

# **Lecture 10**

## **Conditioning and Stability II**

MIT 18.335J / 6.337J

Introduction to Numerical Methods

Per-Olof Persson

October 12, 2006

# Backward Stability of Householder QR

- For a QR factorization  $A = QR$  computed by Householder triangularization, the factors  $\tilde{Q}$  and  $\tilde{R}$  satisfy

$$\tilde{Q}\tilde{R} = A + \delta A, \quad \frac{\|\delta A\|}{\|A\|} = O(\epsilon_{\text{machine}})$$

- Exactly the right  $QR$  factorization of a slightly perturbed  $A$
- As usual,  $\tilde{R}$  is the  $R$  computed by the algorithm using floating points
- However,  $\tilde{Q}$  is a product of *exactly unitary* reflectors:

$$\tilde{Q} = \tilde{Q}_1 \tilde{Q}_2 \cdots \tilde{Q}_n$$

where  $\tilde{Q}_k$  is implicitly given by the computed  $\tilde{v}_k$  (since  $Q$  is generally not formed explicitly)

# Backward Stability of Solving $Ax = b$ with QR

## Algorithm: Solving $Ax = b$ by QR Factorization

1.  $QR = A$  using Householder, represent  $Q$  by reflectors
2.  $y = Q^*b$  implicitly using reflectors
3.  $x = R^{-1}y$  by back substitution

- Step 1 is backward stable (from previous slide)
- Step 2 can be shown to be backward stable:

$$(\tilde{Q} + \delta Q)\tilde{y} = b, \quad \|\delta Q\| = O(\epsilon_{\text{machine}})$$

- Step 3 is backward stable (will be shown later):

$$(\tilde{R} + \delta R)\tilde{x} = \tilde{y}, \quad \frac{\|\delta R\|}{\|\tilde{R}\|} = O(\epsilon_{\text{machine}})$$

# Backward Stability of Solving $Ax = b$ with QR

- Put the three steps together to show backward stability of the algorithm:

$$(A + \Delta A)\tilde{x} = b, \quad \frac{\|\Delta A\|}{\|A\|} = O(\epsilon_{\text{machine}})$$

- Proof.* Steps 2 and 3 give

$$b = (\tilde{Q} + \delta Q)(\tilde{R} + \delta R)\tilde{x} = \left[ \tilde{Q}\tilde{R} + (\delta Q)\tilde{R} + \tilde{Q}(\delta R) + (\delta Q)(\delta R) \right] \tilde{x}$$

Step 1 (backward stability of Householder) gives

$$\begin{aligned} b &= \left[ A + \delta A + (\delta Q)\tilde{R} + \tilde{Q}(\delta R) + (\delta Q)(\delta R) \right] \tilde{x} \\ &= (A + \Delta A)\tilde{x} \end{aligned}$$

# Backward Stability of Solving $Ax = b$ with QR

$\delta A$  is small compared to  $A$ , therefore

$$\frac{\|\tilde{R}\|}{\|A\|} \leq \|\tilde{Q}^*\| \frac{\|A + \delta A\|}{\|A\|} = O(1)$$

Now show that each term in  $\Delta A$  is small:

$$\frac{\|(\delta Q)\tilde{R}\|}{\|A\|} \leq \|\delta Q\| \frac{\|\tilde{R}\|}{\|A\|} = O(\epsilon_{\text{machine}})$$

$$\frac{\|\tilde{Q}(\delta R)\|}{\|A\|} \leq \|\tilde{Q}\| \frac{\|\delta R\|}{\|\tilde{R}\|} \frac{\|\tilde{R}\|}{\|A\|} = O(\epsilon_{\text{machine}})$$

$$\frac{\|(\delta Q)(\delta R)\|}{\|A\|} \leq \|\delta Q\| \frac{\|\delta R\|}{\|A\|} = O(\epsilon_{\text{machine}}^2)$$

# Backward Stability of Solving $Ax = b$ with QR

Add the terms to show that  $\Delta A$  is small:

$$\begin{aligned}\frac{\|\Delta A\|}{\|A\|} &\leq \frac{\|\delta A\|}{\|A\|} + \frac{\|(\delta Q)\tilde{R}\|}{\|A\|} + \frac{\|\tilde{Q}(\delta R)\|}{\|A\|} + \frac{\|(\delta Q)(\delta R)\|}{\|A\|} \\ &= O(\epsilon_{\text{machine}})\end{aligned}$$

- Since the algorithm is backward stable, it is also accurate:

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O(\kappa(A)\epsilon_{\text{machine}})$$

# Backward Stability of Back Substitution

- Solve  $Rx = b$  using back substitution:

$$\begin{bmatrix} r_{11} & \cdots & r_{1m} \\ & \ddots & \vdots \\ & & r_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$$x_m = b_m / r_{mm}$$

$$x_{m-1} = (b_{m-1} - x_m r_{m-1,m}) / r_{m-1,m-1}$$

$$x_{m-2} = (b_{m-2} - x_{m-1} r_{m-2,m-1} - x_m r_{m-2,m}) / r_{m-2,m-2}$$

$\vdots$

$$x_j = \left( b_j - \sum_{k=j+1}^m x_k r_{jk} \right) / r_{jj}$$

# Backward Stability of Back Substitution

- Back substitution is backward stable:

$$(R + \delta R)\tilde{x} = b, \quad \frac{\|\delta R\|}{\|R\|} = O(\epsilon_{\text{machine}})$$

