$\boxed{\text{Power method}}$ for computing eigenvalues and eigenvectors.

Let's assume $A \in \mathbb{R}^{m \times m}$ is a real, symmetric matrix.

Definition: (Rayleigh Quotient).

$$r(x) = \frac{x^T A x}{x^T x}.$$

Note that if x is an eigenvector of A, then $r(x) = \lambda$ is its corresponding eigenvalue.

Via taking derivatives, we can see that $\nabla r(x) = \dfrac{2}{x^T x} \left(Ax - r(x)x \right)$.

So, if x is an eigenvector of A, then $\nabla r(x) = 0$. And, if $\nabla r(x) = 0$ with $x \neq 0$, then x is an eigenvector of A, with eigenvalue $r(x)$.

Remark: let q_J be an eigenvector of A. We know that $\nabla r(q_J) = 0$.

By Taylor expansion, we have:

$$r(x) - r(q_J) = \nabla r(q_J)(x - q_J) + (x - q_J)^T \nabla^2 r(\xi)(x - q_J)$$

for some ξ on the line between x and q_J.

This shows that

$$|r(x) - r(q_J)| \le C \| x - q_J \|^2,$$

since $\nabla r(q_J) = 0$.

The idea is if we have a good estimate for an ~~eigenvalue~~ eigenvector, we can get a quadratically accurate estimate for the corresponding eigenvalue.

Power Method

let $v^{(0)}$ with $\|v^{(0)}\| = 1$ be some initial guess for an eigenvector of A. The power method, or power iteration, is defined as:

for $k = 1, 2, \ldots$

$$w = A v^{(k-1)} \qquad \text{apply } A$$
$$v^{(k)} = w / \|w\| \qquad \text{Normalize}$$
$$\lambda^{(k)} = (v^{(k)})^T A v^{(k)} \qquad \text{Rayleigh Quotient.}$$

The claim is that $\lambda^{(k)}$ converges to an eigenvalue of A and $v^{(k)}$ converges to the corresponding eigenvector, as $k \to \infty$. This iteration should converge to the largest eigenvalue in absolute value.

Thm: Suppose $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_m| \geq 0$

and $q_1^T v^{(0)} \neq 0$. Then the power iteration above creates iterates satisfying

$$\|v^{(k)} - (\pm q_1)\| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right), \quad |\lambda^{(k)} - \lambda_1| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right).$$

Pf: Take the 1^{st} iterate, $v^{(0)}$, and expand in the basis of eigenvectors of A:

$$v^{(0)} = a_1 q_1 + \cdots + a_m q_m .$$

$w = A v^{(k-1)}$ is actually a multiple of $A^k v^{(0)}$, up to normalizing constants c_k, so:

$$\begin{aligned} w &= c_k A^k v^{(0)} \\ &= c_k (a_1 \lambda_1^k q_1 + a_2 \lambda_2^k q_2 + \cdots + a_m \lambda_m^k q_m) \\ &= c_k a_1 \lambda_1^k \Big(q_1 + \underbrace{\frac{a_2}{a_1}\Big(\frac{\lambda_2}{\lambda_1}\Big)^k q_2 + \cdots + \frac{a_m}{a_1}\Big(\frac{\lambda_m}{\lambda_1}\Big)^k q_m}_{z_k} \Big). \\ &= c_k a_1 \lambda_1^k (q_1 + z_k). \end{aligned}$$

Now, we normalize:

Why? Because we don't know the sign of c_k. So:

$$v^{(k)} = \frac{w}{\|w\|} = (\pm)\left(\frac{q_1 + z_k}{\|q_1 + z_k\|}\right)$$

$$\begin{aligned} v^{(k)} - (\pm) q_1 &= \frac{(q_1 + z_k)}{\|q_1 + z_k\|} - q_1 \frac{\|q_1 + z_k\|}{\|q_1 + z_k\|} \\ &= \frac{(1 - \|q_1 + z_k\|)\, q_1}{\|q_1 + z_k\|} + \frac{z_k}{\|q_1 + z_k\|}. \end{aligned}$$

