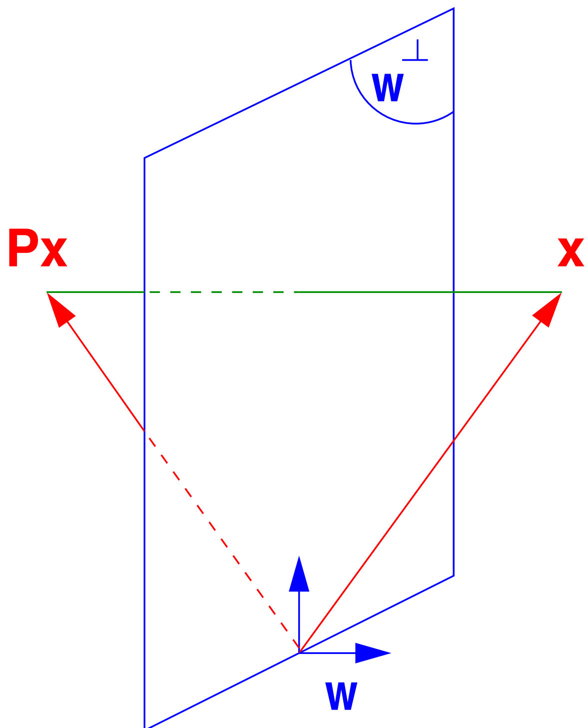
Householder QR

➤ Householder reflectors are matrices of the form

$$P = I - 2ww^T,$$

where w is a unit vector (a vector of 2-norm unity)



Geometrically, Px represents a mirror image of x with respect to the hyperplane $\mathrm{span}\{w\}^{\perp}$.

A few simple properties:

- For real w: P is symmetric – It is also orthogonal $(P^T P = I)$.
- In the complex case $P = I - 2ww^H$ is Hermitian and unitary.
- P can be written as $P = I - \beta vv^T$ with $\beta = 2/\|v\|_2^2$, where v is a multiple of w. [storage: v and β]
- Px can be evaluated $x - \beta(x^T v) \times v$ (op count?)
- Similarly: $PA = A - vz^T$ where $z^T = \beta * v^T * A$

➤ NOTE: we work in $\mathbb{R}^m$, so all vectors are of length m, P is of size $m \times m$, etc.

Problem 1: Given a vector $x \neq 0$, find w such that

$$(I - 2ww^T)x = \alpha e_1,$$

where α is a (free) scalar.

Writing $(I - \beta vv^T)x = \alpha e_1$ yields $\boxed{\beta(v^T x)\ v = x - \alpha e_1.}$

- Desired w is a multiple of $x - \alpha e_1$, i.e., we can take

$$v = x - \alpha e_1$$

- To determine α we just recall that

$$\|(I - 2ww^T)x\|_2 = \|x\|_2$$

- As a result: $|\alpha| = \|x\|_2$, or $\alpha = \pm\|x\|_2$

➤ Should verify that both signs work, i.e., that in both cases we indeed get $Px = \alpha e_1$ [exercise]

➤ Which sign is best? To reduce cancellation, the resulting $x - \alpha e_1$ should not be small. So, $\alpha = -\mathbf{sign}(\xi_1)\|x\|_2$.

$$\boxed{v = x + \mathbf{sign}(\xi_1)\|x\|_2 e_1 \text{ and } \beta = 2/\|v\|_2^2}$$

$$v = \begin{pmatrix} \hat{\xi}_1 \\ \xi_2 \\ \vdots \\ \xi_{m-1} \\ \xi_m \end{pmatrix} \quad \text{with} \quad \hat{\xi}_1 = \begin{cases} \xi_1 + \|x\|_2 & \text{if } \xi_1 > 0 \\ \xi_1 - \|x\|_2 & \text{if } \xi_1 \le 0 \end{cases}$$

➤ OK, but will yield a negative multiple of e_1 if $\xi_1 > 0$.

✍ .. Show that $(I - \beta v v^T)x = \alpha e_1$ when $v = x - \alpha e_1$ and $\alpha = \pm\|x\|_2$.

