

# Householder Triangularization

Note Title

4/26/2012

Before discussing Householder triangularization, we need to interpret:

## \* MGS as Triangular Orthogonalization

Recall the modified Gram-Schmidt (MGS) algorithm.

• Initial set up :  $v_j^{(1)} = a_j \quad 1 \leq j \leq n$

• The first step in the outer for loop:

$$\underbrace{[v_1^{(1)} \ v_2^{(1)} \ \dots \ v_n^{(1)}]}_{= A} \begin{bmatrix} \frac{1}{r_{11}} & -\frac{r_{12}}{r_{11}} & \dots & -\frac{r_{1n}}{r_{11}} \\ 0 & & & \\ \vdots & & I_{n-1 \times n-1} & \\ 0 & & & \end{bmatrix} = \underbrace{[\xi_1 \ v_2^{(2)} \ \dots \ v_n^{(2)}]}_{= AR_1}$$

call this  $R_1$

• The second step:

$$\underbrace{[\xi_1 \ v_2^{(2)} \ \dots \ v_n^{(2)}]}_{= AR_1} \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{r_{22}} & -\frac{r_{23}}{r_{22}} & \dots & -\frac{r_{2n}}{r_{22}} \\ 0 & & & & \\ \vdots & & & & \\ 0 & & & & I_{n-2 \times n-2} \end{bmatrix} = \underbrace{[\xi_1 \ \xi_2 \ v_3^{(3)} \ \dots \ v_n^{(3)}]}_{= AR_1 R_2}$$

call this  $R_2$

⋮

In the end, we get

$$A R_1 R_2 \dots R_n = \hat{Q} = [\xi_1 \ \dots \ \xi_n]$$

call this  $\hat{R}^{-1}$

$$\text{Then } A = \hat{Q} \hat{R} !$$

Hence, we can view

MGS = triangular orthogonalization

meaning: applying triangular operations to reduce to orthonormal col' vectors.

Note: In practice, we do **not** form matrices  $R_i$ ,  $i=1, \dots, n$ .

These are used to interpret the meaning of the MGS algorithm.

★ Householder Triangularization

= orthogonal triangularization!

Instead of triangular orthogonalization.

$$\text{MGS: } A \underbrace{R_1 R_2 \dots R_n}_{= \hat{R}^{-1}} = Q \quad \begin{array}{l} \text{reduced} \\ \text{QR} \end{array}$$

$$\text{Householder: } \underbrace{Q_n Q_{n-1} \dots Q_2 Q_1}_{= Q^T} A = R \quad \begin{array}{l} \text{full QR} \end{array}$$

In essence, it's a triangularization by introducing zeros (Householder, 1958)

$$\begin{array}{c} \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \xrightarrow{Q_1} \begin{bmatrix} \Delta & \Delta & \Delta \\ 0 & \Delta & \Delta \\ 0 & \Delta & \Delta \\ 0 & \Delta & \Delta \\ 0 & \Delta & \Delta \end{bmatrix} \xrightarrow{Q_2} \begin{bmatrix} \Delta & \Delta & \Delta \\ 0 & + & + \\ 0 & 0 & + \\ 0 & 0 & + \\ 0 & 0 & + \end{bmatrix} \xrightarrow{Q_3} \begin{bmatrix} \Delta & \Delta & \Delta \\ 0 & + & + \\ 0 & 0 & \nabla \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ A \qquad \qquad Q_1 A \qquad \qquad Q_2 Q_1 A \qquad \qquad Q_3 Q_2 Q_1 A \end{array}$$

How to construct  $Q_k$  ?  
 $\Rightarrow$  Householder Reflector

$I_{k-1}$   
 is a  
 short  
 notation  
 for  $I_{k-1 \times k-1}$

$$Q_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & F \end{bmatrix} \begin{matrix} \} k-1 \\ \} m-(k-1) \end{matrix}$$

$F \in \mathbb{R}^{(m-k+1) \times (m-k+1)}$ ,  $F^T F = F F^T = I_{m-k+1}$   
 i.e., ortho. mat.  
Householder Reflector

• What  $F$  does is the following:

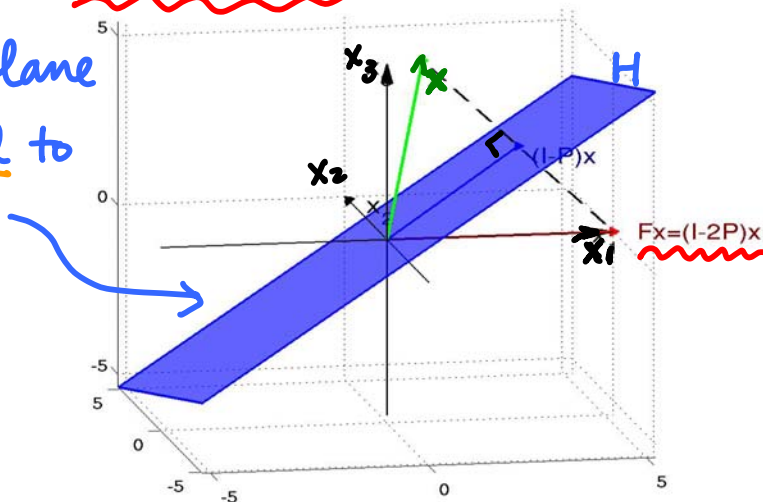
Suppose  $x \in \mathbb{R}^{m-k+1}$ , i.e.,

$$x = \begin{bmatrix} * \\ * \\ * \\ \vdots \\ * \end{bmatrix} \xrightarrow{F} Fx = \begin{bmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \|x\| e_1$$

• Why  $F$  is called a reflector?

Define  $v = \|x\| e_1 - x$

$H$ : Hyperplane  
normal to  
 $v$



$$P = P_v = \frac{v v^T}{v^T v} : \text{ortho. proj. onto } \langle v \rangle$$

But  $v \perp H$ .

Hence  $I - P_v$  : ortho. proj. onto  $H$ .

Now  $F$  can be written as

$$F = I - 2P_v = I - 2 \frac{v v^T}{v^T v}$$

Note  $F$  is not a projector.

but  $F$  is an orthogonal matrix.

why?  $(I - 2P_v)^2 = (I - 2P_v)(I - 2P_v)$

$$= I - 2P_v - 2P_v + 4P_v^2$$

$$= I - 4P_v + 4P_v = I$$

$$\neq I - 2P_v \checkmark$$

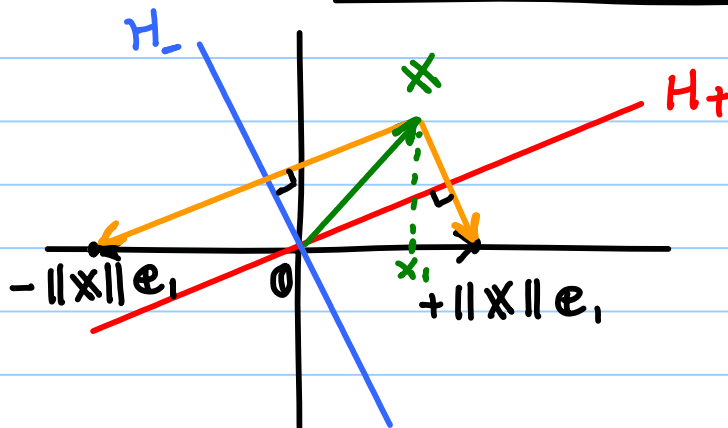
$$(I - 2P_v)^T (I - 2P_v) = (I^T - 2P_v^T)(I - 2P_v)$$

$$P_v^T = P_v \rightarrow = (I - 2P_v)(I - 2P_v)$$

$$= I \checkmark$$

$$\text{Similarly } (I - 2P_v)(I - 2P_v)^T = I. \checkmark$$

★ Two possible reflectors :  
which is better?



Answer: In the above figure,

$-\|x\|e_1$  is a better choice.

Why? It's better numerically to move  $Fx$  as far from  $x$  as possible.

↳ Why? Computing  $v$  involves

$v = \pm \|x\|e_1 - x$ . If  $\|x\|e_1 \approx x$

then  $\|x\|e_1 - x$  loses numerical accuracy called cancellation error.