So: $\| v^{(k)} - q_1 \|$

$$\leq \frac{\left| 1 - \| q_1 + z_k \| \right| \| q_1 \|}{\| q_1 + z_k \|} + \frac{\| z_k \|}{\| q_1 + z_k \|}$$

$$= \left(\left| 1 - \| q_1 + z_k \| \right| + \| z_k \| \right) \frac{1}{\| q_1 + z_k \|}$$

$$= \left(\left| \| q_1 \| - \| q_1 + z_k \| \right| + \| z_k \| \right) \frac{1}{\| q_1 + z_k \|}$$

Back half of Δ-inequality $\leq \dfrac{2 \| z_k \|}{\| q_1 + z_k \|}$,

i.e. $\| v^{(k)} - q_1 \| \to 0$ as $\| z_k \| \to 0$.

Note that $\| z_k \| \to 0$ with a rate that depends on the ratio $\left(\frac{\lambda_2}{\lambda_1} \right)^k$.

The rate of convergence for the eigenvalues approximations $\lambda^{(k)}$ is given by the properties of the Rayleigh quotient.

Inverse Iteration:

What about finding other eigenvalues and eigenvectors?

Idea: "Shift" the matrix A so that we change the dominant eigenvalue:

Take $\mu \in \mathbb{R}$, μ not an eigenvalue of A.

Let $\{\lambda_j\}$ be the eigenvalues of A.

Then $\{\lambda_j - \mu\}$ are the eigenvalues of

$$(A - \mu I).$$

and $\{(\lambda_j - \mu)^{-1}\}$ are the eigenvalues of

$$(A - \mu I)^{-1}.$$

If $\mu \approx \lambda_j$, then $(\lambda_j - \mu)^{-1}$ is really big, and probably the largest eigenvalue of $(A - \mu I)^{-1}$ in magnitude.

$\leadsto$ Apply power method to $(A - \mu I)^{-1}$.

Inverse Iteration Algorithm:

$v^{(0)}$, $\|v^{(0)}\| = 1$.

for $k = 1, 2, \ldots$

Solve $(A - \mu I) w = v^{(k-1)}$ for w.

$v^{(k)} = w / \|w\|$

$\lambda^{(k)} = v^{(k)T} A v^{(k)}$.

Householder Reflectors and the QR Factorization

Recall: When we discussed solving the least squares problem, i.e. solving the normal equations, we talked about using the QR Factorization of A.

Thm 2.12: Suppose $A \in \mathbb{R}^{m \times n}$, $m \geq n$. Then $A = \hat{Q}\hat{R}$,

with $\hat{R}$ = upper triangular $n \times n$ matrix
$\hat{Q}$ = ~~xxx~~ $m \times n$ matrix

with also $\hat{Q}^T \hat{Q} = I_n$.

If $\mathrm{rank}(A) = n$, then $\hat{R}$ is invertible.

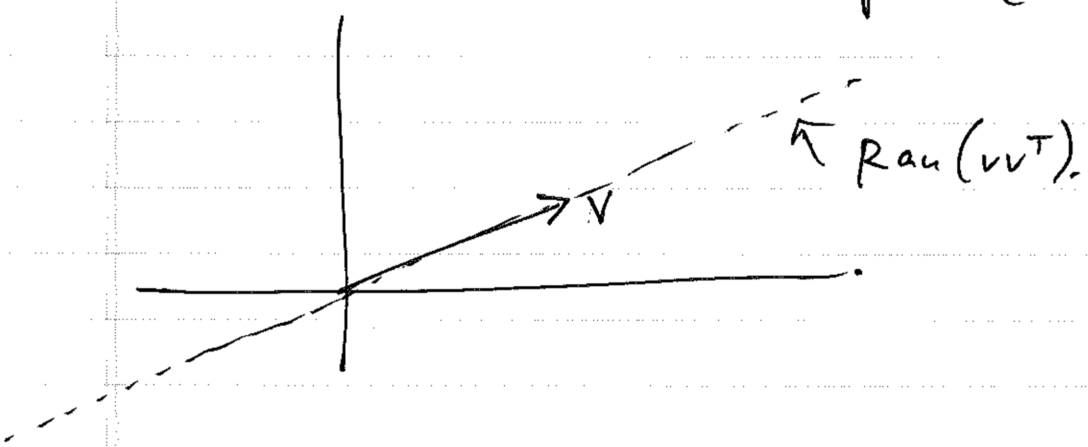
How do we explicitly build the QR factorization? One approach is with matrices called Householder reflectors.

