Norms on $\mathbb{R}^{n\times n}$ (Matrix Norms)

Given a norm $\|\cdot\|$ on $\mathbb{R}^n$,
the corresponding subordinate matrix norm
(or the induced matrix norm)

$$(*) \quad \|A\| := \max_{\substack{v\in\mathbb{R}^n \\ v\neq 0}} \|Av\| / \|v\|$$

$$= \max_{\substack{v\in\mathbb{R}^n \\ \|v\|=1}} \|Av\|$$

Exercise

Verify $\|\cdot\|$ as in $(*)$
is indeed a norm on $\mathbb{R}^{n\times n}$.

Matrix p-norm $(p \geqslant 1)$
subordinate to $\|\cdot\|_p$ on $\mathbb{R}^n$.

$$\|A\|_p := \max_{\substack{v\in\mathbb{R}^n \\ \|v\|_p=1}} \|Av\|_p$$

Frobenius (Euclidean) norm on $\mathbb{R}^{n\times n}$

(not subordinate to <u>any</u> norm on $\mathbb{R}^n$)

$$\|A\|_F := \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2}$$

e.g. $$\left\| \begin{bmatrix} -3 & 1 \\ +1 & 3 \end{bmatrix} \right\|_F = \sqrt{(-3)^2 + (+1)^2 + (+1)^2 + (3)^2}$$

$$= \sqrt{20}$$

Matrix 2-norm (spectral norm)

$$\|A\|_2 := \max_{\substack{v\in\mathbb{R}^n \\ \|v\|_2 = 1}} \|Av\|_2$$

e.g. $$\left\| \begin{bmatrix} -3 & 1 \\ +1 & 3 \end{bmatrix} \right\|_2$$

$$= \max \left\{ \left\| \begin{bmatrix} -3 & 1 \\ +1 & 3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right\|_2 \;\middle|\; \begin{matrix} w_1, w_2 \in \mathbb{R} \\ w_1^2 + w_2^2 = 1 \end{matrix} \right\}$$

$$= \max \left\{ \left\| \begin{bmatrix} w_2 - 3w_1 \\ 3w_2 + w_1 \end{bmatrix} \right\|_2 \;\middle|\; \begin{matrix} w_1, w_2 \in \mathbb{R} \\ w_1^2 + w_2^2 = 1 \end{matrix} \right\}$$

$$= \max \left\{ \sqrt{(w_2 - 3w_1)^2 + (3w_2 + w_1)^2} \;\middle|\; \begin{matrix} w_1, w_2 \in \mathbb{R} \\ w_1^2 + w_2^2 = 1 \end{matrix} \right\}$$

$$= \max \left\{ \sqrt{10w_1^2 + 10w_2^2} \;\middle|\; \begin{matrix} w_1, w_2 \in \mathbb{R} \\ w_1^2 + w_2^2 = 1 \end{matrix} \right\}$$

$$= \sqrt{10}$$

Characterizations of Matrix
1-, ∞-, 2-norms

THM

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

$$(+) \quad \|A\|_1 = \max_{k=1,\ldots,n} \sum_{j=1}^{n} |a_{jk}| = \max_{k=1,\ldots,n} \|a_k\|_1$$

Proof

Let $w \in \mathbb{R}^n$, $\|w\|_1 = 1$.

$$\begin{aligned} \|Aw\|_1 &= \|a_1 w_1 + a_2 w_2 + \cdots + a_n w_n\|_1 \\ &\leq |w_1|\,\|a_1\|_1 + |w_2|\,\|a_2\|_1 + \cdots + |w_n|\,\|a_n\| \\ &\leq \{|w_1| + \cdots + |w_n|\} \max_{k=1,\ldots,n} \|a_k\|_1 \\ &= \max_{k=1,\ldots,n} \|a_k\|_1 . \end{aligned}$$

Furthermore, letting $k_* := \arg\max_{k=1,\ldots,n} \|a_k\|_1$,

$$\begin{aligned} \|Ae_{k_*}\|_1 &= \|a_{k_*}\|_1 \\ &= \max_{k=1,\ldots,n} \|a_k\|_1 \end{aligned}$$

implying (+).