Solution: Equivalent to showing that

$$x - (\beta x^T v)v = \alpha e_1 \quad \text{i.e.,} \quad x - \alpha e_1 = (\beta x^T v)v$$

but recall that $v = x - \alpha e_1$ so we need to show that

$$\beta x^T v = 1 \quad \text{i.e., that} \quad \frac{2x^T v}{\|x - \alpha e_1\|_2^2} = 1$$

- Denominator $= \|x\|_2^2 + \alpha^2 - 2\alpha e_1^T x = 2(\|x\|_2^2 - \alpha e_1^T x)$
- Numerator $= 2x^T v = 2x^T(x - \alpha e_1) = 2(\|x\|_2^2 - \alpha x^T e_1)$

Numerator/ Denominator $= 1$. ■

Alternative:

- Define $\sigma = \sum_{i=2}^{m} \xi_i^2$.
- Always set $\hat{\xi}_1 = \xi_1 - \|x\|_2$. Update OK when $\xi_1 \leq 0$
- When $\xi_1 > 0$ compute $\hat{x}_1$ as

$$\hat{\xi}_1 = \xi_1 - \|x\|_2 = \frac{\xi_1^2 - \|x\|_2^2}{\xi_1 + \|x\|_2} = \frac{-\sigma}{\xi_1 + \|x\|_2}$$

So: $$\hat{\xi}_1 = \begin{cases} \frac{-\sigma}{\xi_1 + \|x\|_2} & \text{if } \xi_1 > 0 \\ \xi_1 - \|x\|_2 & \text{if } \xi_1 \leq 0 \end{cases}$$

- It is customary to compute a vector v such that $v_1 = 1$. So v is scaled by its first component.
- If $\sigma == 0$, wll get $v = [1; x(2:m)]$ and $\beta = 0$.

➤ Matlab function:

```
function [v,bet] = house (x)
%% computes the householder vector for x
m = length(x);
v = [1 ; x(2:m)];
sigma = v(2:m)' * v(2:m);
if (sigma == 0)
   bet = 0;
else
   xnrm = sqrt(x(1)^2 + sigma) ;
   if (x(1) <= 0)
      v(1) = x(1) - xnrm;
   else
      v(1) = -sigma / (x(1) + xnrm) ;
   end
   bet = 2 / (1+sigma/v(1)^2);
   v = v/v(1) ;
end
```

Problem 2: Generalization.

Given an $m \times n$ matrix X, find $w_1, w_2, \ldots, w_n$ such that

$$(I - 2w_n w_n^T) \cdots (I - 2w_2 w_2^T)(I - 2w_1 w_1^T)X = R$$

where $r_{ij} = 0$ for $i > j$

- First step is easy : select w_1 so that the first column of X becomes αe_1
- Second step: select w_2 so that x_2 has zeros below 2nd component.
- etc.. After $k-1$ steps: $X_k \equiv P_{k-1} \ldots P_1 X$ has the following shape:

$$
X_k = \begin{pmatrix}
x_{11} & x_{12} & x_{13} & \cdots & \cdots & \cdots & x_{1n} \\
 & x_{22} & x_{23} & \cdots & \cdots & \cdots & x_{2n} \\
 & & x_{33} & \cdots & \cdots & \cdots & x_{3n} \\
 & & & \ddots & \cdots & \cdots & \vdots \\
 & & & & x_{kk} & \cdots & \vdots \\
 & & & & x_{k+1,k} & \cdots & x_{k+1,n} \\
 & & & & \vdots & \vdots & \vdots \\
 & & & & x_{m,k} & \cdots & x_{m,n}
\end{pmatrix}.
$$

➤ To do: transform this matrix into one which is upper triangular up to the k-th column...

➤ ... while leaving the previous columns untouched.

- To leave the first $k-1$ columns unchanged w must have zeros in positions 1 through $k-1$.

$$P_k = I - 2w_k w_k^T, \quad w_k = \frac{v}{\|v\|_2},$$

where the vector v can be expressed as a Householder vector for a shorter vector using the matlab function house,

$$v = \begin{pmatrix} 0 \\ house(X(k:m,k)) \end{pmatrix}$$

- The result is that work is done on the $(k:m, k:n)$ submatrix.