Before getting to Householder reflectors, some review of special matrices formed by a single vector in $\mathbb{R}^n$, v, $v \neq 0$.

Example: $vv^T = v \otimes v \in \mathbb{R}^n$.

Note that $\mathrm{Ran}(vv^T) = \{ y \in \mathbb{R}^n \mid y = v^T x \text{ for } x \in \mathbb{R}^n \}$

$$= \{ \alpha v \mid \alpha \in \mathbb{R} \}.$$
$$= \mathrm{span}(v).$$

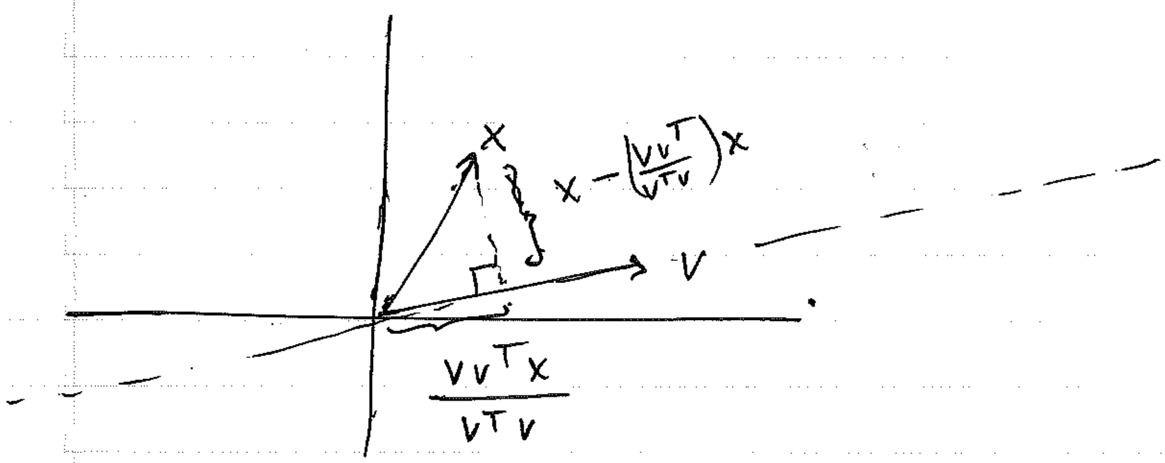


So, the Range of vv^T is the collection of all scalar multiples of v.

Example: $$\frac{vv^T}{v^Tv} = \frac{vv^T}{\|v\|^2} = \left(\frac{v}{\|v\|}\right)\left(\frac{v^T}{\|v\|}\right).$$

Note that $$\left(\frac{vv^T}{\|v\|^2}\right)\left(\frac{vv^T}{\|v\|^2}\right) = \frac{1}{\|v\|^4}\left(v \underbrace{v^Tv}_{=\|v\|^2} v^T\right)$$
$$= \frac{1}{\|v\|^2} vv^T.$$

So, the matrix $\left(\frac{vv^T}{v^Tv}\right)$ is called a projector. Geometrically, it takes a vector and projects it onto the $\text{Ran}\left(\frac{vv^T}{v^Tv}\right) = \text{span}(v)$.



let $P = \frac{vv^T}{v^Tv}$. We just showed that $P^2 = P$. Also note that P is symmetric, $P = P^T$. Then we can see:

$$
\begin{aligned}
&(Px)^T(x - Px) \\
&= x^TP^Tx - x^TP^TPx \\
&= x^TPx - x^TP^2x \\
&= x^TPx - x^TPx = 0,
\end{aligned}
$$

i.e. the vectors Px and $x - Px$ are orthogonal.

