BLUE (wrap-up) QR decomposition

ROB 501

Necmiye Ozay

- **BLUE: Best Linear Unbiased Estimator**
- **Back substitution and QR decomposition**

Recap: on non-stochastic setting:

Underdetermined case: $y = Cx \quad x \in \mathbb{R}^n, \; y \in \mathbb{R}^m$

rows are linearly indep.

$n > m$.

Given weight matrix $S > 0$, $\|x\|_S^2 = x^T S x$

$$\left(\hat{x} := \underset{x \in \mathbb{R}^n}{\arg\min} \; \|x\|_S^2 \quad \text{s.t. } y = Cx \right) = S^{-1} C^T (C S^{-1} C^T)^{-1} y$$

Overdetermined case: $y = Cx + e$

cols are lin. indep

$x \in \mathbb{R}^n, \; y \in \mathbb{R}^m$

$m > n$

$S > 0$

$$\left(\hat{x} := \underset{e \in \mathbb{R}^m,\, x \in \mathbb{R}^n}{\arg\min} \; \|e\|_S^2 \quad \text{s.t. } y = Cx + e \right) = \underset{x \in \mathbb{R}^n}{\arg\min} \; \|y - Cx\|_S^2$$

$$= (C^T S C)^{-1} C^T S y$$

BLUE

m>n, rank(C) = n

$$y = C\,x + \varepsilon\,, \quad y \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad \varepsilon \in \mathbb{R}^m$$

Noise model: Assume ε is a R.V.

with $m = E\{\varepsilon\} = 0$, $Q = \mathrm{cov}(\varepsilon) > 0$

x is unknown constant, no probabilistic model for x.

A linear estimator: $\hat{x} = \underbrace{K}_{\mathbb{R}^{n\times m}} y$

Unbiased estimator: $E\{\hat{x}\} = x$

Best in terms of minimizing the variance

$$\mathrm{var}(\hat{x}) = E\{(\hat{x} - m_{\hat{x}})^T(\hat{x} - m_{\hat{x}})\}$$

$$E\{(\hat{x}-x)^T(\hat{x}-x)\}$$

$$\hat{x} = K(Cx + \varepsilon)$$

$$\hat{x} - x = \underbrace{KC}_{I}\,x + K\varepsilon - x$$

$$\hat{x} - x = K\varepsilon$$

Theorem: Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $y = Cx + \epsilon$, $E\{\epsilon\} = 0$, $E\{\epsilon\epsilon^\top\} =: Q > 0$, and $\mathbf{rank}(C) = n$. The Best Linear Unbiased Estimator (BLUE) is $\hat{x} = \hat{K}y$ where

$$\hat{K} = \left(C^\top Q^{-1} C\right)^{-1} C^\top Q^{-1}.$$

$$\boxed{\hat{K} = (C^T Q^{-1} C)^{-1} C^T Q^{-1}}$$

Moreover, the covariance of the error is

$$\mathrm{var}(\hat{x}) = E\{(\hat{x} - x)(\hat{x} - x)^\top\} = \left(C^\top Q^{-1} C\right)^{-1}.$$

$$\mathrm{var}(\hat{x}) = E((K\varepsilon)^T K\varepsilon) = \mathrm{tr}(K\ Q\ K^T)$$

$$\boxed{\hat{K} = \underset{K}{\mathrm{argmin}}\ \underbrace{\mathrm{tr}(KQK^T)} \quad \text{s.t.} \quad KC = I}$$

$y \in \mathbb{R}^m$

$x \in \mathbb{R}^n$

$\hat{x} = K y$

$\mathbb{R}^{n \times m}$

Proof:

k_i is ith row of K

$$K = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix}$$

$$\hat{K} = \underset{K}{\operatorname{argmin}} \; tr(KQK^T) \quad \text{s.t.} \quad KC = I$$

$$K^T = [k_1^T \,|\, k_2^T \,|\, \dots \,|\, k_n^T]$$

$$tr(KQK^T) = tr\left(\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} Q \, [k_1^T \,|\, k_2^T \,|\, \dots \,|\, k_n^T] \right)$$

$$= tr\left(\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} [Qk_1^T \,|\, Qk_2^T \,|\, \dots \,|\, Qk_n^T] \right)$$

$$= \sum_{i=1}^{n} k_i Q k_i^T = \underbrace{\boxed{\sum_{i=1}^{n} \|k_i\|_Q^2}}$$

objective function can be separated into a sum of functions of each k_i

$$\boxed{KC = I}$$
$$\Updownarrow$$
$$C^T K^T = I \iff \boxed{C^T [k_1^T \,|\, k_2^T \,|\, \dots \,|\, k_n^T] = I} = \underbrace{[C^T k_1^T \,|\, C^T k_2^T \,|\, \dots \,|\, C^T k_n^T]}$$

