QR Decomposition in Practice.

We have seen that the Gram-Schmidt algorithm can be unstable when the input vectors are close to being linearly dependent.

How can we find another method?

Definition. A Householder reflection is a matrix of the form \( P = I - 2uu^T \) where \( u \) is a vector of length 1.

\[
\begin{align*}
ux = u(u \cdot x) \\
Px = x - 2u(u \cdot x)
\end{align*}
\]
Lemma. A Householder reflection is a symmetric orthogonal matrix (and hence $P^2 = I$).

Proof.

$$P^T = (I - 2uu^T)^T$$

$$= I - 2u^Tu^T = I - 2uu^T = P.$$  

$$PP^T = (I - 2uu^T)(I - 2uu^T)$$

$$= I - 4uu^T + 4u(u^Tu)u^T$$

$$= I. \quad \square$$

Now given $\tilde{x}$, we want to find a Householder transformation $P$ so that

$$Px = \tilde{c}e_1,$$ or $P$ zeros all but 1st coord.
We can do this by writing

\[ P_x = x - 2u (u^T x) = c e_1 \]

so

\[ u = \frac{1}{2(u^T x)} (x - ce_1) \]

Thus \( u \) is a linear combination of \( x \) and \( e_1 \). Now \( \|x\|_2 = \|P_x\|_2 = \|ce_1\|_2 = |c| \), so

\[ u = \frac{\|x\|_2 e_1}{\|x + \|x\|_2 e_1\|_2} \quad \text{or} \quad u = \frac{\|x - \|x\|_2 e_1\|_2}{\|x - \|x\|_2 e_1\|_2} \]
We will choose

\[ \text{House}(x) := \begin{bmatrix} x_1 + \text{sign}(x_4) \|x\|_2 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} / \|x\|_2 \]

This defines a unit vector associated to \( x \).

We now show how to perform QR via Householder transformations. Suppose \( A \) is a general \( 5 \times 4 \) matrix.

1. Choose \( P_1 \) so \( P_1A = \begin{bmatrix} x \ x \ x \ x \\ x \ x \ x \ x \\ 0 \ x \ x \ x \\ 0 \ x \ x \ x \\ 0 \ x \ x \ x \end{bmatrix} \)

2. Let \( P_2 = \begin{bmatrix} a \\ \text{vec}(P_2) \end{bmatrix} \) \( P_2P_2A = \begin{bmatrix} x \ x \ x \ x \\ x \ x \ x \ x \\ 0 \ x \ x \ x \\ 0 \ x \ x \ x \\ 0 \ x \ x \ x \end{bmatrix} \)

   \[ \text{etc...} \]

4. \( P_4P_3P_2P_1A = \begin{bmatrix} x \ x \ x \ x \\ 0 \ x \ x \ x \\ 0 \ 0 \ x \ x \\ 0 \ 0 \ 0 \ 0 \end{bmatrix} := \tilde{R} \)
Now we know

\[ QR = A = P_4^T P_3^T P_2^T P_1^T (P_{\tilde{A}} P_{\tilde{A}}^\top)^{-1} P_{\tilde{A}} P_{\tilde{A}}^\top A \]

\[ = P_4^T P_3^T P_2^T P_1^T \tilde{R}. \]

We can't quite conclude that \( Q = P_4^T \ldots P_1^T \)

since \( Q \) is supposed to be \( m \times n \) and

the \( P_i \) are \( m \times m \). But note that if we

write

\[ P_4^T P_3^T P_2^T P_1^T = P_4^T P_3^T P_2^T P_1^T = E = \begin{bmatrix} Q & \tilde{P}_2 \\ \tilde{P}_4 & \end{bmatrix} \]

and \( \tilde{R} = \begin{bmatrix} R \\ 0 \end{bmatrix} \) then \[ Q \tilde{P} \begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} QR \end{bmatrix}. \]
so it will suffice to let

\[ Q = \text{first } n \text{ columns of } P_1 \ldots P_q \]
\[ R = \text{first } n \text{ rows of } \tilde{R} \]

In general, we have

Algorithm. (QR using Householder)

for \( i = 1 \) to \( \min(m-1,n) \)

\[ u_i = \text{House}( A(i:m,i) ) \]

\[ P_i = I - u_i u_i^T \]

\[ A(i:m,i:n) = P_i A(i:m,i:n) \]

end
What about stability?

Lemma. Let $P$ be an exact Householder transformation and $\tilde{P}$ be $fl(P)$. Then

$$fl(\tilde{PA}) = P(A+E), \quad \|E\|_2 = O(\epsilon) \cdot \|A\|_2$$

$$fl(\tilde{AP}) = (A+E)P, \quad \|F\|_2 = O(\epsilon) \cdot \|A\|_2.$$

This comes from the roundoff error analysis for dot products. Given this lemma, we can show

well, for arbitrary orthogonal matrices

Theorem. For any collection of orthogonal matrices $P_i, Q_i$ we have

$$fl(\tilde{P}_j \cdots \tilde{P}_1 A \tilde{Q}_1 \cdots \tilde{Q}_j) = P_j \cdots P_1(A+E)Q_1 \cdots Q_j$$

where $\tilde{P}_i = fl(P_i), \tilde{Q}_i = fl(Q_i)$, and $\|E\|_2 = \sum j O(\epsilon) \|A\|_2$. 
Proof. Suppose $\overline{P}_j = P_j \ldots P_1$, $\overline{Q}_j = Q_j \ldots Q_1$.

We want to show

$$A_j := fI(\tilde{P}_j A_{j-1} \tilde{Q}_j)$$

$$= \overline{P}_j (A+E_j) \overline{Q}_j$$

(where $A_j$ is the result of the iteration $\overline{P}_j$.)

For $j=0$, there's nothing to show. So suppose the result is true for $j-1$. Then

$$B = fI(\tilde{P}_j A_{j-1})$$

$$= P_j (A_{j-1} + E') \text{ by Lemma}$$

$$= P_j (\overline{P}_{j-1} (A+E_{j-1}) \overline{Q}_{j-1} + E') \text{ by induction}$$

$$= \overline{P}_j (A + E_{j-1} + \overline{P}_{j-1} E' \overline{Q}_{j-1}^T) \overline{Q}_{j-1}$$

we now want to estimate $\|E''\|$. 

We have
\[
\| E'' \|_2 = \| E_{j-1} + P_j^T E'^j Q_{j-1} \|_2 \\
\leq \| E_{j-1} \|_2 + \| E' \|_2 \\
\overset{\text{by Lemma}}{\leq} O(e) \| AA \|_2 \\
\overset{\text{induction}}{\leq} (j-1) O(e) \| AA \|_2 \\
\leq j O(e) \| AA \|_2
\]
as desired. (Estimating \( BQ_j \) is similar.) \( \square \)

To compare, suppose we had used non orthogonal matrices. Let \( X \) be any matrix and \( \tilde{X} = f_1(X) \).

\[
f_1(\tilde{X}A) = XA + E' \overset{\text{by Lemma}}{=} X(A^e + X^{-1}E) \\
= X(A + F)
\]

where \( \| E \|_2 \leq O(e) \| XX \|_2 \| AA \|_2 \), so \( \| F \|_2 \leq \| XX \|_2 \| AA \|_2 \) or \( \| F \|_2 \leq O(e) \| XX \|_2 \| X \|_2 \).
So if we multiply by a collection of such transforms, error is magnified by the product of condition numbers, which is 1 if the matrices are orthogonal.