Definition: A matrix P is called an <u>orthogonal projector</u> if $P^2 = P$ and $P = P^T$.

<u>Example</u>: Take $v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

So, $P = \dfrac{vv^T}{v^Tv} = \dfrac{1}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \end{bmatrix} = \dfrac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$.

look at $P^2 = \dfrac{1}{25}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

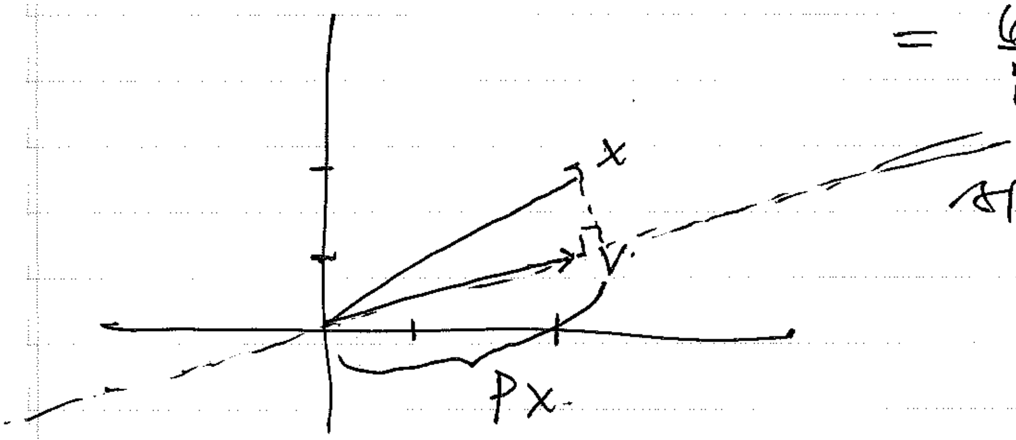
$$= \frac{1}{25}\begin{bmatrix} 20 & 10 \\ 10 & 5 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}.$$

$$= P.$$

Take $x = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

$$Px = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 12 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{12}{5} \\ \frac{6}{5} \end{bmatrix}$$

$$= \frac{6}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \text{span}(v).$$



span(v).

Let's consider another projector $\tilde{P}$:

$$\tilde{P} = I - P = I - \frac{vv^T}{v^Tv}.$$

Notice that $Ran(\tilde{P}) = \{ y \in \mathbb{R}^n \mid y = x - Px \text{ for some } x \in \mathbb{R}^n \}$.

So we have $(x - Px)^T (Pz)$

$$= \underbrace{(\tilde{P}x)^T}_{\text{arbitrary element in } Ran(\tilde{P})} \underbrace{(Pz)}_{\text{arbitrary element in the } Ran(P)}$$

$$= x^T P z - x^T P^T P z$$
$$= x^T P z - x^T P^2 z$$
$$= x^T P z - x^T P z = 0.$$

So $Ran(\tilde{P}) \perp Ran(P)$.

In other words, $\tilde{P} = I - P$ projects onto the subspace orthogonal to the $Ran(P)$.

<u>Check:</u> $\tilde{P} = I - P$ is also an orthogonal projector:

$$\tilde{P}^2 = (I-P)^2 = (I-P)(I-P)$$
$$= I - 2P + P^2$$
$$= I - 2P + P$$
$$= I - P = \tilde{P}.$$

Also, clearly $\tilde{P}^2 = \tilde{P}$.

<u>Back to our example.</u>

$$v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad P = \frac{vv^T}{v^Tv} = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}.$$

and $\tilde{P} = I - P = I - \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{4}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}.$$

$$\longrightarrow \quad \tilde{P} = I - P = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}.$$

let's look at $\operatorname{Ran}(\tilde{P}) = \{ y \in \mathbb{R}^n \mid y = \tilde{P}x, \ x \in \mathbb{R}^n \}$.

