Orthonormal Vectors and QR Factorization

Gabriel Hugh Elkaim

Winter 2016

Gram-Schmidt Procedure (2.3)

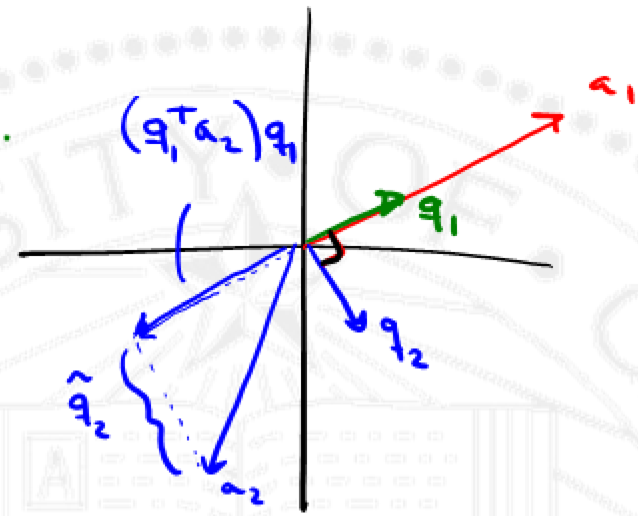
$\tilde{q}_1 := a_1$

$q_1 := \tilde{q}_1 / \|\hat{q}_1\|$ ← normalize.

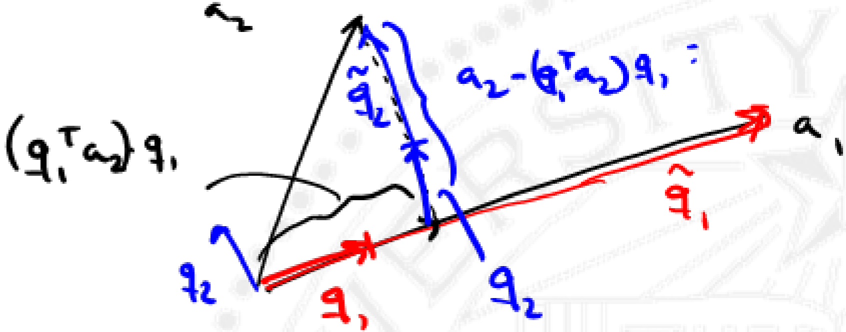
$\tilde{q}_2 := a_2 - (q_1^T a_2) \cdot q_1$

$q_2 := \tilde{q}_2 / \|q_2\|$

$q_3 := a_3 - (q_1^T a_3)\, q_1 - (q_2^T a_3)\, q_2$



Gram-Schmidt Procedure (3.3)



$$a_i = \underbrace{(q_1^T a_i)}\, q_1 + \underbrace{(q_2^T a_i)}\, q_2 + \dots + (q_{i-1}^T a_i) + \|\hat{q}_i\| a_i$$

$$r_{1i} q_1 + r_{2i} q_2 + \dots + r_{(i-1)i} q_{i-1} + r_{ii} q_i$$

$qr(A)$

QR decomposition

$A \in \mathbb{R}^{n\times k}$

$A = QR \qquad Q \in \mathbb{R}^{n\times k} \qquad R \in \mathbb{R}^{k\times k}$

upper triangular
+ on the diagonals

Invertible

R^{-1} exists

$$\underbrace{\begin{bmatrix} a_1 & \cdots & a_k \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} q_1 & \cdots & q_k \end{bmatrix}}_{Q} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ 0 & r_{22} & \cdots & r_{2k} \\ & & \ddots & \vdots \\ 0 & & & r_{kk} \end{bmatrix}$$

orthonormal basis
for span(A)

$Q^TQ = I_k$

$\mathcal{R}(Q) = \mathcal{R}(A)$

General Gram-Schmidt Procedure

$a_1 \cdots a_k$ are dependent $\quad \tilde{q}_j = 0$ for some j

a_j is linearly dependent on $(a_1 - a_{j-1})$

$r = 0$
for $i = 1 \ldots k$
{
$\tilde{a} = a_i - \sum_{j=1}^{r} q_j q_j^T a_i$
if $(\|\tilde{a}\| > eps)$ {
$r := r+1$
$q_r = \tilde{a} / \|\tilde{a}\|$
}
}

$[q_1 \cdots q_r]$ basis for $\mathcal{R}(A)$.