$[e^1 | \dots | e^n]$

$$\iff C^T k_i^T = e^i \quad \text{for all } i = 1, \dots, n$$

e^i: ith col'n of the identity matrix

→ Since both the objective function and constraints are decomposed, we can solve n independent smaller optimization problems of the form:

$$\hat{k}_i^T = \operatorname{argmin} \|k_i^T\|_Q \quad \text{s.t.} \quad C^T k_i^T = e^i \qquad \text{for } i = 1, \dots, n$$

Minimum norm solution of an undetermined set of equations:

$$\iff \hat{k}_i^T = Q^{-1} C (C^T Q^{-1} C)^{-1} e^i$$

$$\hat{K}^T = \left[k_1^T \;\vdots\; k_2^T \;\vdots\; \cdots \;\vdots\; k_n^T \right]$$

$$= \left[Q^{-1} C (C^T Q^{-1} C)^{-1} e^1 \;\vdots\; Q^{-1} C (C^T Q^{-1} C)^{-1} e^2 \;\vdots\; \cdots \;\vdots\; Q^{-1} C (C^T Q^{-1} C)^{-1} e^n \right]$$

b.c. $\left[e^1 \;\vdots\; e^2 \;\vdots\; \cdots \;\vdots\; e^n \right] = I$

$$= \left[Q^{-1} C (C^T Q^{-1} C)^{-1} \right] \cdot \cancel{I}$$

$$\boxed{\hat{K} = (C^T Q^{-1} C)^{-1} C^T Q^{-1}}$$

(b.c. Q and Q^{-1} and $C^T Q^{-1} C$ are symmetric matrices)

Error Covariance Computation for BLUE

$$
\begin{aligned}
E\{(\hat{x}-x)\,(\hat{x}-x)^\top\} &= KQK^\top \\
&= \overbrace{\boxed{\left(C^\top Q^{-1} C\right)^{-1} C^\top Q^{-1}}}^{K}\,\cancel{Q}\,\overbrace{\boxed{\cancel{Q^{-1}}C\left(C^\top Q^{-1} C\right)^{-1}}}^{K^\top} \\
&= \left(C^\top Q^{-1} C\right)^{-1} \cancel{\left[C^\top Q^{-1} C\right]\left(C^\top Q^{-1} C\right)^{-1}} \\
&= \underline{\left(C^\top Q^{-1} C\right)^{-1}}
\end{aligned}
$$

<u>Indeed</u>

$$
\begin{aligned}
\hat{x}-x &= Ky - x \\
&= KCx + K\epsilon - x \\
&= K\epsilon \text{ (because } KC = I) \\
\therefore E\{(\hat{x}-x)(\hat{x}-x)^\top\} &= E\{(K\epsilon)(K\epsilon)^\top\} \\
&= E\{K\epsilon\epsilon^\top K^\top\} \\
&= KQK^\top
\end{aligned}
$$

Back substitution

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

$$\boxed{Ax=b}$$

$$x = A^{-1}b$$

$$\boxed{\begin{aligned} 2x_1 + 2x_2 + x_3 &= 10 \\ x_2 - x_3 &= 2 \\ \boxed{4x_3} &\boxed{= 12} \end{aligned}}$$

③ $x_1 = -3/2$ $\leftarrow$ $2x_1 + 10 + 3 = 10$

② $x_2 = 5$ $\leftarrow$ $x_2 - 3 = 2$

① $x_3 = 3$

upper triangular

$$\begin{bmatrix} 2 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 12 \end{bmatrix}$$

$$\boxed{Rx = r}$$

Proposition: Let A be an $m \times n$ matrix w/ linearly independent cols. Then, $\exists$ $m \times n$ matrix Q w/ orthonormal cols and an upper triangular $n \times n$ matrix R s.t. $A = QR$.

This is QR decomposition.