ALGORITHM : 1. *Householder QR*

1. For $k = 1 : n$ do
2. $\quad [v, \beta] = house(X(k : m, k))$
3. $\quad X(k : m, k : n) = (I - \beta vv^T) X(k : m, k : n)$
4. $\quad$ If $(k < m)$
5. $\quad\quad X(k + 1 : m, k) = v(2 : m - k + 1)$
6. $\quad$ end
7. end

➤ In the end:

$$X_n = P_n P_{n-1} \dots P_1 X = \text{upper triangular}$$

Yields the factorization:

$$\boldsymbol{X = QR}$$

where

$$\boldsymbol{Q = P_1 P_2 \ldots P_n} \text{ and } \boldsymbol{R = X_n}$$

Example:

Reduce the system of vectors:

$$\boldsymbol{X = [x_1, x_2, x_3] = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 4 \end{pmatrix}}$$

Answer:

$$x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \|x_1\|_2 = 2, v_1 = \begin{pmatrix} 1+2 \\ 1 \\ 1 \\ 1 \end{pmatrix}, w_1 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 1+2 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$P_1 = I - 2w_1w_1^T = \frac{1}{6} \begin{pmatrix} -3 & -3 & -3 & -3 \\ -3 & 5 & -1 & -1 \\ -3 & -1 & 5 & -1 \\ -3 & -1 & -1 & 5 \end{pmatrix},$$

$$P_1X = \begin{pmatrix} -2 & -1 & -2 \\ 0 & 1/3 & -1 \\ 0 & -2/3 & -2 \\ 0 & -2/3 & 3 \end{pmatrix}$$

<u>Next stage:</u>

$$\tilde{x}_2 = \begin{pmatrix} 0 \\ 1/3 \\ -2/3 \\ -2/3 \end{pmatrix}, \ \|\tilde{x}_2\|_2 = 1, \ v_2 = \begin{pmatrix} 0 \\ 1/3 + 1 \\ -2/3 \\ -2/3 \end{pmatrix},$$

$$P_2 = I - \frac{2}{v_2^T v_2} v_2 v_2^T = \frac{1}{3} \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & -1 & 2 & 2 \\ 0 & 2 & 2 & -1 \\ 0 & 2 & -1 & 2 \end{pmatrix},$$

$$
P_2P_1X = \begin{pmatrix} -2 & -1 & -2 \\ 0 & -1 & 1 \\ 0 & 0 & -3 \\ 0 & 0 & 2 \end{pmatrix}
$$

Last stage:

$$
\tilde{x}_3 = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 3 \end{pmatrix}, \ \|\tilde{x}_3\|_2 = \sqrt{13}, \ v_1 = \begin{pmatrix} 0 \\ 0 \\ -2-\sqrt{13} \\ 3 \end{pmatrix},
$$

$$
P_2 = I - \frac{2}{v_3^T v_3} v_3 v_3^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -.83205 & .55470 \\ 0 & 0 & .55470 & .83205 \end{pmatrix},
$$

$$P_3P_2P_1X = \begin{pmatrix} -2 & -1 & -2 \\ 0 & -1 & 1 \\ 0 & 0 & \sqrt{13} \\ 0 & 0 & 0 \end{pmatrix} = R,$$

$$P_3P_2P_1 = \begin{pmatrix} -.50000 & -.50000 & -.50000 & -.50000 \\ -.50000 & -.50000 & .50000 & .50000 \\ .13868 & -.13868 & -.69338 & .69338 \\ -.69338 & .69338 & -.13868 & .13868 \end{pmatrix}$$

➤ So we end up with the factorization

$$\boxed{X = \underbrace{P_1P_2P_3}_{Q} R}$$

MAJOR difference with Gram-Schmidt: Q is $m \times m$ and R is $m \times n$ (same as X). The matrix R has zeros below the n-th row. Note also : this factorization always exists.