An arbitrary vector in the $\operatorname{Ran}(\tilde{P})$ takes the form:

$$\tilde{P}x = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

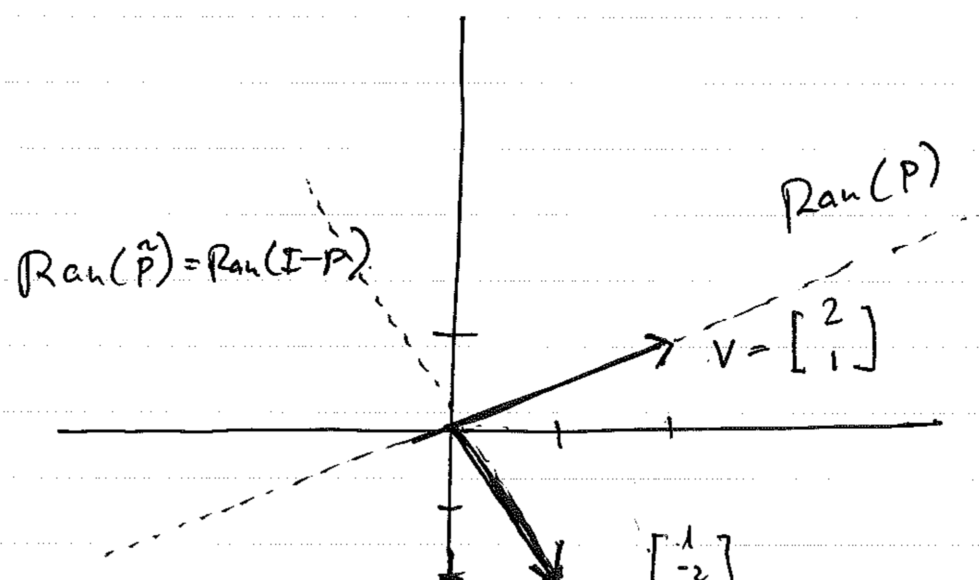
$$= \frac{1}{5}\begin{bmatrix} x_1 - 2x_2 \\ -2x_1 + 4x_2 \end{bmatrix}$$

$$= \frac{1}{5}x_1\begin{bmatrix} 1 \\ -2 \end{bmatrix} - \frac{2}{5}x_2\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$= \left(\frac{x_1}{5} - \frac{2x_2}{5} \right)\begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

$$\in \operatorname{Span}\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix} \right).$$

i.e. $\operatorname{Ran}(\tilde{P}) = \operatorname{Span}\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix} \right)$



Summary so far: For a vector $v \in \mathbb{R}^n, v \neq 0$.

$$P = \frac{vv^T}{v^Tv} \quad \text{and} \quad \tilde{P} = I - \frac{vv^T}{v^Tv}$$

are orthogonal projectors.

P takes vectors and orthogonally projects them onto $\text{span}(v)$

$\tilde{P}$ takes vectors and orthogonally projects them onto $\text{span}(v)^\perp = \begin{pmatrix}\text{subspace of} \\ \text{vectors orthogonal} \\ \text{to } v.\end{pmatrix}$.

Definition: The Householder reflector corresponding to $v \in \mathbb{R}^n$, $v \neq 0$, is the matrix $H = I - 2P = I - 2\frac{vv^T}{v^Tv}$.

Remark: H is a symmetric matrix, and it is also orthogonal, i.e.

$$H^2 = HH = (I-2P)(I-2P)$$
$$= I - 4P + 4P^2 = I - 4P + 4P = I.$$

In particular, $\|Hx\|_2^2 = \|x\|_2^2$, from your exam!