It can similarly be shown that

$$\|A\|_\infty = \max_{j=1,\dots,n} \sum_{k=1}^{n} |a_{jk}|.$$

Ex

$$A = \begin{bmatrix} 1 & 5 \\ -6 & 1 \end{bmatrix}$$

$$\|A\|_1 = 8 \qquad \|A\|_\infty = 7$$

2-norm of A can be characterized in terms of eigenvalues of $A^T A$.

Proposition

Let $B \in \mathbb{R}^{n \times n}$ s.t. $B^T = B$.

(i) B has n real eigenvalues $\lambda_1, \dots, \lambda_n$.

(ii) There exists an orthonormal set $\{q_1, \dots, q_n\}$ in $\mathbb{R}^n$ with q_j denoting an eigenvector corr. λ_j s.t. $q_k^T q_j = 0$ $k \neq j$ and $q_k^T q_k = 1$.

Ex

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$$

Eigenvalues are roots of

$$\det(A - \lambda I) = (2-\lambda)^2 - 16$$
$$= \lambda^2 - 4\lambda - 12$$

$$\lambda_2 = 6 \qquad \lambda_1 = -2$$

Eigenvector corr. $\lambda_1 = -2$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} q_1 = 0, \quad \text{e.g. } q_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

Eigenvector corr. $\lambda_2 = 6$

$$\begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} q_2 = 0, \quad \text{e.g. } q_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$\{q_1, q_2\}$ is orthonormal.

THM

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

$$\|A\|_2 = \lambda_n^{1/2}$$

where λ_n is the largest eigenvalue of $A^T A$.

Proof

Observe that eigenvalues of $A^T A$ are nonnegative, as

$$A^T A q = \lambda q$$

$$\Longrightarrow \lambda = \frac{q^T A^T A q}{q^T q} = \frac{\|Aq\|_2^2}{\|q\|_2^2} \geqslant 0.$$

Sort these eigenvalues from smallest to largest, that is

$$0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \dots \leqslant \lambda_n,$$

let q_j be an eigenvector corr. λ_j s.t. $\{q_1, q_2, \dots, q_n\}$ is orthonormal.

Let $w \in \mathbb{R}^n$, $\|w\|_2 = 1$. Then

$$w = \alpha_1 q_1 + \alpha_2 q_2 + \cdots + \alpha_n q_n$$

for some $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$ s.t.

$$\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2 = 1$$

(as $1 = w^T w = (\alpha_1 q_1 + \cdots + \alpha_n q_n)^T (\alpha_1 q_1 + \cdots + \alpha_n q_n)$
$= \alpha_1^2 + \cdots + \alpha_n^2$)

It follows that

$$\|Aw\|_2 = \sqrt{w^T A^T A w}$$

$$= \sqrt{(\alpha_1 q_1 + \cdots + \alpha_n q_n)^T (\alpha_1 \lambda_1 q_1 + \cdots + \alpha_n \lambda_n q_n)}$$

$$= \sqrt{\alpha_1^2 \lambda_1 + \cdots + \alpha_n^2 \lambda_n}$$

$$\leq \sqrt{(\alpha_1^2 + \cdots + \alpha_n^2)\lambda_n} = \lambda_n^{1/2}.$$

Additionally,

$$\|A q_n\|_2 = \sqrt{q_n^T A^T A q_n}$$

$$= \sqrt{q_n^T (\lambda_n q_n)} = \lambda_n^{1/2}$$

hence the desired result.

Ex

$$A = \begin{bmatrix} 2 & 5 \\ -6 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 2 & -6 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 5 \\ -6 & 1 \end{bmatrix} = \begin{bmatrix} 40 & 4 \\ 4 & 26 \end{bmatrix}$$

$$\det(A^T A - \lambda I) = (40-\lambda)(26-\lambda) - 16$$

$$= \lambda^2 - 66\lambda + 1024$$

$$\|A\|_2 = \lambda_2^{1/2} = \sqrt{\frac{66 + \sqrt{66^2 - 4 \cdot 1024}}{2}} = \sqrt{33 + \sqrt{65}}$$

Submultiplicative properties

① $\|Aw\| \leq \|A\| \, \|w\|$.