Hence in general,  $v$  should be chosen as

The first entry of  $x = \begin{bmatrix} x_1 \\ \vdots \end{bmatrix}$

$$v = -\operatorname{sgn}(x_1) \|x\| e_1 - x$$

$$\text{where } \operatorname{sgn}(x_1) := \begin{cases} 1 & \text{if } x_1 \geq 0 \\ -1 & \text{if } x_1 < 0 \end{cases}$$

or equivalently,

$$v = \operatorname{sgn}(x_1) \|x\| e_1 + x.$$

★ Algorithm (Householder QR fact.)

for  $k=1:n$

$$\left\{ \begin{array}{l} x = A(k:m, k) \\ v_k = \operatorname{sgn}(x_1) \|x\| e_1 + x \\ v_k = v_k / \|v_k\| \\ A(k:m, k:n) = A(k:m, k:n) \\ \quad - 2v_k (v_k^T A(k:m, k:n)) \\ \quad = FA(k:m, k:n) \end{array} \right.$$

Note: In the end,  $A$  is replaced by the final  $R$  of  $A = QR$ .

$Q$  has not been constructed explicitly.

This is OK since QR fact. is usually used as an intermediate process for solving some other problem, e.g.,  $Ax = b$  or  $\min \|b - Ax\|^2$

As we showed before, this leads to  $Rx = Q^T b$

Hence, as long as  $Q^T b$  is computed we often do not need  $Q$  itself.

Algorithm (Implicit calculation of  $Q^T b$ )

for  $k = 1:n$

$$b(k:m) = b(k:m) - 2v_k(v_k^T b(k:m))$$

↳ This should be included in the Householder QR fact. algorithm in the previous page. Or, just store  $v_k$ ,  $k = 1:n$ .

Do not store  $Q_k = I - 2v_k v_k^T$

No need to store these.

$v_k$ ,  $k = 1:n$  suffice.

Example Let's consider the familiar example matrix.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{bmatrix} \quad \begin{array}{l} \varepsilon : \text{small} \\ \text{s.t. } \varepsilon^2 \text{ can} \\ \text{be ignored.} \end{array}$$

I'm not going through the detailed computations by the Householder reflectors here (it would take several pages to do this by hand). But in the end, you get:

$$\begin{matrix} Q_3 & Q_2 & Q_1 \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\ 0 & 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & -\varepsilon & 0 & 0 \\ -\varepsilon & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & A \end{matrix}$$

$$= \begin{bmatrix} -1 & -1 & -1 \\ 0 & \sqrt{2}\varepsilon & \varepsilon/\sqrt{2} \\ 0 & 0 & (\frac{1}{\sqrt{3}} + \frac{\sqrt{2}}{\sqrt{3}})\varepsilon \\ 0 & 0 & 0 \end{bmatrix}$$

$R$

If we want  $A = QR$ ,

then  $\underbrace{Q_3 Q_2 Q_1}_{Q^T} A = R$

note each  $Q_i$  ortho mat.

$$\begin{aligned}
 \text{So, } Q &= (Q_3 Q_2 Q_1)^T \\
 &= Q_1^T Q_2^T Q_3^T \\
 &= Q_1 Q_2 Q_3
 \end{aligned}$$

In this example, we have

$$Q = \begin{bmatrix} -1 & \epsilon/\sqrt{2} & \epsilon/\sqrt{6} & -\epsilon/\sqrt{3} \\ -\epsilon & -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 0 & \sqrt{2/3} & 1/\sqrt{3} \end{bmatrix}$$

Now let's check  $Q^T Q$

$$Q^T Q = \begin{bmatrix} 1+\epsilon^2 & 0 & 0 & 0 \\ 0 & 1+\epsilon^2/2 & \epsilon^2/2\sqrt{3} & -\epsilon^2/\sqrt{6} \\ 0 & \epsilon^2/2\sqrt{3} & 1+\epsilon^2/6 & -\epsilon^2/3\sqrt{2} \\ 0 & -\epsilon^2/\sqrt{6} & -\epsilon^2/3\sqrt{2} & 1+\epsilon^2/3 \end{bmatrix}$$

$$\approx I_{4 \times 4} !$$

*no loss of  
orthogonality!*

This is a big difference from  
CGS and even MGS !