✍ Cost of Householder QR? Compare with Gram-Schmidt

Question: How to obtain $X = Q_1 R_1$ where $Q_1 =$ same size as X and R_1 is $n \times n$ (as in MGS)?

Answer: simply use the partitioning

$$X = (Q_1 \; Q_2) \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \quad \rightarrow \quad X = Q_1 R_1$$

- Referred to as the "thin" QR factorization (or "economy-size QR" factorization in matlab)
- How to solve a least-squares problem $Ax = b$ using the Householder factorization?
- Answer: no need to compute Q_1. Just apply Q^T to b.
- This entails applying the successive Householder reflections to b

The rank-deficient case

- Result of Householder QR: $\boldsymbol{Q_1}$ and $\boldsymbol{R_1}$ such that $\boldsymbol{Q_1 R_1 = X}$. In the rank-deficient case, can have $\mathbf{span}\{\boldsymbol{Q_1}\} \neq \mathbf{span}\{\boldsymbol{X}\}$ because $\boldsymbol{R_1}$ may be singular.

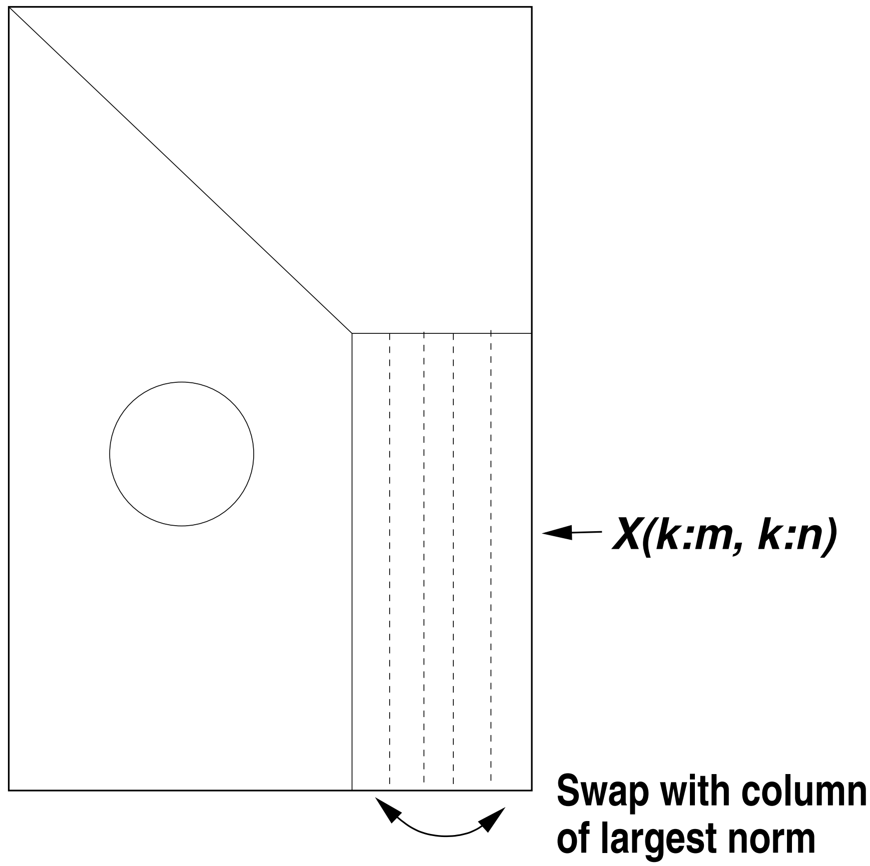
- Remedy: Householder QR with column pivoting. Result will be:

$$\boldsymbol{A\Pi} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{R_{11}} & \boldsymbol{R_{12}} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

- $\boldsymbol{R_{11}}$ is nonsingular. So rank$(\boldsymbol{X})$ = size of $\boldsymbol{R_{11}}$ = rank$(\boldsymbol{Q_1})$ and $\boldsymbol{Q_1}$ and $\boldsymbol{X}$ span the same subspace.