for $\underline{\text{any}}$ matrix norm $\|\cdot\|$
induced by the vector norm $\|\cdot\|$

$$\|A\| = \max_{\substack{v \in \mathbb{R}^n \\ v \neq 0}} \|Av\| / \|v\| \geq \|Aw\| / \|w\|$$

$$\Longrightarrow \|A\| \, \|w\| \geq \|Aw\|$$

② $\|AB\| \leq \|A\| \, \|B\|$

$$\|AB\| = \max_{\substack{w \in \mathbb{R}^n \\ \|w\|_2 = 1}} \|ABw\|$$

From ①

$$\leq \max_{\substack{w \in \mathbb{R}^n \\ \|w\|_2 = 1}} \|A\| \, \|Bw\|$$

$$= \|A\| \max_{\substack{w \in \mathbb{R}^n \\ \|w\|_2 = 1}} \|Bw\| = \|A\| \, \|B\|$$

Condition Numbers

$f : V \longrightarrow W$

V, W vector spaces

$\|\cdot\|_V, \|\cdot\|_W$ norms on V, W

Absolute local condition number

$$\text{Cond}_x f := \sup_{\substack{\delta x \in V \setminus \{0\} \\ x + \delta x \in D}} \frac{\| f(x+\delta x) - f(x) \|_W}{\| \delta x \|_V}$$

where D is a subset of V.

__Ex__

(1) Given $\varepsilon > 0$, $\varepsilon \approx 0$, $f(x) = \sin x$
$D = (0, \pi/2)$

$$\text{Cond}_\varepsilon \sin = \sup_{\varepsilon + h \in D} \frac{\sin(\varepsilon + h) - \sin \varepsilon}{h}$$

$$= \cos \varepsilon$$

$$\lim_{\varepsilon \to 0^+} \text{Cond}_\varepsilon \sin = 1$$

② ε as in ①, $f(x) = \sqrt{x}$

$D = (0,1)$

$$\mathrm{Cond}_{\varepsilon} f = \sup_{\varepsilon + h \in D} \frac{\sqrt{\varepsilon + h} - \sqrt{\varepsilon}}{h}$$

$$= 1/_{2\sqrt{\varepsilon}}$$

$$\lim_{\varepsilon \to 0^+} \mathrm{Cond}_{\varepsilon} f = \infty.$$

Relative local condition number

$$\mathrm{cond}_x f := \sup_{\substack{\delta x \in V \setminus \{0\} \\ x + \delta x \in D}} \frac{\|f(x+\delta x) - f(x)\|_W / \|f(x)\|_W}{\|\delta x\|_V / \|x\|_V}$$

$$= \frac{\|x\|_V}{\|f(x)\|_W} \mathrm{Cond}_x f$$

Regarding Ex ① on previous page

$$\mathrm{cond}_{\varepsilon} \sin = \frac{\varepsilon \cos \varepsilon}{\sin \varepsilon} = \frac{\varepsilon}{\tan \varepsilon}$$

$$\lim_{\varepsilon \to 0^+} \mathrm{cond}_{\varepsilon} \sin = 1$$

Regarding Ex ② above $(f(x) = \sqrt{x})$

$$\mathrm{cond}_{\varepsilon} f = \frac{\varepsilon}{\sqrt{\varepsilon}} \frac{1}{2\sqrt{\varepsilon}} = 1/2$$

$$\lim_{\varepsilon \to 0^+} \mathrm{cond}_{\varepsilon} f = 1/2$$

Condition Number for a Linear System

$$Ax = b \quad \left(\begin{array}{l} \text{assume } A \in \mathbb{R}^{n \times n} \\ \text{and invertible} \end{array} \right.$$

$$A^{-1}\cdot \; : \; b \longmapsto A^{-1} b$$

$$\text{Cond}_{\hat{b}} \; A^{-1}\cdot \; = \; \sup_{\substack{\delta b \in \mathbb{R}^n \\ \delta b \neq 0}} \frac{\| A^{-1}(\hat{b} + \delta b) - A^{-1}\hat{b} \|}{\| \delta b \|}$$

(Note: $D = \mathbb{R}^n$ is used)

$$\text{Cond}_{\hat{b}} \; A^{-1}\cdot \; = \; \sup_{\substack{\delta b \in \mathbb{R}^n \\ \delta b \neq 0}} \frac{\| A^{-1} \delta b \|}{\| \delta b \|} = \| A^{-1} \|$$

