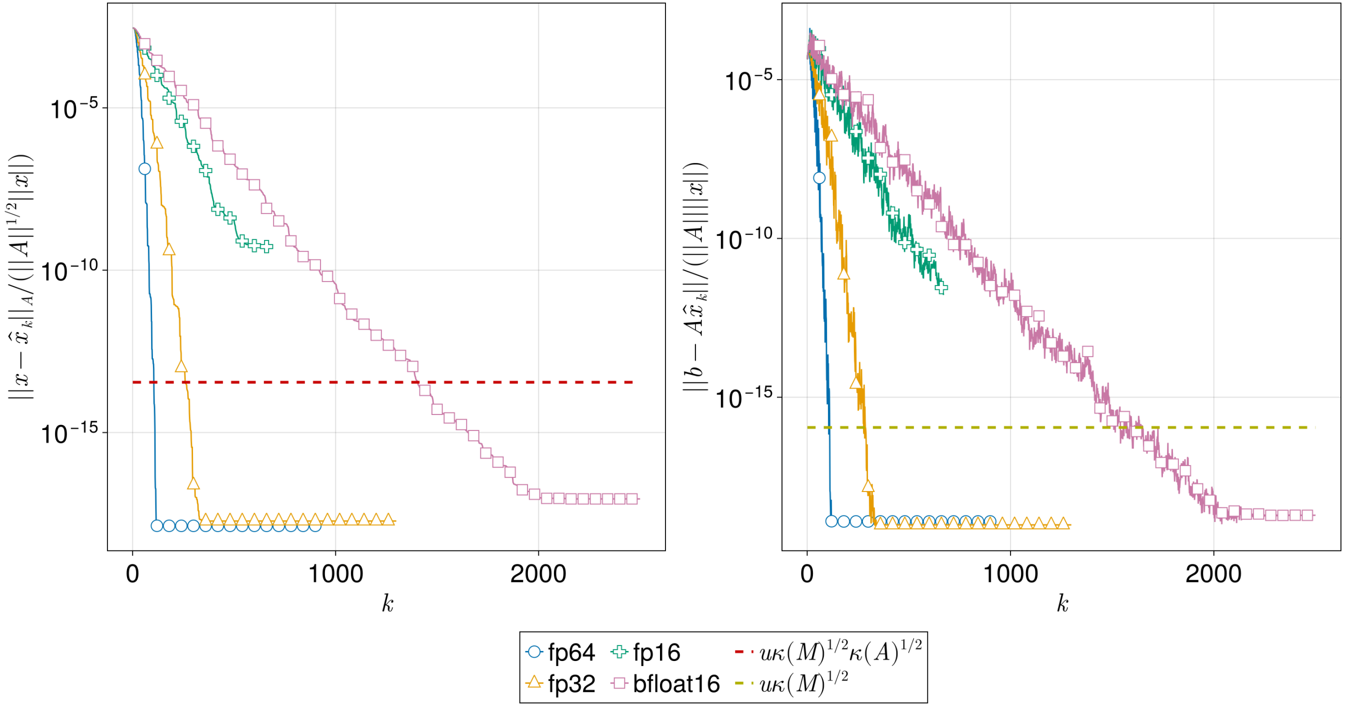
- We set $n = 85$, $b = \frac{1}{\sqrt{n}} \cdot [1, \ldots, 1]^\top$, $x_0 = [0, \ldots, 0]^\top$ for all the experiments.

- We also fix $\lambda_1 = 1, \lambda_n = 10^5$, and $\rho = 0.6 \implies$ accumulation of eigenvalues to the left.

[1]We follow the settings described in (Carson, Liesen, and Strakoš 2024), and (Vieublé 2022) for these constructions.