- $\boldsymbol{\Pi}$ permutes columns of $\boldsymbol{X}$.

Algorithm: At step k, active matrix is $X(k:m, k:n)$. Swap k-th column with column of largest 2-norm in $X(k:m, k:n)$. If all the columns have zero norm, stop.



Practical Question: How to implement this ???

✍ Suppose you know the norms of each column of X at the start. What happens to each of the norms of $X(2:m,j)$ for $j = 2,\cdots,n$? Generalize this to step k and obtain a procedure to inexpensively compute the desired norms at each step.

Properties of the QR factorization

Consider the 'thin' factorization $A = QR$, (size(Q) = [m,n] = size (A)). Assume $r_{ii} > 0$, $i = 1, \ldots, n$

1. When A is of full column rank this factorization exists and is unique

2. It satisfies:

$$\text{span}\{a_1, \cdots, a_k\} = \text{span}\{q_1, \cdots, q_k\}, \quad k = 1, \ldots, n$$

3. R is identical with the Cholesky factor G^T of $A^T A$.

➤ When A in rank-deficient and Householder with pivoting is used, then

$$Ran\{Q_1\} = Ran\{A\}$$

Givens Rotations

➤ Matrices of the form

$$G(i,k,\theta) = \begin{array}{@{}c@{}c@{}} \begin{pmatrix} 1 & \ldots & 0 & & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & \cdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \ldots & 0 & & \ldots & & 1 \end{pmatrix} & \begin{array}{c} \\ \\ i \\ \\ k \\ \\ \\ \end{array} \end{array}$$

with $c = \cos\theta$ and $s = \sin\theta$

➤ represents a rotation in the span of e_i and e_k.

Main idea of Givens rotations consider $y = Gx$ then

$$y_i = c * x_i + s * x_k$$
$$y_k = -s * x_i + c * x_k$$
$$y_j = x_j \quad \text{for} \quad j \neq i, k$$

- Can make $y_k = 0$ by selecting

$$s = x_k / t; \quad c = x_i / t; \quad t = \sqrt{x_i^2 + x_k^2}$$

- This is used to introduce zeros in the first column of a matrix A (for example $G(m-1, m)$, $G(m-2, m-1)$ etc..$G(1, 2)$)..
- See text for details

Orthogonal projectors and subspaces

Notation: Given a supspace $\mathcal{X}$ of $\mathbb{R}^m$ define

$$\mathcal{X}^\perp = \{y \mid y \perp x, \quad \forall\, x \in \mathcal{X}\}$$

➤ Let $Q = [q_1, \cdots, q_r]$ an orthonormal basis of $\mathcal{X}$

✍ How would you obtain such a basis?

➤ Then define orthogonal projector $P = QQ^T$

Properties

(a) $P^2 = P$ (b) $(I - P)^2 = I - P$
(c) $Ran(P) = \mathcal{X}$ (d) $Null(P) = \mathcal{X}^\perp$
(e) $Ran(I - P) = Null(P) = \mathcal{X}^\perp$

➤ Note that (b) means that $I - P$ is also a projector

Proof. (a), (b) are trivial

(c): Clearly $Ran(P) = \{x |\ x = QQ^T y, y \in \mathbb{R}^m\} \subseteq \mathcal{X}$. Any $x \in \mathcal{X}$ is of the form $x = Qy, y \in \mathbb{R}^m$. Take $Px = QQ^T(Qy) = Qy = x$. Since $x = Px$, $x \in Ran(P)$. So $\mathcal{X} \subseteq Ran(P)$. In the end $\mathcal{X} = Ran(P)$.

(d): $x \in \mathcal{X}^\perp \leftrightarrow (x, y) = 0, \forall y \in \mathcal{X} \leftrightarrow (x, Qz) = 0, \forall z \in \mathbb{R}^r \leftrightarrow (Q^T x, z) = 0, \forall z \in \mathbb{R}^r \leftrightarrow Q^T x = 0 \leftrightarrow QQ^T x = 0 \leftrightarrow Px = 0$.

(e): Need to show inclusion both ways.

