Householder's method for QR factorization

Thursday, March 13, 2025 5:43 PM

Refs: ch 4.3 of Olver et al. (APPM 3310 textbook), and also see Golub and van Loan

Example Usage / Motivation

[2025: Suggestion: combine w/ ch 9.4–9.5 notes where we use Householder to reduce to upper Hessenberg]

note we could use it to solve a square linear system $A\cdot\vec{x}=\vec{b}$.

How? $A=QR$, so

$$Q\cdot R\cdot\vec{x}=\vec{b} \iff R\vec{x}=Q^{-1}\vec{b}$$
$$=Q^T\vec{b} \quad \text{via orthogonality of } Q$$

Solving $R\vec{x}=Q^T\vec{b}$ via backsubstitution since R is upper triangular.

When should you do this? It's a little slower than LU factorization, but slightly better if A is "ill-conditioned"
to be defined in a numerics class.

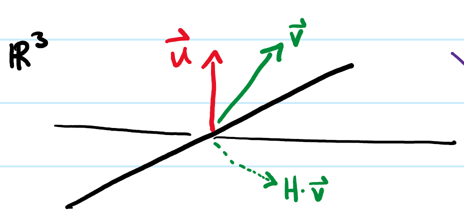
So, today: QR factorization of A if A is a bit ill-conditioned.

- Don't do Gram-Schmidt in this case
- Even modified Gram-Schmidt isn't great
- Book suggests "Householder's Method", which we'll discuss
  ... but if A is ill-conditioned, this should be combined with a column pivoting strategy. i.e. re-order vectors "Column pivoted QR"
- If matrix is sparse, use Givens rotations (faster, also stable)

Householder Reflectors aka "elementary reflection matrix"

Given $\vec{u}$ with $\|\vec{u}\|=1$ (in Euclidean norm),

$H=I-2\cdot\vec{u}\vec{u}^T$ is a Householder reflection



this plane is orthogonal to $\vec{u}$
aka $u^\perp=\{\vec{v}: \vec{v}\cdot\vec{u}=0\}$

$H\cdot\vec{v}$ reflects $\vec{v}$ about the line/plane/hyperplane $u^\perp$

Thursday, March 13, 2025 6:00 PM

Note $H = H^T$ and

$$H \cdot H^T = (I - 2\vec{u}\vec{u}^T)(I - 2\vec{u}\vec{u}^T) = I - 4\vec{u}\vec{u}^T + 4\vec{u}\underbrace{\vec{u}^T\vec{u}}_{=1 \text{ since } \|\vec{u}\|=1}\vec{u}^T = I$$

... so H is orthogonal (and symmetric)

i.e. $H^{-1} = H$!

We can choose $\vec{u}$ (hence H) to map a particular vector $\vec{v}$ to a target output $\vec{w}$ as long as $\|\vec{v}\| = \|\vec{w}\|$:

Lemma 4.28: Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ w/ $\|\vec{v}\| = \|\vec{w}\|$, then if $\vec{u} := \frac{\vec{v} - \vec{w}}{\|\vec{v} - \vec{w}\|}$ and $H = I - 2\vec{u}\vec{u}^T$,

then $H \cdot \vec{v} = \vec{w}$ (and since $H^2 = I$, $H\vec{w} = \vec{v}$)

proof: just calculate,

$$H \cdot \vec{v} = (I - 2\vec{u}\vec{u}^T)\vec{v} = \vec{v} - \frac{2}{\|\vec{v}-\vec{w}\|^2} \cdot (\vec{v} - \vec{w}) \cdot (\vec{v} - \vec{w})^T \vec{v}$$

$$= \vec{v} - \frac{2 \; (\vec{v} - \vec{w})}{\|\vec{v}\|^2 - 2\langle \vec{v}, \vec{w} \rangle + \|\vec{w}\|^2} \cdot \left(\|\vec{v}\|^2 - \langle \vec{w}, \vec{v} \rangle \right)$$

$$= \vec{v} - \frac{2 \left(\|\vec{v}\|^2 - \langle \vec{w}, \vec{v} \rangle \right)}{2\|\vec{v}\|^2 - 2\langle \vec{v}, \vec{w} \rangle} \cdot (\vec{v} - \vec{w}) \qquad \text{since } \|\vec{w}\|^2 = \|\vec{v}\|^2$$

$$= \vec{v} - (\vec{v} - \vec{w}) = \vec{w}. \qquad \square$$

Householder QR method

For now, doing full QR: Q will be square

Goal is $A = QR$ or... $Q^T A = R$.

Recall (orthogonal matrix) $\cdot$ (orthogonal matrix) = (orthogonal matrix)

so we'll build Q^T as the product of Householder reflectors

Step 1: $A = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{bmatrix}$ $\nwarrow \vec{v}_1$

Choose $\vec{u}_1$ (hence H_1) so that

$$H_1 \cdot A = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet \\ \boxed{0} & \bullet & \bullet & \bullet \\ \boxed{0} & \bullet & \bullet & \bullet \\ \boxed{0} & \bullet & \bullet & \bullet \end{bmatrix}$$

zero-out

Then...