Furthermore, each component of  $\delta R$  satisfies

$$\frac{|\delta r_{ij}|}{|r_{ij}|} \leq m\epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$$

- Show in full detail for  $m = 1, 2, 3$  as well as general  $m$

# Proof for Back Substitution ( $m = 1$ )

- For  $m = 1$ , the algorithm is simply one floating point division. Use the floating points axiom to get

$$\tilde{x}_1 = b_1 \oslash r_{11} = \frac{b_1}{r_{11}}(1 + \epsilon_1) = \frac{b_1}{r_{11}(1 + \epsilon'_1)}$$

where  $|\epsilon_1| \leq \epsilon_{\text{machine}}$  and  $|\epsilon'_1| \leq \epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$

Therefore, we solved a perturbed problem exactly:

$$(r_{11} + \delta r_{11})\tilde{x}_1 = b_1 \quad \text{with} \quad \frac{|\delta r_{11}|}{|r_{11}|} \leq \epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$$

## Proof for Back Substitution ( $m = 2$ )

- For  $m = 2$ , we first solve for  $\tilde{x}_2$  as before. Next we compute  $\tilde{x}_1$ :

$$\begin{aligned}\tilde{x}_1 &= (b_1 \ominus (\tilde{x}_2 \otimes r_{12})) \oslash r_{11} = \frac{(b_1 - \tilde{x}_2 r_{12}(1 + \epsilon_2))(1 + \epsilon_3)}{r_{11}}(1 + \epsilon_4) \\ &= \frac{b_1 - \tilde{x}_2 r_{12}(1 + \epsilon_2)}{r_{11}(1 + \epsilon'_3)(1 + \epsilon'_4)} = \frac{b_1 - \tilde{x}_2 r_{12}(1 + \epsilon_2)}{r_{11}(1 + 2\epsilon_5)}\end{aligned}$$

where

$$|\epsilon_2|, |\epsilon_3|, |\epsilon_4| \leq \epsilon_{\text{machine}} \quad \text{and} \quad |\epsilon'_3|, |\epsilon'_4|, |\epsilon_5| \leq \epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$$

Again this is an exact solution to  $(R + \delta R)\tilde{x} = b$  with

$$\begin{bmatrix} \frac{|\delta r_{11}|}{|r_{11}|} & \frac{|\delta r_{12}|}{|r_{12}|} \\ & \frac{|\delta r_{22}|}{|r_{22}|} \end{bmatrix} = \begin{bmatrix} 2|\epsilon_5| & |\epsilon_2| \\ & |\epsilon_1| \end{bmatrix} \leq \begin{bmatrix} 2 & 1 \\ & 1 \end{bmatrix} \epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$$

# Proof for Back Substitution ( $m = 3$ )

- For  $m = 3$ , compute  $\tilde{x}_3$  and  $\tilde{x}_2$  as before. Then compute

$$\begin{aligned}
 \tilde{x}_1 &= [(b_1 \ominus (\tilde{x}_2 \otimes r_{12})) \ominus (\tilde{x}_3 \otimes r_{13})] \oslash r_{11} \\
 &= \frac{[(b_1 - \tilde{x}_2 r_{12}(1 + \epsilon_4))(1 + \epsilon_6) - \tilde{x}_3 r_{13}(1 + \epsilon_5)](1 + \epsilon_7)}{r_{11}(1 + \epsilon'_8)} \\
 &= \frac{b_1 - \tilde{x}_2 r_{12}(1 + \epsilon_4) - \tilde{x}_3 r_{13}(1 + \epsilon_5)(1 + \epsilon'_6)}{r_{11}(1 + \epsilon'_6)(1 + \epsilon'_7)(1 + \epsilon'_8)}
 \end{aligned}$$

That is,  $(R + \delta R)\tilde{x} = b$  with

$$\begin{bmatrix} \frac{|\delta r_{11}|}{|r_{11}|} & \frac{|\delta r_{12}|}{|r_{12}|} & \frac{|\delta r_{13}|}{|r_{13}|} \\ & \frac{|\delta r_{22}|}{|r_{22}|} & \frac{|\delta r_{23}|}{|r_{23}|} \\ & & \frac{|\delta r_{33}|}{|r_{33}|} \end{bmatrix} \leq \begin{bmatrix} 3 & 1 & 2 \\ & 2 & 1 \\ & & 1 \end{bmatrix} \epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$$

# Proof for Back Substitution (general $m$ )

- Similar analysis for general  $m$  gives the pattern (shown for  $m = 5$ ):

$$\frac{|\delta R|}{|R|} \leq W \epsilon_{\text{machine}} + O(\epsilon_{\text{machine}}^2)$$

where  $W =$

$$\underbrace{\begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ & 0 & 1 & 1 & 1 \\ & & 0 & 1 & 1 \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}}_{\otimes} + \underbrace{\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}}_{\oplus} + \underbrace{\begin{bmatrix} 4 & 0 & 1 & 2 & 3 \\ & 3 & 0 & 1 & 2 \\ & & 2 & 0 & 1 \\ & & & 1 & 0 \\ & & & & 0 \end{bmatrix}}_{\ominus} = \begin{bmatrix} 5 & 1 & 2 & 3 & 4 \\ & 4 & 1 & 2 & 3 \\ & & 3 & 1 & 2 \\ & & & 2 & 1 \\ & & & & 1 \end{bmatrix}$$