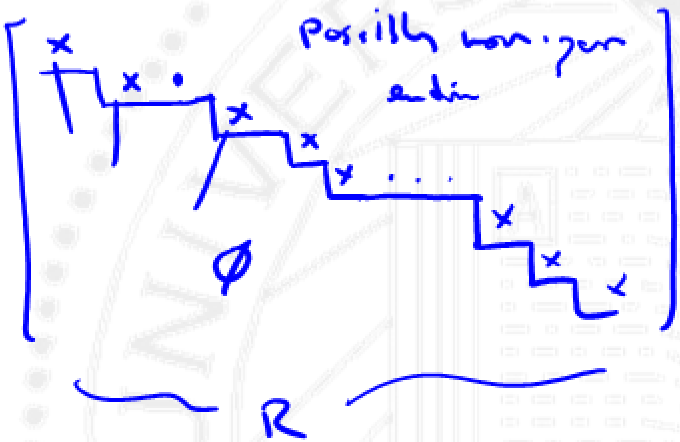
$r = \text{rank}(A)$.

General Gram-Schmidt Procedure

$[q_1 \cdots q_r]$ orthonormal basis for $\mathcal{R}(A)$

$r = \operatorname{rank}(A)$

$A = QR \quad Q^TQ = I_r \quad R \in \mathbb{R}^{r \times n}$ — $\operatorname{rank}(R) = r$



General Gram-Schmidt Procedure

$\cancel{A = QR}$

$A = Q\,[\tilde{R} \mid S]\,P \qquad Q^TQ = I_r$

$\tilde{R}$ upper triangular, $\tilde{R}^{-1}$ exists

P - permutation matrix $\in \mathbb{R}^{k\times k}$

$$AP = QR$$

rank revealing QR factorization

Applications of G-S

- directly yeilds an orthonormal basis for $\mathcal{R}(A)$
- yeilds a factorization $A = BC$ $\quad B \in \mathbb{R}^{n \times r} \quad C \in \mathbb{R}^{r \times k}$

$y = Ax$ $\qquad x \xrightarrow{n} \boxed{C} \xrightarrow{r} \boxed{B} \xrightarrow{k} y$

If I want to check if b is in span of $\{a_1 \ldots a_k\}$

$b \in \mathrm{span}\{a_1 \ldots a_k\}$ is $b \in \mathcal{R}(A)$? $\quad GS \to [a_1 \cdots a_k \; b]$

Staircase pattern in R shows which columns of A are independent from previous ones.

G-S — works incrementally on for $(a_1 \ldots a_p)$ $\quad p = 1 \ldots k$

$[a_1 \ldots a_p] = [q_1 \ldots q_p] R_p$

GS — $[a_1 \cdots a_p] \quad p = 1 \ldots k$

$$[a_1 \cdots a_p] = [q_1 \cdots q_p] R_p$$

$s = \mathrm{rank}(a_1 \cdots a_p)$ $\quad R_p$ is the leading $s \times p$

subblock of R

"Full" *QR* Factorization (1.2)

$$A = Q_1 R_1 = \left[Q_1 \,\vdots\, Q_2 \right] \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$ ← upper triangular

not unique

$[Q_1 \; Q_2]$ square orthogonal spans $\mathbb{R}^n$

$\mathcal{R}(Q_1) = \mathcal{R}(A)$ columns of $Q_2 \in \mathbb{R}^{n \times (n-r)}$ $\perp$ Q_1