Choose $\vec{u}_2$ (hence H_2) so that

$$H_2 \cdot (H_1 A) = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet \\ \boxed{0} & \bullet & \bullet & \bullet \\ \boxed{0} & \boxed{0} & \bullet & \bullet \\ \boxed{0} & \boxed{0} & \bullet & \bullet \end{bmatrix}$$

still 0 new 0

etc.

p. 3

Friday, March 14, 2025 6:19 AM

So for step 1, want to turn $\vec v_1 = \begin{bmatrix}\bullet\\\bullet\\\bullet\\\bullet\end{bmatrix}$ into $\vec\omega_1 = \begin{bmatrix}\bullet\\0\\0\\0\end{bmatrix}$ ← anything.

So... define $\vec\omega_1 = \|\vec v_1\|\cdot\vec e_1$ has the right zeros, and $\|\vec\omega_1\| = \|\vec v_1\|$

hence use our lemma: (choose $\vec u = \dfrac{\vec v - \vec\omega}{\|\vec v - \vec\omega\|}$ so that $H\vec v = \vec\omega$)

So, set $\vec u_1 = \dfrac{\vec v_1 - \|\vec v_1\|\cdot\vec e_1}{\|\vec v_1 - \|v_1\|\cdot\vec e_1\|}$

note: if numerator (hence also denominator) is zero, it means $\vec v_1 = c\cdot\vec e_1$ for some c, so in that case choose $\vec u_1 = \vec 0$, i.e. $H = I$

$$H_1 = I - 2\vec u_1\vec u_1^T$$

i.e. $\vec\omega_1 := \vec v_1 - \|\vec v_1\|\cdot\vec e_1$, $\vec u_1 = \vec\omega_1/\|\vec\omega_1\|$

thus

$$H_1\cdot A = \begin{bmatrix} r_{11} & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \end{bmatrix}$$

not the same as in original matrix

Step 2 Want to turn into $\begin{bmatrix}\bullet\\\bullet\\0\\0\end{bmatrix}$ and not ruin 1st column.

$\vec v_2$

two ways to think of it:

Way #1: define $H_2 = \left[\begin{array}{c|c}1 & 0\\\hline 0 & I - 2\vec u_2\cdot\vec u_2^T\end{array}\right]$ block matrix, i.e. at step k

(work in a smaller space)

$$\left[\begin{array}{c|c} (k-1)\times(k-1) \text{ identity} & 0\\\hline 0 & I - 2\vec u_k\vec u_k^T\end{array}\right]$$

where $\vec u_2 = \dfrac{\vec v_2 - \|\vec v_2\|\cdot\vec e_1}{\|\vec v_2 - \|\vec v_2\|\vec e_1\|}$ $\vec v_2$ are rows 2 through n of 2nd column of $\underbrace{H_1 A}_{\text{not of } A}$

$\vec e_1 \in \mathbb{R}^{n-1}$ $\vec v_2 \in \mathbb{R}^{n-1}$ $u_k \in \mathbb{R}^{n-k+1}$

Way #2: (equivalent) define $\tilde v_2 = \begin{bmatrix}0\\\vec v_2\end{bmatrix}$ } append zeros (at step k, this is $k-1$ zeros), $H_2 = I - 2\cdot\vec u_2\vec u_2^T$

(work in original space but append zeros)

where $\vec u_2 = \dfrac{\tilde v_2 - \|\tilde v_2\|\cdot\vec e_2}{\|\tilde v_2 - \|\tilde v_2\|\vec e_2\|}$ $\vec e_2 \in \mathbb{R}^n$ i.e. $\vec\omega_2 = \tilde v_2 - \|\tilde v_2\|\vec e_2$, $\vec u_2 = \vec\omega_2/\|\vec\omega_2\|$

either way, we know have

$$H_2\cdot(H_1 A) = \begin{bmatrix} r_{11} & r_{12} & \bullet & \bullet \\ 0 & r_{22} & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet \end{bmatrix}$$

Steps 3...n−1 follow same pattern

Friday, March 14, 2025 6:37 AM

→ we don't need step "n" since nothing left to do then

at the end of step $n-1$

we have $\underbrace{H_{n-1} \cdot H_{n-2} \cdot \ldots \cdot H_2 \cdot H_1}_{Q^T} A = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \vdots \\ 0 & & \ddots & \\ & & & r_{nn} \end{bmatrix} =: R$

So $Q = H_1^T H_2^T \cdot \ldots \cdot H_{n-1}^T$

$= \underbrace{H_1 \cdot H_2 \cdot \ldots \cdot H_{n-1}}_{\text{product of orthogonal matrices is orthogonal} \checkmark}$ since $H_i = H_i^T$