Remark: look at $v^T H x = v^T (I - 2P) x$

$$= v^T \left(I - \frac{2vv^T}{v^Tv} \right) x$$
$$= v^T x - \frac{2 v^T v v^T x}{v^T v}$$

$$= v^T x - 2v^T x = -v^T x.$$

Recall that $v^T H x = \|v\|_2 \, \|Hx\|_2 \cos\theta_1$, with θ_1 the angle between the vectors Hx and v. Furthermore:

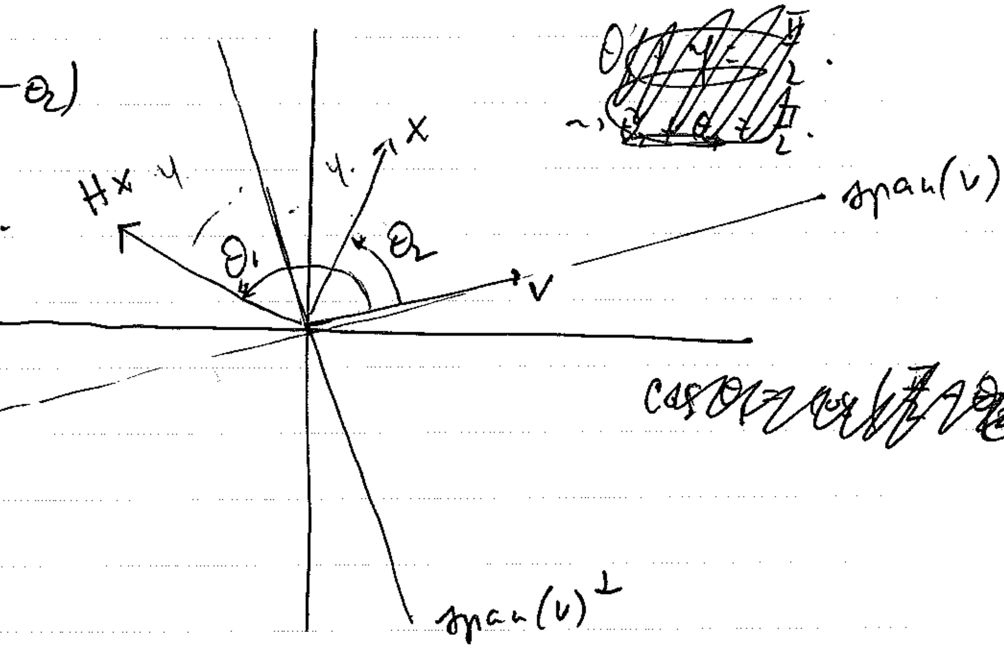
$$v^T H x = \|v\|_2 \, \|Hx\|_2 \cos\theta_1 = \|v\|_2 \, \|x\|_2 \cos\theta,$$
since $\|Hx\|_2 = \|x\|_2$.

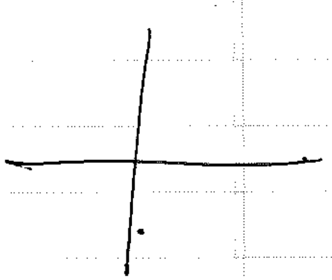
Also, we have $v^T x = \|v\|_2 \, \|x\|_2 \cos\theta_2$, with θ_2 the angle between v and x.

Since $v^T H x = -v^T x$, $\Rightarrow \cos\theta_1 = -\cos\theta_2$.

So, $-\cos\theta_2 = \cos(\pi - \theta_2)$

$\rightsquigarrow$ $\theta_1 = \pi - \theta_2$.





<u>Check:</u> $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta.$

with $\alpha = \pi$, $\beta = -\theta_2$, then:

$$\cos(\pi - \theta_2) = \cos(\pi)\cos(-\theta_2) - \underbrace{\sin(\pi)}_{0}\sin(\theta_2)$$

$$= -1\cos(-\theta_2) - 0.$$

$$= -\cos(-\theta_2) = -\cos(\theta_2).$$

since $\cos(x) = \cos(-x)$.

$$\leadsto \cos(\pi - \theta_2) = -\cos(\theta_2).$$

Lemma 5.3 Let $1 \le k < n$, let H_k be a $k \times k$ Householder matrix (reflection).

Define $H \in \mathbb{R}^{n \times n}$ as.