any matrix $[A | \tilde{A}]$ that is full rank → $\tilde{A} = I$

$[A | I] \to GS$ — Q_1 orthonormal basis built from A

Q_2 orthonormal basis built from I

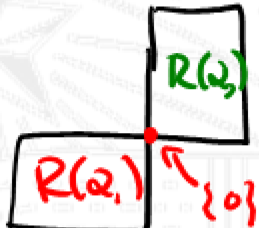
"Full" *QR* Factorization (2.2)

extend any orthogonal basis to cover all of $\mathbb{R}^n$

$\mathcal{R}(Q_1) \perp \mathcal{R}(Q_2)$ complementary subspaces

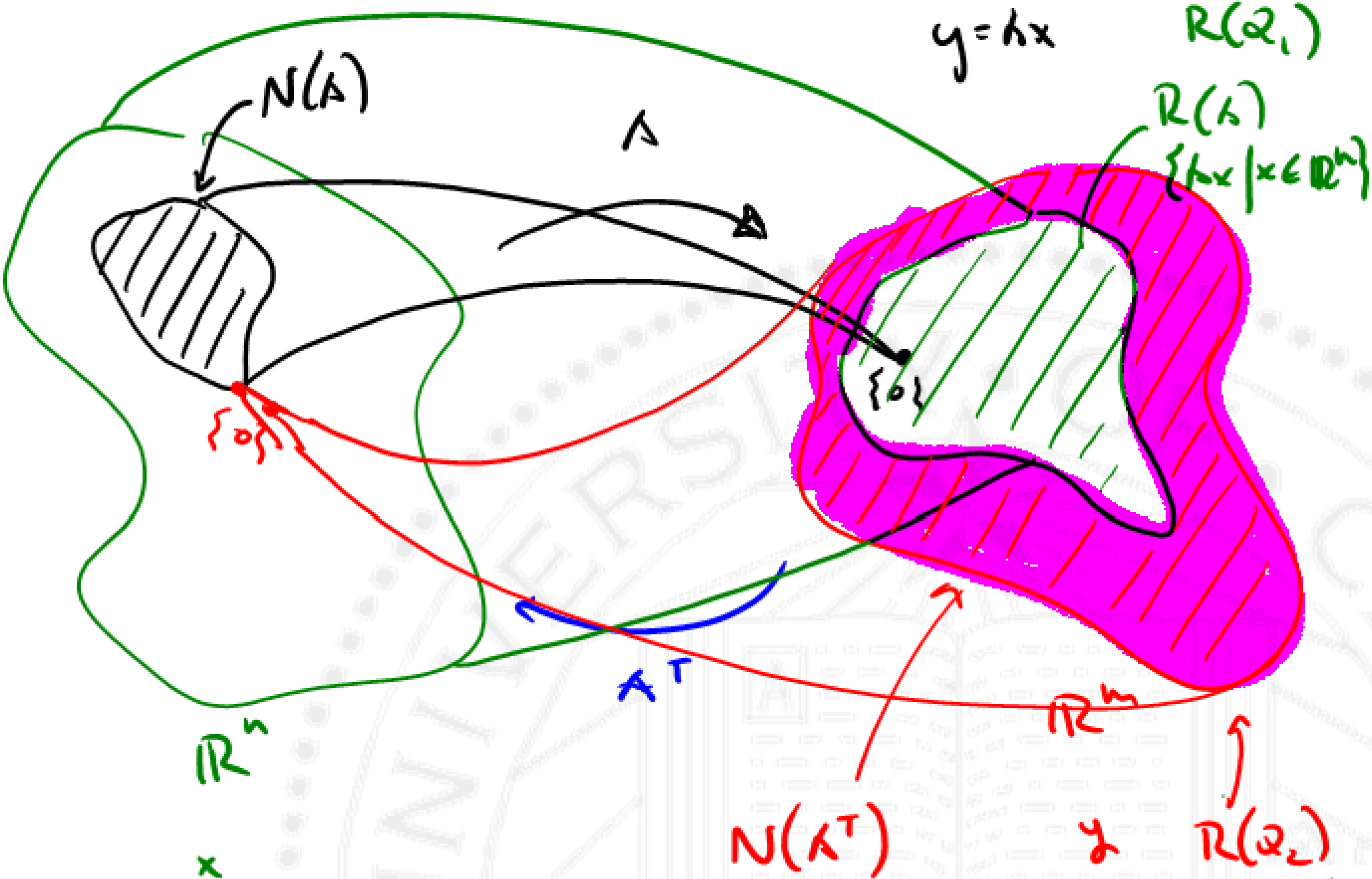
$\langle Q_1 u, Q_2 v\rangle = 0$

$\mathcal{R}(Q_1) \perp \mathcal{R}(Q_2) = \mathbb{R}^n$



$\mathcal{R}(Q_1) \overset{\perp}{+} \mathcal{R}(Q_2) = \mathbb{R}^n$

$\mathcal{R}(Q_2) = \mathcal{R}(Q_1)^\perp$ $\qquad$ $\mathcal{R}(Q_1) = \mathcal{R}(Q_2)^\perp$


y = Ax
R(Q1)
N(A)
A
R(A)
{Ax | x ∈ ℝⁿ}
{0}
{0}
Aᵀ
ℝⁿ
ℝᵐ
N(Aᵀ)
y
R(Q2)
x

Orthogonal Decomposition Induced by A

$A = [Q_1 \vdots Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ $\qquad A^T = [R_1^T \ \ 0] \begin{bmatrix} Q_1^T \\ \cdots \\ Q_2^T \end{bmatrix}$

$z \in N(A^T) \quad Q_1^T z = 0 \quad z \in R(Q_2)$

colmns of $Q_2 \to$ basis for $N(A^T)$

$0 = A^T z = [R_1^T \ \ 0] \begin{bmatrix} Q_1^T z \\ Q_2^T z \end{bmatrix} = R_1^T Q_1^T z.$ $\quad Q_1^T z.$

$(R_1^T)^{-1} 0 = (R_1^T)^{-1} A^T z = (R_1^T)^{-1} [R_1^T \ \ 0] \begin{bmatrix} Q_1^T z \\ Q_2^T z \end{bmatrix} = (R_1^T)^{-1} (R_1^T)$