Notes: 1) $Q^T Q = I_{n\times n}$

2) $[R]_{ij} = 0 \quad i > j$

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

3) Col's A are lin indep $\iff$ R invertible.

Why do we care about QR factorization?

1) Suppose $Ax = b$ is overdetermined, col's A are linearly independent. $A = \underline{QR}$ and consider:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|Ax - b\|$$

from normal equations: $\boxed{\underbrace{A^T A}\, \hat{x} = \underbrace{A^T b}}$ *

$$A^T A = R^T \underbrace{Q^T Q}_{I} R = R^T R$$

$$A^T b = R^T Q^T b$$

$$A^T A \hat{x} = A^T b \iff R^T R \hat{x} = R^T Q^T b$$

b.c. R is invertible R^T is invertible and we can multiply both sides with $(R^T)^{-1}$

$$\iff \boxed{R\hat{x} = Q^T b} \; *$$

R is upper triangular

$\rightarrow$ We can solve for $\hat{x}$ by back substitution.

2) Suppose A is square and invertible.

Write $A = QR \implies A^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T$

Question 1: How can we invert an upper triangular matrix?

$$R R^{-1} = I$$

let $[R^{-1}]_i$ be the ith col of R^{-1}. Then

$$R \underbrace{[R^{-1}]_i}_{\text{solve by back substitution}} = e^i \qquad i = 1, \dots, n \quad (\text{assume } R \in \mathbb{R}^{n \times n})$$

R: upper triangular matrix

e^i: ith col of I ($\hat{=}$ ith standard basis element)

(MATLAB \ command uses QR decomposition e.g. $A \backslash B = A^{-1}B$)

3) Suppose $Ax=b$ is underdetermined w/ the rows of A lin. indep. $\Leftrightarrow$ cols of A^T lin. indep.

min norm solution $\hat{x} = A^T(AA^T)^{-1}b$

$(MN)^{-1} = N^{-1}M^{-1}$ if M,N square invertible

Let $A^T = QR$

$\downarrow$

QR decomp. of A^T

$$AA^T = R^TQ^TQR = R^TR$$

$$\begin{aligned}\hat{x} &= QR(R^TR)^{-1}b \\ &= Q\cancel{R}\cancel{R^{-1}}(R^T)^{-1}b \\ &= Q(R^T)^{-1}b\end{aligned}$$

The above examples show that QR factorization ($\Leftrightarrow$ decomposition) simplifies many different operations.

How to compute QR factorization?

$A = [A_1 \vdots A_2 \vdots \cdots \vdots A_n]$, $A_i \in \mathbb{R}^m$

G-S w/ normalization: $\{A_1, \dots, A_n\} \to \{v^1, v^2, \dots, v^n\}$

by $v^1 = \dfrac{A_1}{\|A_1\|}$

$v^2 = A_2 - \langle A_2, v^1\rangle \cdot v^1$, $v^2 = \dfrac{v^2}{\|v^2\|}$

$\vdots$

$\forall\ 1 \leq k \leq n \quad \text{span}\{A_1, \dots, A_k\} = \text{span}\{v^1, \dots, v^k\}$

$Q = [v^1 \mid v^2 \mid \dots \mid v^k]$, where $[Q^T Q]_{ij} = \langle v^i, v^j \rangle$

$$= \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$Q^T Q = I$

What about R? (we want $A = QR$)

$A_i \in \text{span}\{\underbrace{v^1, \dots, v^i}\}$

for $k \in \{1, \dots, i\}$

$(v^k)^T A_i = \underline{\langle A_i, v^k \rangle}$

$= \underline{\underline{\alpha_k}}$

$\rightarrow A_i = \overset{(v^k)^T}{}\ \underbrace{\alpha_1 v^1} + \dots + \underbrace{\alpha_i v^i}$

$A_i = \underline{\langle A_i, v^1 \rangle} v^1 + \underline{\langle A_i, v^2 \rangle} v^2 + \dots + \underline{\langle A_i, v^i \rangle} v^i$

$$[A_i]_V = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_i \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \langle A_i, v^1 \rangle \\ \langle A_i, v^2 \rangle \\ \vdots \\ \langle A_i, v^i \rangle \\ 0 \\ \vdots \\ 0 \end{bmatrix} = R_i$$

$\therefore \boxed{A_i = Q R_i}$

$\Updownarrow$

$[A_1 \mid \dots \mid A_n] = [v^1 \mid v^2 \mid \dots \mid v^n][R_1 \mid R_2 \mid \dots \mid R_n]$ $QR = A$

$A_i = [v^1 \mid v^2 \mid \dots \mid v^n] R_i$

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

$Q \in \mathbb{R}^{m \times n}$

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

$A_i \in \text{span}\{\underbrace{v^1, \dots, v^i}\} \rightarrow$ guarantees that $[R_i]_j = 0$ if $j > i$.