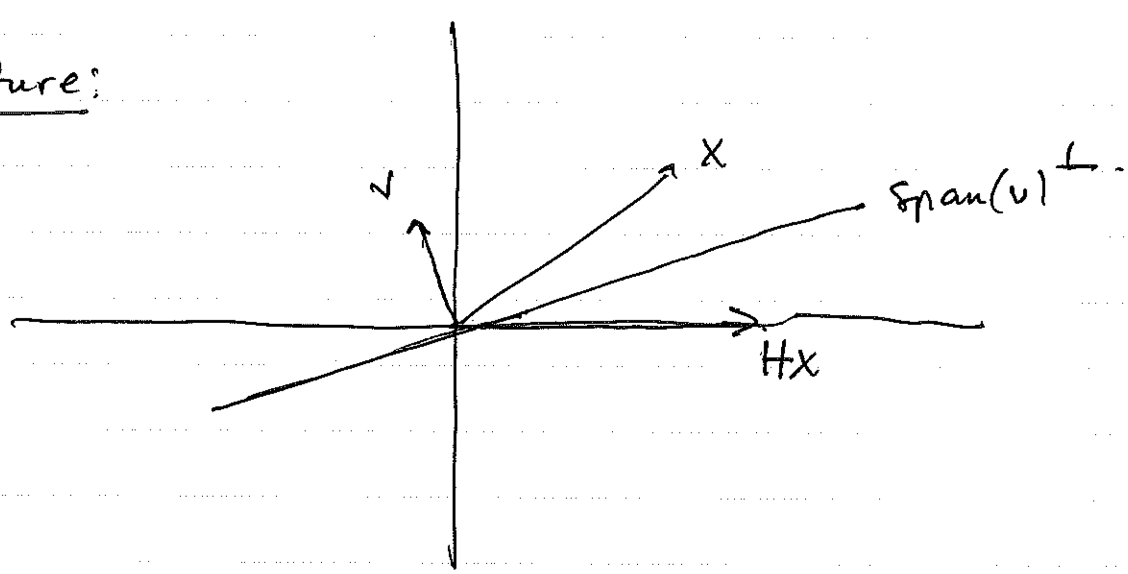
$$H = \begin{pmatrix} I_{n-k} & 0 \\ 0 & H_k \end{pmatrix}.$$

I_{n-k} is the $(n-k) \times (n-k)$ identity matrix.

Then: H is also a Householder matrix.

Lemma 5.4: Given any vector $x \in \mathbb{R}^n$, $x \neq 0$, there exists a Householder matrix H so that all the elements of Hx are zero, except the first, i.e. Hx is a nonzero multiple of e_1, the 1st column of the identity matrix.

Picture:



Pf: Idea, try to define v as a linear combination of x and e_1.

Take $v = x + ce_1$, c to be determined.

$$v^Tx = (x+ce_1)^Tx = x^Tx + ce_1^Tx = x^Tx + c\beta,$$

$$\beta = e_1^Tx.$$

$$v^Tv = (x+ce_1)^T(x+ce_1) = x^Tx + 2c\beta + c^2.$$

So then
$$Hx = x - \frac{2vv^Tx}{v^Tv}$$

$$= x - \frac{2v(x^Tx + c\beta)}{x^Tx + 2c\beta + c^2}$$

$$= \frac{(x^Tx + 2c\beta + c^2)x - 2v(x^Tx + c\beta)}{x^Tx + 2c\beta + c^2}$$

$$= \frac{(x^Tx + 2c\beta + c^2)x - 2(x + ce_1)(x^Tx + c\beta)}{x^Tx + 2c\beta + c^2}$$

$$= \frac{(x^Tx + \cancel{2c\beta} + c^2)x - 2x(x^Tx + \cancel{\beta c}) - 2ce_1(x^Tx + c\beta)}{x^Tx + 2c\beta + c^2}$$

$$= \frac{(c^2 - x^T x)x - 2c(x^T x + c\beta)e_1}{x^T x + 2c\beta + c^2}.$$

Notice that if $c^2 - x^T x = 0$, then Hx is a multiple of e_1. $\Rightarrow$ $\boxed{\text{Take } c^2 = x^T x}$

Also, we want $x^T x + 2c\beta + c^2 \neq 0$.