$= Q_1^T z.$

Orthogonal Decomposition Induced by *A*

$\mathcal{R}(Q_2) = \mathcal{N}(A^T)$ columns of Q_2 are a basis for $\mathcal{N}(A^T)$

$\mathcal{R}(A) \perp \mathcal{N}(A^T)$ perpendicular subspaces

$\mathcal{R}(Q_1) \overset{\perp}{+} \mathcal{R}(Q_2) \in \mathbb{R}^n$

$\mathcal{R}(A)^{\perp} = \mathcal{N}(A^T)$ $\mathcal{N}(A^T)^{\perp} = \mathcal{R}(A)$

$$\boxed{\mathcal{R}(A) \overset{\perp}{+} \mathcal{N}(A^T) = \mathbb{R}^n}$$

orthogonal decomposition of $\mathbb{R}^n$ induced by A

Questions?

$$\mathcal{R}(A) \overset{\perp}{+} \mathcal{N}(A^T) = \mathbb{R}^n$$

$$\downarrow \qquad\qquad \downarrow$$

$$\mathcal{R}(Q_1) + \mathcal{R}(Q_2) = \mathbb{R}^n$$

$y = Ax \longrightarrow x = A \backslash y$ ← introduced by MATLAB

m divide ..

Least-Squares

Gabriel Hugh Elkaim

Winter 2016

Least-Squares

- Least-squares (approximate) solution of overdetermined equations
- Projection and orthagonality principle
- Least-squares estimation
- BLUE (Best Linear Unbiased Estimator) property

Overdeterimined Linear Equations

$y = Ax \quad A \in \mathbb{R}^{m \times n} \quad m > n$ strictly "skinny"

overdetermined: more equations than unknowns

for most y, I cannot solve for x.