Relative local condition number

$$\text{cond}_{\hat{b}} \; A^{-1}\cdot \; = \frac{\| \hat{b} \|}{\| A^{-1} \hat{b} \|} \, \| A^{-1} \|$$

Observe

$$\| \hat{b} \| = \| A A^{-1} \hat{b} \| \leqslant \| A \| \, \| A^{-1} \hat{b} \| ,$$

so

$$\text{cond}_{\hat{b}} \; A^{-1}\cdot \; \leqslant \| A \| \, \| A^{-1} \|$$

<u>Defn</u>

$$\kappa(A) = \| A \| \, \| A^{-1} \|$$

is called the condition number of $A \in \mathbb{R}^{n \times n}$.

Ex

$$A = \begin{bmatrix} 1 & & & & 0 \\ 2 & 1 & & & \\ \vdots & 0 & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ n & 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 1 & & & & 0 \\ -2 & 1 & & & \\ \vdots & 0 & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ -n & 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$\kappa_1(A) = \|A\|_1 \, \|A^{-1}\|_1$$

$$= \left(\sum_{j=1}^{n} j \right)^2 = \left(\frac{n(n+1)}{2} \right)^2$$

$$\kappa_\infty(A) = \|A\|_\infty \, \|A^{-1}\|_\infty$$

$$= (n+1)^2$$

Exercise

Show that

$$\kappa_2(A) = \lambda_n^{1/2} \,/\, \lambda_1^{1/2}$$

where λ_n, λ_1 are largest, smallest eigenvalues of $A^T A$.

Ex

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 10^{-8} \end{bmatrix} \qquad \begin{aligned} \lambda_2 &\approx 2 \\ \lambda_1 &\approx 10^{-16}/2 \end{aligned}$$

$$\kappa_2(A) \approx 2/10^{-8} = 2 \cdot 10^{8}$$

Suppose we wish to solve

$$A x = b,$$

but because of rounding errors we solve

$$A(x+\widehat{\delta x}) = (b+\widehat{\delta b}).$$

$$\begin{pmatrix}\text{Absolute}\\ \text{error}\end{pmatrix} \frac{\|\widehat{\delta x}\|}{\|\widehat{\delta b}\|} = \frac{\|(x+\widehat{\delta x}) - x\|}{\|\widehat{\delta b}\|} = \frac{\|A^{-1}(b+\widehat{\delta b}) - A^{-1}b\|}{\|\widehat{\delta b}\|}$$

$$\leqslant \sup_{\substack{\delta b \in \mathbb{R}^n \\ \delta b \neq 0}} \frac{\|A^{-1}(b+\delta b) - A^{-1}b\|}{\|\delta b\|}$$

$$= \mathrm{Cond}_b\, A^{-1}. = \|A^{-1}\|$$

$$\begin{pmatrix}\text{Relative}\\ \text{error}\end{pmatrix} \frac{\|\widehat{\delta x}\| / \|x\|}{\|\widehat{\delta b}\| / \|b\|} \leq \frac{\|b\|}{\|x\|} \mathrm{Cond}_b\, A^{-1}.$$

$$= \mathrm{cond}_b\, A^{-1}. \leq \|A^{-1}\| \|A\|$$

CH 2 - P7

Least squares problem

$(1,1)$, $(2,4)$, $(3,8)$

Find line $\ell(t) = x_2 t + x_1$ passing through these points.

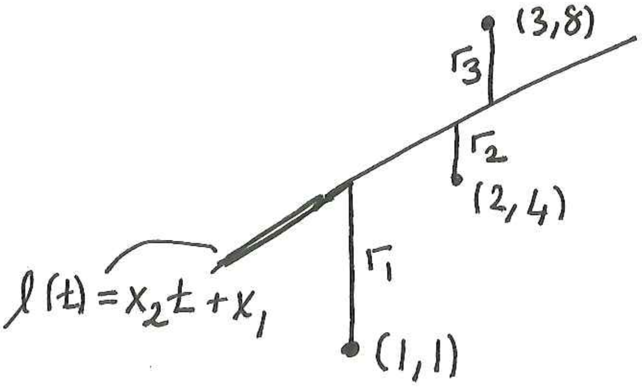
$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 + 2x_2 &= 4 \\ x_1 + 3x_2 &= 8 \end{aligned}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix}$$

No solution, no such line.