Recall $c^2 = x^T x$, $\beta = e_1^T x$.

By Cauchy-Schwarz, $e_1^T x \leq \|e_1\|_2 \|x\|_2$

implying $\beta^2 \leq \|x\|_2^2 = c^2$.

Then
$$\begin{aligned} x^T x + 2c\beta + c^2 &= c^2 + 2c\beta + c^2 \\ &\geq \beta^2 + 2c\beta + c^2 \\ &= (\beta + c)^2 \end{aligned}$$

We need to make sure $\beta + c \neq 0$ so that

$$x^T x + 2c\beta + c^2 \neq 0.$$

This can be done by choosing:

$$c = \begin{cases} (\operatorname{sign} \beta) \sqrt{x^T x} & \text{when } \beta \neq 0. \\ \sqrt{x^T x} & \text{when } \beta = 0. \end{cases}$$

Plugging this back in, we see that $Hx = -ce_1$.

<u>Example:</u> Compute a Householder matrix for the vector $x = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ so that Hx is <u>a multiple</u> of e_1.

$$x^T x = 1^2 + 2^2 + 2^2 = 9. \Rightarrow c^2 = 9.$$

$$\beta = e_1^T x = 1.$$

So
$$c = 3. = \begin{cases} (\text{sign}\,\beta)\sqrt{x^T x} & \text{if } \beta \neq 0 \\ \sqrt{x^T x} & \text{if } \beta = 0. \end{cases}$$

Then we define
$$v = x + c e_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 2 \end{bmatrix}.$$

<u>We compute:</u>
$$v v^T = \begin{bmatrix} 4 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 4 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 16 & 8 & 8 \\ 8 & 4 & 4 \\ 8 & 4 & 4 \end{bmatrix}$$

and $v^T v = 16 + 4 + 4 = 24$

So $H = I - 2\dfrac{vv^T}{v^Tv}$

$$= I - \frac{1}{12}\begin{bmatrix} 16 & 8 & 8 \\ 8 & 4 & 4 \\ 8 & 4 & 4 \end{bmatrix}$$

$$= I - \frac{1}{3}\begin{bmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}.$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}.$$

let's compute:

$$\frac{1}{3}\begin{bmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 12 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 2 \end{bmatrix}.$$

and then $Hx = x - \begin{bmatrix} 4 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}.$

Building the QR Factorization using Householder matrices.

Let's assume $A \in \mathbb{R}^{n \times n}$.

→ Given A, construct H_1 so that

$$H_1 A = \begin{pmatrix} * & \\ 0 & \\ | & \ast \\ 0 & \end{pmatrix},$$

i.e. $H_1 A$ has zeros in the first column, except the 1st entry.

→ Construct H_2 so that

$$H_2 (H_1 A) = \begin{pmatrix} * & * & \\ 0 & * & \\ | & 0 & \ast \\ | & | & \\ 0 & 0 & \end{pmatrix},$$

i.e. $H_2 H_1 A$ has zeros in the 1st and 2nd columns below the diagonal. Note that

$$H_2 = \begin{pmatrix} 1 & 0 \text{ --- } 0 \\ 0 & \\ | & \tilde{H}_2 \\ 0 & \end{pmatrix}$$

with $\tilde{H}_2 \in \mathbb{R}^{(n-1) \times (n-1)}$ Householder reflector.

Continue this process until

$$H_{n-1} H_{n-2} \cdots H_2 H_1 A = \begin{pmatrix} \ast & \cdots & \ast \\ & \ddots & \vdots \\ 0 & & \ast \end{pmatrix}.$$

By construction, H_i, $i = 1, \ldots, n-1$ are orthogonal, so $H_i H_i = I$.

Then we have:

$$A = \underbrace{H_1 H_2 \cdots H_{n-1}}_{Q} \underbrace{\begin{pmatrix} \ast & \cdots & \ast \\ & \ddots & \vdots \\ 0 & & \ast \end{pmatrix}}_{R}.$$