Recall Gram-Schmidt

Gram-Schmidt Process: Two steps: orthogonalize, then normalize.

Let $\{y_i \mid i = 1, ..., n\}$ be a *linearly independent* set of vectors. Define a set of vectors $\{v_i \mid i = 1, ..., n\}$ by:

$$
\begin{aligned}
v_1 &= y_1 \\
v_2 &= y_2 - a_{21} v_1,
\end{aligned}
$$

and **choose** a_{21} so that $< v_1, v_2 >= 0$.

$$
\begin{aligned}
\therefore 0 =< v_1, v_2 > &=< v_1, y_2 - a_{21} v_1 > \\
&=< v_1, y_2 > - a_{21} < v_1, v_1 > \\
\therefore a_{21} &= \frac{< v_1, y_2 >}{\|v_1\|^2}
\end{aligned}
$$

Write $v_3 = y_3 - a_{31} v_1 - a_{32} v_2$

$$
\begin{aligned}
\longrightarrow \quad 0 &=< v_1, v_3 >=< v_1, y_3 - a_{31} v_1 - a_{32} v_2 > \\
&=< v_1, y_3 > - a_{31} < v_1, v_1 > - a_{32} \underbrace{< v_1, v_2 >}_{=0} \\
\therefore a_{31} &= \frac{< v_1, y_3 >}{\|v_1\|^2} \\
\longrightarrow \quad 0 &=< v_2, v_3 >=< v_2, y_3 - a_{31} v_1 - a_{32} v_2 > \\
&=< v_2, v_3 >=< v_2, y_3 > - a_{31} \underbrace{< v_2, v_1 >}_{=0} - a_{32} < v_2, v_2 > \\
\therefore a_{32} &= \frac{< v_2, y_3 >}{\|v_2\|^2}
\end{aligned}
$$

In general, one obtains:

$$
v_k = y_k - \sum_{j=1}^{k-1} \underbrace{\frac{< v_j, y_k >}{\|v_j\|^2}}_{a_{kj}} \cdot v_j .
$$

Now, $\{v_k \mid k = 1, ..., n\}$ is an **orthogonal** set.

Define: $\tilde{v}_i = \dfrac{v_i}{\|v_i\|} \Rightarrow \{\tilde{v}_i \mid i = 1, ..., n\}$ is **orthonormal**.

$$\left(\hat{x} := \underset{x \in \mathbb{R}^n}{\arg\min} \; \|x\|^2_S \;\; \text{s.t.} \; y = Cx \right) = S^{-1}C^T(CS^{-1}C^T)^{-1}y$$

cols are lin. indep

Min weighted norm solution to $y = Cx$ where C is full row rank.

$$\left(\hat{k}_i^T = \arg\min \|k_i^T\|_Q \;\; \text{s.t.} \; \boxed{C^T} k_i^T = \boxed{e^i} \right)$$

A subproblem for solving BLUE.

In BLUE, we assumed C is full col. rank

→ C^T is full row rank

$$\hat{k}_i^T = Q^{-1}C(C^TQ^{-1}C)^{-1}e^i$$

$$\hat{k}_i^T = Q^{-1}C(C^TQ^{-1}C)^{-1}e^i$$

Min weighted norm solution to $e^i = C^T k_i^T$ where C^T is full row rank.

$Ax = b$ has an exact solution
iff $b \in \text{col span}(A)$

if $A \in \mathbb{R}^{m \times n}$ with $m < n$ (underdetermined case):

- col's of A are always linearly dependent (because we have n col's in $\mathbb{R}^m$ with $m < n$)

- If rows of A are linearly independent, then this implies $\text{colspan}(A) = \mathbb{R}^m$
  this also implies $b \in \text{colspan}(A)$ is trivially satisfied.

$$y = C x$$

sensor model

the quantity we want to estimate

measurements