Instead

$$(*) \quad \underset{x_1, x_2 \in \mathbb{R}}{\text{minimize}} \left\| \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_2$$



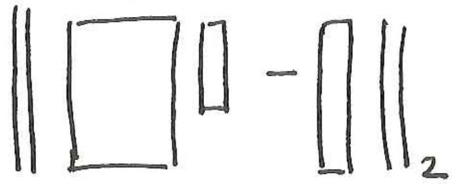
$$\underset{x_1, x_2}{\text{minimize}} \sqrt{r_1^2 + r_2^2 + r_3^2}$$

Problem statement

Given $A \in \mathbb{R}^{m \times n}$ $m > n$ and $b \in \mathbb{R}^m$

$$\min_{x \in \mathbb{R}^n} \| Ax - b \|_2$$

$\downarrow$



Find also a minimizing $x \in \mathbb{R}^n$.

An approach based on normal eqns

$$\text{objective} - \| Ax - b \|_2^2$$

$$= (Ax - b)^T (Ax - b)$$

$$f(x) = x^T A^T A x - 2 b^T A x + b^T b$$

Set $\partial f(x) / \partial x_i = 0$ $i = 1, \dots, n$

(Normal eqn) $A^T A x = 2 A^T b$

Normal eqn. for (*) on previous page

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 8 \end{bmatrix}$$

The minimizing x satisfies

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} x = \begin{bmatrix} 15 \\ 33 \end{bmatrix}$$

Not very desirable, because

(i) $\kappa_2(A^TA) = \kappa_2(A)^2$,

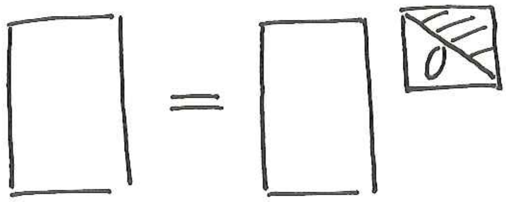
(ii) A^TA is expensive to compute.

A QR Factorization based approach

Every $A \in \mathbb{R}^{m\times n}$ with $m > n$

has a (reduced) QR factorization

$$A = \hat{Q}\hat{R}$$



where $\hat{Q} \in \mathbb{R}^{m\times n}$ s.t. $\hat{Q}^T\hat{Q} = I_n$

$\hat{R} \in \mathbb{R}^{m\times m}$ is upper triangular.

Ex

$$\begin{bmatrix} 2 & 1 \\ 2 & -5 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 2\sqrt{3} & -\sqrt{3} \\ 0 & 2\sqrt{6} \end{bmatrix}$$

THM

Let $A \in \mathbb{R}^{m \times n}$ with $m > n$ and with a (reduced) QR factorization $A = \hat{Q}\hat{R}$, let $b \in \mathbb{R}^m$.

Furthermore, assume $\text{rank}(A) = n$.

(i) $\text{rank}(\hat{R}) = m$

(ii) $\hat{x}$ s.t. $\hat{R}\hat{x} = \hat{Q}^T b$ satisfies

$$\| A\hat{x} - b \|_2 \leqslant \| Ax - b \|_2 \quad \forall x \in \mathbb{R}^n.$$

Proof

(i) $$\hat{R}\hat{x} = 0 \implies \hat{Q}\hat{R}x = Ax = 0$$

$$\implies x = 0$$

Hence, columns of $\hat{R}$ are linearly independent implying $\text{rank}(\hat{R}) = m$.

(ii) Decompose b into

$$b = b_Q + b_P$$

where $b_Q \in \text{Col}(\hat{Q}) := \text{span}\{\hat{q}_1, \dots, \hat{q}_m\}$ and $b_P \in \text{Col}(\hat{Q})^\perp$.