$\mathcal{R}(A)$ is at most n

inconsistent answers

Approximately solve $y = Ax$

error $r \triangleq Ax - y$ ← discrepancy

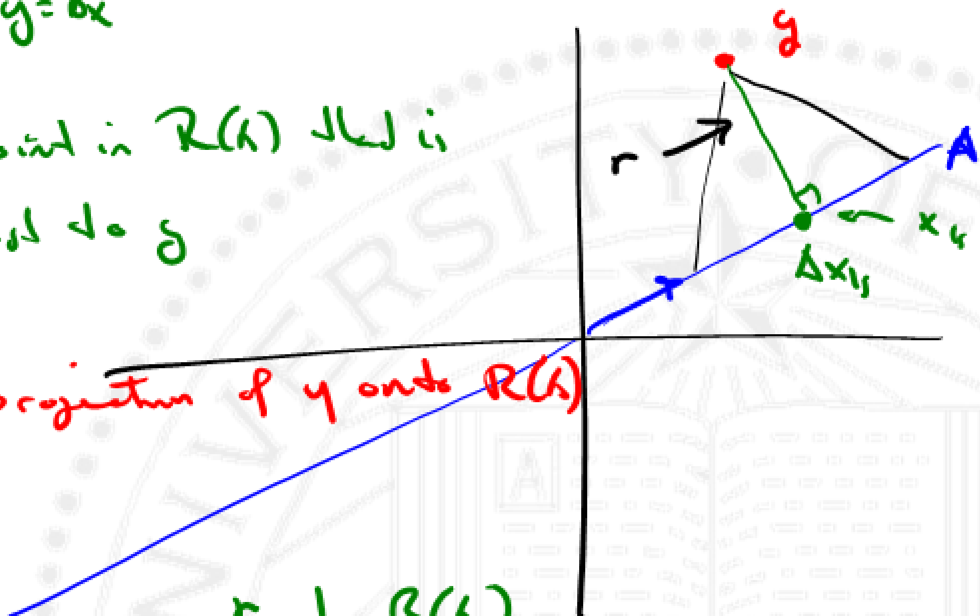
$\boxed{\text{find } x = x_{ls}, \ \min \|r\|_2}$ ← least squares

Geometric Interpretation

$A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$y = Ax$

Ax_{ls} point in $R(A)$ that is closest to y



Ax_{ls} projection of y onto $R(A)$

$r \perp R(A)$

Least-Squares (approximate) solution

Assume that A is skinny and full rank

$$\|r\|_2^2 = (y-Ax)^T(y-Ax) = x^TA^TAx - 2y^TAx + y^Ty$$

$$\frac{\partial \|r\|_2^2}{\partial x} = 0 = \cancel{2}x^TA^TA - \cancel{2}y^TA = \phi$$

$$x^TA^TA - y^TA = \phi$$

$$\left[x^TA^TA = y^TA\right]^T$$

$$[A^TA]^{-1}\underbrace{A^TAx}_{I} = [A^TA]^{-1}A^Ty$$

Normal Equation

$$\boxed{x_{ls} = \underbrace{[A^TA]^{-1}A^T}_{\text{pseudo-inverse } A^+}y}$$

Least-Squares (approximate) solution

$$x_{ls} = [A^TA]^{-1}A^T y$$

x_{ls} - linear function of y $\quad x_{ls} = By \quad B = [A^TA]^{-1}A^T$

$$A^{+} \triangleq [A^TA]^{-1}A^T$$ pseudo-inverse

A^TA small, square, invertible

A is square and invertible

then $A^{+} = A^{-1}$

$$[A^TA]^{-1}(A^T \cdot A) = A^{-1}\underbrace{A^{-T}A^T}_{I}A = A^{+} = A^{-1}$$

I

A^{+} is a left inverse of A } skinny, full rank

$$A^{+}A = I$$

Projection on $R(A)$

Ax_{ls} by definition point in $R(A)$ that is closest to y.