* numerically, we don't always need the matrix Q, just to be able to apply $Q^T \cdot \vec{v}$ for some $\vec{v}$, so in that case we use the $\{\vec{u}_i\}$ vectors for an efficient implementation. (Not important for our class)

Ex. 4.29 Do Householder QR on $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & -2 \\ -1 & 2 & 3 \end{bmatrix}$

Step 1: $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$, $\|\vec{v}_1\| = \sqrt{3}$, $\vec{w}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} - \sqrt{3}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\vec{u}_1 = \frac{\vec{w}_1}{\|\vec{w}_1\|}$ (messy)

$H_1 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} - 2 \cdot \vec{u}_1 \vec{u}_1^T$ so $H_1 A = \begin{bmatrix} 1.7 & \boxed{-.57} & -1.73 \\ 0 & \boxed{2.15} & 3.1 \\ 0 & \boxed{-.15} & -2.1 \end{bmatrix}$

Comment:

To multiply $H_1 \cdot A$, do not form H_1

Instead: $(\underbrace{I - 2 \cdot \vec{u}_1 \vec{u}_1^T}_{H_1}) \cdot A = A - 2 \cdot \vec{u}_1 \cdot (\underbrace{\vec{u}_1^T A}_{\text{compute this first!}})$

Step 2 $\tilde{v}_2 = \begin{bmatrix} 0 \\ 2.15 \\ -.15 \end{bmatrix}$ $\|\tilde{v}_2\| = 2.16$, $\vec{w}_2 = \begin{bmatrix} 0 \\ 2.15 \\ -.15 \end{bmatrix} - 2.16\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

$\vec{u}_2 = \vec{w}_2 / \|\vec{w}_2\|$, $H_2 = I - 2 \cdot \vec{u}_2 \vec{u}_2^T$

so $H_2 H_1 A = \begin{bmatrix} \boxed{1.7 \quad -.57 \quad -1.73} \\ 0 \quad 2.16 \quad 3.2 \\ 0 \quad 0 \quad 1.8 \end{bmatrix} = R$

Same as in $H_1 \cdot A$

$Q = H_1 \cdot H_2 = \begin{bmatrix} .57 & .61 & .53 \\ .57 & .15 & -.80 \\ -.57 & .77 & -.26 \end{bmatrix}$ Done!

Some remarks on computational complexity

Square case $m=n$

$$\underset{n\times n}{A} = Q \; R$$

Gram-Schmidt and Modified Gram-Schmidt take $O(n^3)$ time

Householder takes $O(n^3)$ time to create R and all the reflection vectors $\vec{u}_1, \dots, \vec{u}_{n-1}$

$$Q = H_1 \cdot H_2 \cdot \dots \cdot H_{n-1}, \quad H_k = I - 2\vec{u}_k \vec{u}_k^T$$

Creating Q from $\{\vec{u}_1, \dots, \vec{u}_{n-1}\}$ takes $O(n^3)$ flops too

Tall case $m>n$

(1) $$\underset{m\times n}{A} = \underset{m\times m}{Q} \; \underset{m\times n}{R}$$

Gram-Schmidt doesn't compute this directly, you'd have to do postprocessing

Householder takes $O(mn^2)$ time to create R and all the reflection vectors $\vec{u}_1, \dots, \vec{u}_{n-1}$

but creating Q from $\{\vec{u}_1, \dots, \vec{u}_{n-1}\}$ takes $O(m^2 n)$ time. Expensive!

Alternative: keep Q implicit

to multiply $Q \cdot \vec{b}$,

use $$Q\vec{b} = \underbrace{H_1 \cdot \underbrace{H_2 \cdot \dots \cdot \underbrace{H_{n-1} \cdot \vec{b}}_{O(m)}}_{O(m)}}_{O(m)}$$

Similarly for $Q^T \cdot \vec{b}$

so $O(mn)$ flops and we didn't have to spend $O(m^2 n)$ flops building Q explicitly

(2) "thin QR"

$$\underset{m\times n}{A} = \underset{m\times n}{\tilde{Q}} \; \underset{n\times n}{\tilde{R}}$$

Gram-Schmidt and Modified Gram-Schmidt take $O(mn^2)$ time

Householder doesn't give $\tilde{Q}$ but it gives $Q = \left[\, \tilde{Q} \;\vdots\; \tilde{\tilde{Q}} \,\right]$ from case (1)

so if you want $\vec{y}_1 = \tilde{Q}^T \cdot \vec{b}$

just do $\vec{y} = Q^T \vec{b} = \begin{bmatrix} \tilde{Q}^T \vec{b} \\ \tilde{\tilde{Q}}^T \vec{b} \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} \vec{y}_1 \\ \vec{y}_2 \end{bmatrix}$ so return $\vec{y}_1$

i.e. $\vec{y}_1 = y[1:n]$

so $O(mn)$ to compute $\tilde{Q}^T b$

plus $O(mn^2)$ from case (1)