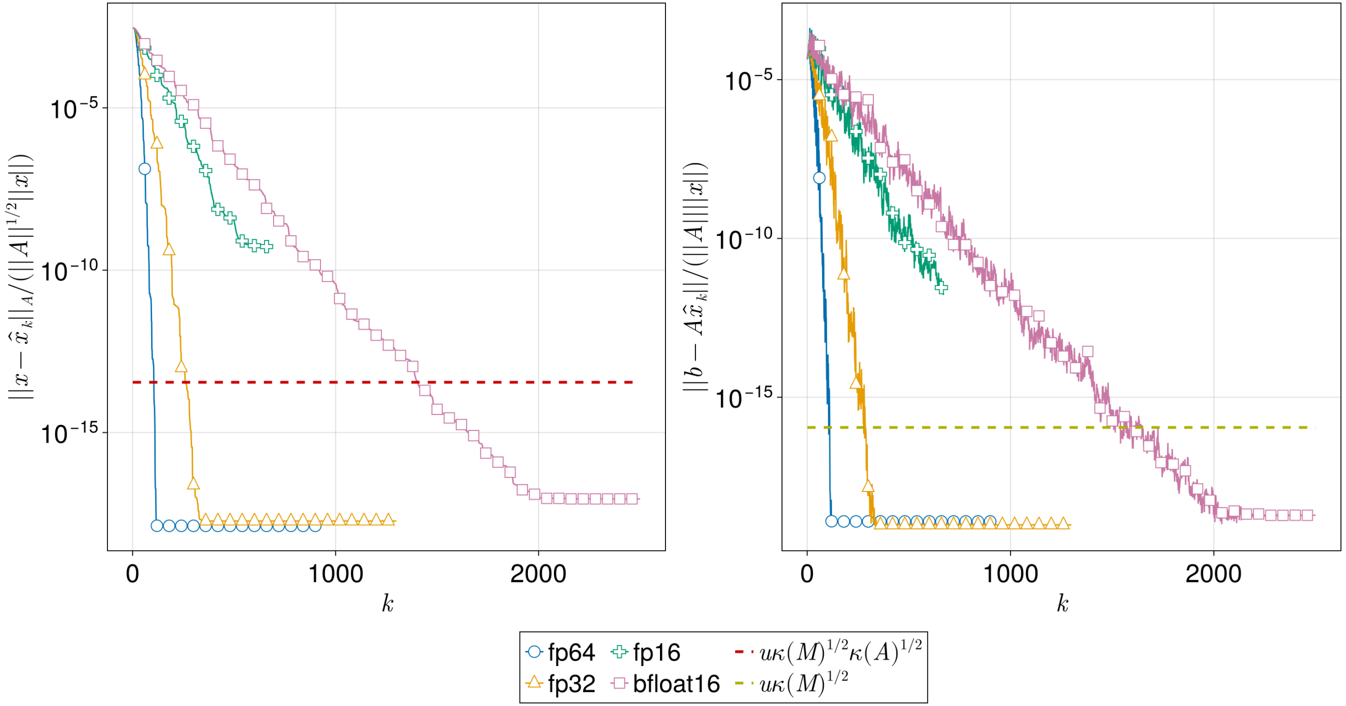
Left PCG

- Set $j = 55$



Left PCG

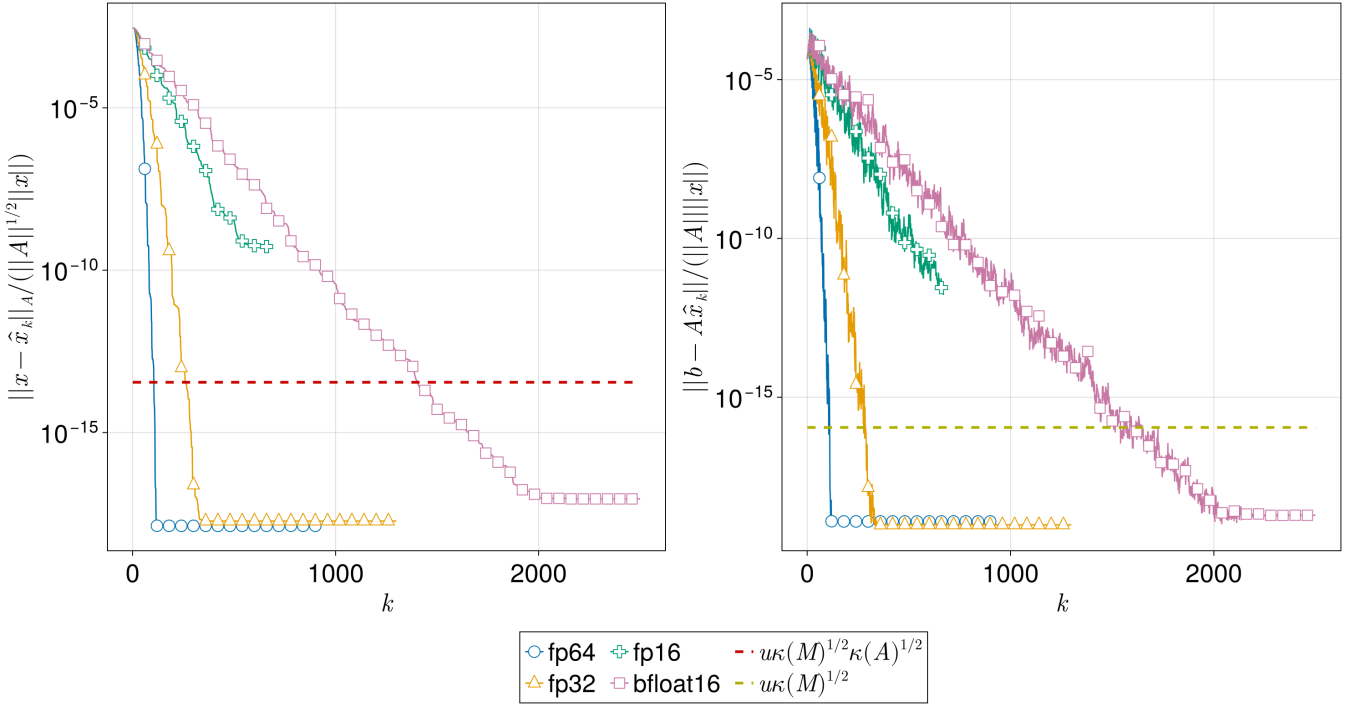
- Set $j = 55$



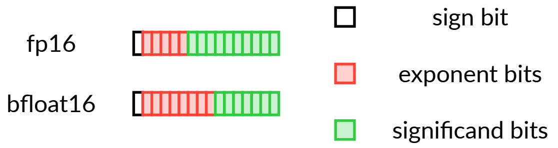
Arithmetic	Unit roundoff	Smallest subnormal number
bfloat16	3.91×10^{-3}	9.18×10^{-41}
fp16 (half)	4.88×10^{-4}	5.96×10^{-8}
fp32 (single)	5.96×10^{-8}	1.40×10^{-45}
fp64 (double)	1.11×10^{-16}	4.94×10^{-324}