By the Pythagorean thm

$$\| Ax - b \|_2^2 = \| \hat{Q}\hat{R}x - b_Q \|_2^2 + \| b_P \|_2^2$$

$$\geqslant \| b_P \|_2^2 .$$

Observe that, as $b_Q \in Col(Q)$,

$$\hat{Q} z = b_Q$$
$$\Longleftrightarrow$$
$$z = \hat{Q}^T b_Q.$$

Additionally,

$$(+) \quad \hat{R}\hat{x} = \hat{Q}^T b_Q$$

is satisfied by a unique $\hat{x} \in \mathbb{R}^m$. For $\hat{x}$ satisfying (+), we must have

$$\hat{Q}\hat{R}\hat{x} = \hat{Q}^T b_Q.$$

Consequently,

$$\begin{aligned} \|A\hat{x} - b\|_2^2 &= \|\hat{Q}\hat{R}\hat{x} - b_Q\|_2^2 + \|b_P\|_2^2 \\ &= \|b_P\|_2^2 \\ &\leq \|Ax - b\|_2^2 \quad \forall x \in \mathbb{R}^n. \end{aligned}$$

The proof is completed by observing

$$\begin{aligned} \hat{Q}^T b_Q &\overset{\text{as } b_P \in Col(\hat{Q})^\perp}{=} \hat{Q}^T b_Q + \hat{Q}^T b_P \\ &= \hat{Q}^T b, \end{aligned}$$

and plugging this in (+).

$\square$

Ex

$$\min_{x \in \mathbb{R}^2} \left\| \begin{bmatrix} 2 & 1 \\ 2 & -5 \\ 2 & 1 \end{bmatrix} x - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\|_2$$

Using the QR factorization on page 3

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

$$= \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{6} \end{bmatrix}$$

$$\Longrightarrow \quad \hat{x} = \begin{bmatrix} 13/24 \\ 1/12 \end{bmatrix}$$

Proof of existence for QR factorization

The proof mimics the Gram-Schmidt procedure and by induction on the columns of A.

Base case

If $A = a_1$,

$$A = [q_1] \, \|a_1\|$$

where $q_1 = a_1/\|a_1\|$. (If $a_1 = 0$, choose q_1 an arbitrary unit vector and $\hat{R} = 0$)

Inductive case

Suppose $A = [a_1 \; a_2 \; \ldots \; a_k] \in \mathbb{R}^{m \times k}$ $(m > k)$

As the inductive hypothesis, assume

$$[a_1 \ a_2 \ \cdots \ a_{k-1}] = \hat{Q}_{k-1} \hat{R}_{k-1}$$

is a (reduced) QR factorization with $\hat{Q}_{k-1} \in \mathbb{R}^{m \times (k-1)}$ s.t. $\hat{Q}_{k-1}^T \hat{Q}_{k-1} = I$ and $\hat{R}_{k-1} \in \mathbb{R}^{(k-1)\times(k-1)}$.

We would like to show

$$(++) \quad A = [\hat{Q}_{k-1} \ q_k] \begin{bmatrix} \hat{R}_{k-1} & r \\ 0 & \alpha \end{bmatrix}$$

for some $q_k \in \mathbb{R}^m$ s.t. $q_k^T q_k = 1$, $Q_{k-1}^T q_k = 0$, as well as $r \in \mathbb{R}^{k-1}$, $\alpha \in \mathbb{R}$.

Existence of (++) is equivalent to

$$a_k = \hat{Q}_{k-1} r + \alpha q_k$$
$$Q_{k-1}^T q_k = 0$$
$$q_k^T q_k = 1.$$

Setting $r = \hat{Q}_{k-1}^T a_k$, $\alpha = \|a_k - \hat{Q}_{k-1} \hat{Q}_{k-1}^T r\|$, $q_k = (a_k - \hat{Q}_{k-1} \hat{Q}_{k-1}^T a_k)/\alpha$ all three equation above, hence (++), are satisfied provided $a_k - \hat{Q}_{k-1} \hat{Q}_{k-1}^T a_k \neq 0$.

If $a_k - \hat{Q}_{k-1} \hat{Q}_{k-1}^T a_k = 0$, choose r as above, q_k any unit vector orthogonal to $\mathrm{Col}(\hat{Q}_{k-1})$ and $\alpha = 0$ to satisfy the three equation and (++).

Now the proof follows from induction. $\square$