- $x \in Null(P) \leftrightarrow Px = 0 \leftrightarrow (I - P)x = x \rightarrow x \in Ran(I - P)$
- $x \in Ran(I - P) \leftrightarrow \exists y \in \mathbb{R}^m | x = (I - P)y \rightarrow Px = P(I - P)y = 0 \rightarrow x \in Null(P)$ ■

Result: Any $x \in \mathbb{R}^m$ can be written in a unique way as

$$x = x_1 + x_2, \quad x_1 \in \mathcal{X}, \quad x_2 \in \mathcal{X}^\perp$$

- Proof: Just set $x_1 = Px, \quad x_2 = (I - P)x$
- Called the *Orthogonal Decomposition*

Orthogonal decomposition

- In other words $\mathbb{R}^m = P\mathbb{R}^m \oplus (I - P)\mathbb{R}^m$ or:

$$\mathbb{R}^m = Ran(P) \oplus Ran(I - P) \text{ or:}$$
$$\mathbb{R}^m = Ran(P) \oplus Null(P) \text{ or:}$$
$$\boxed{\mathbb{R}^m = Ran(P) \oplus Ran(P)^\perp}$$

- Can complete basis $\{q_1, \cdots, q_r\}$ into orthonormal basis of $\mathbb{R}^m$, $q_{r+1}, \cdots, q_m$

- $\{q_{r+1}, \cdots, q_m\}$ = basis of $\mathcal{X}^\perp$. $\rightarrow$ $\boxed{dim(\mathcal{X}^\perp) = m - r.}$

Four fundamental supspaces - URV decomposition

Let $A \in \mathbb{R}^{m \times n}$ and consider $\mathrm{Ran}(A)^{\perp}$

Property 1: $\mathrm{Ran}(A)^{\perp} = Null(A^T)$

Proof: $x \in \mathrm{Ran}(A)^{\perp}$ iff $(Ay, x) = 0$ for all y iff $(y, A^T x) = 0$ for all y ...

Property 2: $\mathrm{Ran}(A^T) = Null(A)^{\perp}$

- Take $\mathcal{X} = \mathrm{Ran}(A)$ in orthogonal decomoposition

- Result:

$$\mathbb{R}^m = \boldsymbol{Ran(A) \oplus Null(A^T)}$$
$$\mathbb{R}^n = \boldsymbol{Ran(A^T) \oplus Null(A)}$$

4 fundamental subspaces

$\boldsymbol{Ran(A)}$ $\boldsymbol{Null(A)}$,
$\boldsymbol{Ran(A^T)}$ $\boldsymbol{Null(A^T)}$

- Express the above with bases for $\mathbb{R}^m$:

$$[\underbrace{u_1, u_2, \cdots, u_r}_{Ran(A)}, \underbrace{u_{r+1}, u_{r+2}, \cdots, u_m}_{Null(A^T)}]$$

and for $\mathbb{R}^n$ $\quad [\underbrace{v_1, v_2, \cdots, v_r}_{Ran(A^T)}, \underbrace{v_{r+1}, v_{r+2}, \cdots, v_n}_{Null(A)}]$

- Observe $u_i^T A v_j = 0$ for $i > r$ or $j > r$. Therefore

$$U^T A V = R = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}_{m \times n} \quad C \in \mathbb{R}^{r \times r} \quad \longrightarrow$$

$$\boxed{A = URV^T}$$

- General class of URV decompositions

➤ Far from unique.

✍ Show how you can get a decomposition in which $\boldsymbol{C}$ is lower (or upper) triangular, from the above factorization.

➤ Can select decomposition so that $\boldsymbol{R}$ is upper triangular $\longrightarrow$ URV decomposition.

➤ Can select decomposition so that $\boldsymbol{R}$ is lower triangular $\longrightarrow$ ULV decomposition.

➤ SVD $=$ special case of URV where $\boldsymbol{R} =$ diagonal

✍ How can you get the ULV decomposition by using only the Householder QR factorization (possibly with pivoting)? [Hint: you must use Householder twice]