Left PCG

- Set $j = 55$

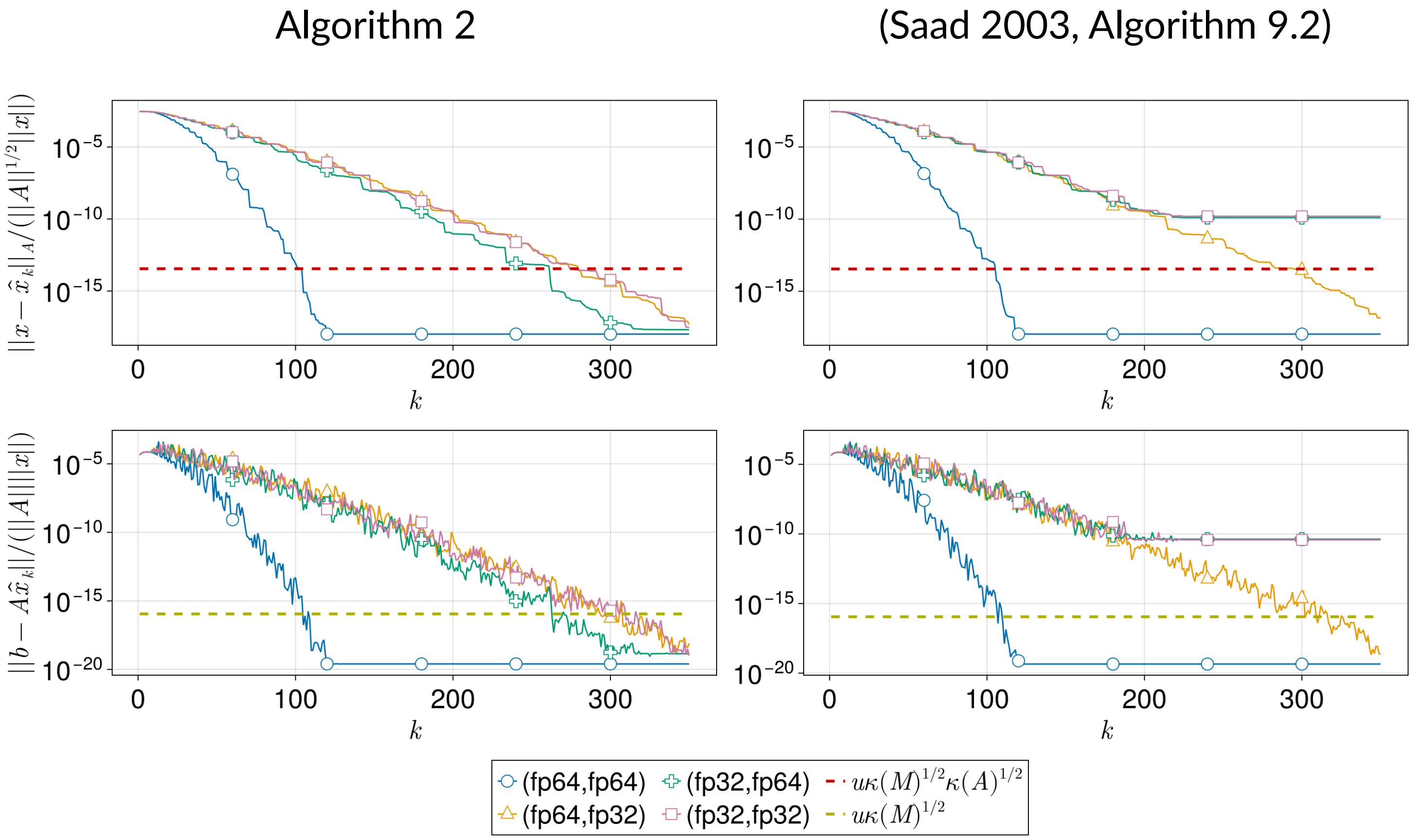


Arithmetic	Unit roundoff	Smallest subnormal number
bfloat16	3.91×10^{-3}	9.18×10^{-41}
fp16 (half)	4.88×10^{-4}	5.96×10^{-8}
fp32 (single)	5.96×10^{-8}	1.40×10^{-45}
fp64 (double)	1.11×10^{-16}	4.94×10^{-324}



Split PCG comparison

- Set $j = 65$.



Split PCG comparison

Algorithm 2

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

$$s_{k+1} = M_L^{-1} r_{k+1}$$

$$\|b - A\hat{x}_k - \hat{r}_k\| \le (nk\boldsymbol{u})\|A\| \max_{j\le k}\Big(\|\hat{x}_j\|, \|x\|\Big)$$

(Saad 2003, Algorithm 9.2)

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

$$\|b - A\hat{x}_k - \hat{r}_k\| \le \|A\| \cdot \max_{j\le k}\Big(\|\hat{x}_j\|, \|x\|\Big)\Big[O(nk\boldsymbol{u}) + O(\boldsymbol{u}_s)\kappa(M_L)\Big]$$

Summary and open questions

Summary and open questions

- Forward and backward error bounds for a mixed precision PCG algorithm.

Summary and open questions

- Forward and backward error bounds for a mixed precision PCG algorithm.
- Achieved using interconnections to other mathematical representations of CG.

Summary and open questions

- Forward and backward error bounds for a mixed precision PCG algorithm.
- Achieved using interconnections to other mathematical representations of CG.
- Can we remove influence or bound the iteration step(s) k?

Summary and open questions

- Forward and backward error bounds for a mixed precision PCG algorithm.
- Achieved using interconnections to other mathematical representations of CG.