Ax_{ls} is the projection of y onto $R(A)$

$$Ax_{ls} = P_{R(A)}\, y = Ax_{ls} = A(A^TA)^{-1}A^T y$$

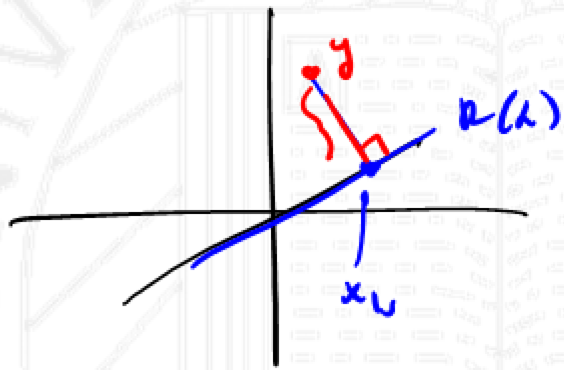
$$\text{Projection Matrix} \triangleq A(A^TA)^{-1}A^T$$

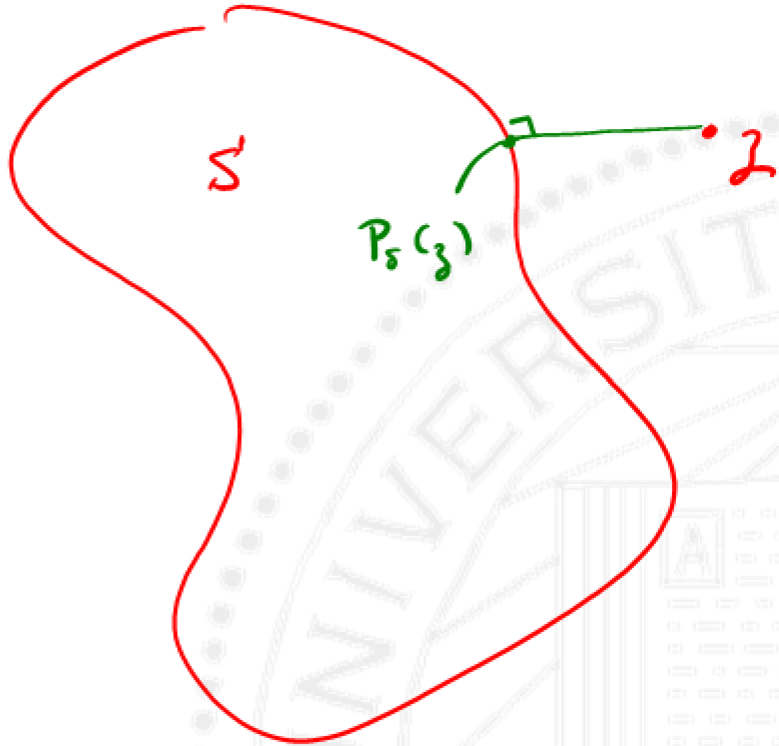
Orthogonality Principle

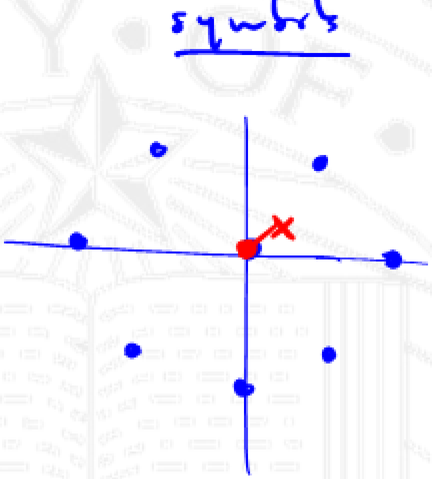
Optimal Residual

$$\|r\|_{min} = Ax_{ls} - y = \underbrace{\left[A^T(A^TA)^{-1}A - I\right]}_{\text{orthogonal to } \mathcal{R}(A)} y$$

$$\langle r, Az \rangle = y^T\left(A(A^TA)^{-1}A^T - I\right)^T Az = 0 \quad \forall\, z \in \mathbb{R}^n$$




S
z
$P_S(z)$


symbols