- Can we remove influence or bound the iteration step(s) k?
- Preprint available:

Summary and open questions

- Forward and backward error bounds for a mixed precision PCG algorithm.
- Achieved using interconnections to other mathematical representations of CG.
- Can we remove influence or bound the iteration step(s) k?
- Preprint available:

Thank you for your attention!

References

B., Thomas, Erin Carson, and Yuxin Ma. 2025. “Forward and Backward Error Bounds for a Mixed Precision Preconditioned Conjugate Gradient Algortihm.” https://arxiv.org/abs/2510.11379.

Carson, Erin, and Theo Mary. 2025. “Mixed-Precision Computing: High Accuracy with Low Precision.” https://www.siam.org/publications/siam-news/articles/mixed-precision-computing-high-accuracy-with-low-precision/.

Carson, Erin, Jörg Liesen, and Zdeněk Strakoš. 2024. “Towards Understanding CG and GMRES Through Examples.” *Linear Algebra Appl.*

Greenbaum, Anne. 1989. “Behavior of Slightly Perturbed Lanczos and Conjugate-Gradient Recurrences.” *Linear Algebra Appl.*

Greenbaum, Anne. 1997. “Estimating the Attainable Accuracy of Recursively Computed Residual Methods.” *SIAM J. Matrix Anal. Appl.*

Paige, C. C. 1976. “Error Analysis of the Lanczos Algorithm for Tridiagonalizing a Symmetric Matrix.” *IMA Journal of Applied Mathematics*.

Saad, Yousef. 2003. *Iterative Methods for Sparse Linear Systems*. 2nd ed. SIAM.

Vieublé, Bastien. 2022. (INP Toulouse) “Mixed precision iterative refinement for the solution of large sparse linear systems.” Theses. https://hal.science/tel-03975935.