MA 1018: Introduction to Scientific Computing
Solution of Linear Least-Squares Problem

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: January 21, 2025/Last updated: February 15, 2026

The linear least-squares problem

- Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m > n$, and $\boldsymbol{b} \in \mathbb{R}^m$. Then the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is called overdetermined. It has more equations than unknowns. *In general, such a system has no solution.*
- The least-squares problem is to find $\boldsymbol{x} \in \mathbb{R}^n$ that solves the minimization problem:

$$\boxed{\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{b} - \boldsymbol{Ax}\|_2. \qquad (\star)}$$

  As the unknown $\boldsymbol{x}$ occurs linearly, this is also referred to as *the linear least-squares problem.*
- Let $\boldsymbol{A} = [\boldsymbol{a}_{\cdot 1} \;\; \boldsymbol{a}_{\cdot 2} \;\; \cdots \;\; \boldsymbol{a}_{\cdot n}]$ and $\boldsymbol{r} := \boldsymbol{b} - \boldsymbol{Ax}$ (residual vector). Since the columns of $\boldsymbol{A}$ span a hyperplane, we get the solution by making $\boldsymbol{r}$ orthogonal to the columns of $\boldsymbol{A}$,

$$\boldsymbol{r}^\top \boldsymbol{a}_{\cdot j} = 0, \quad \forall\, j = 1, 2, \cdots, n.$$

The linear least-squares problem (cont'd)

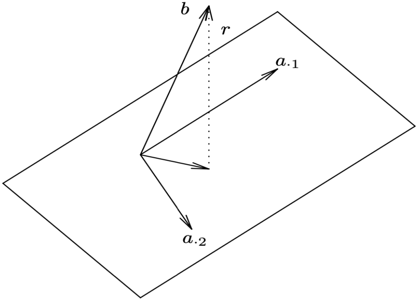
Therefore, we have

$$\mathbf{0} = \boldsymbol{r}^\top [\boldsymbol{a}_{\cdot 1} \;\; \boldsymbol{a}_{\cdot 2} \;\; \cdots \;\; \boldsymbol{a}_{\cdot n}] = \boldsymbol{r}^\top \boldsymbol{A},$$

which implies the following *normal equations:*

$$\mathbf{0} = \boldsymbol{A}^\top \boldsymbol{r} = \boldsymbol{A}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}^\top \boldsymbol{b} - \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}$$

$$\Leftrightarrow \quad \boxed{\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{b}.}$$



Uniqueness of solution of the normal equations

Theorem: *If the column vectors of A are linearly independent, then $A^\top A$ is nonsingular and the normal equations $A^\top A x = A^\top b$ have a unique solution $\widehat{x}$, which is the solution of the linear least-squares problem $(\star)$.*

Proof:

(1) *Show that the symmetric matrix $A^\top A$ is positive definite.*
Let $0 \neq x \in \mathbb{R}^n$. Let $y = Ax$. Since the column vectors of A are linearly independent, we have $y \neq 0$ and

$$x^\top (A^\top A) x = x^\top A^\top A x = y^\top y = \|y\|_2^2 > 0.$$

Hence, $A^\top A$ is nonsingular and $A^\top A x = A^\top b$ has a unique solution $\widehat{x}$.

(2) *Let $\widehat{r} := b - A\widehat{x}$. We want to show that $\|\widehat{r}\|_2 \leq \|r\|_2$.*
Since $r = b - Ax = b - A\widehat{x} + A(\widehat{x} - x) = \widehat{r} + A(\widehat{x} - x)$, we have

$$\begin{aligned}
\|r\|_2^2 &= (\widehat{r} + A(\widehat{x} - x))^\top (\widehat{r} + A(\widehat{x} - x)) \\
&= \widehat{r}^\top \widehat{r} + \widehat{r}^\top A(\widehat{x} - x) + (\widehat{x} - x)^\top A^\top \widehat{r} + (\widehat{x} - x)^\top A^\top A(\widehat{x} - x).
\end{aligned}$$

Since $A^\top \widehat{r} = A^\top (b - A\widehat{x}) = 0$, we obtain

$$\|r\|_2^2 = \widehat{r}^\top \widehat{r} + (\widehat{x} - x)^\top A^\top A(\widehat{x} - x) = \|\widehat{r}\|_2^2 + \|A(\widehat{x} - x)\|_2^2 \geq \|\widehat{r}\|_2^2.$$

Uniqueness of solution of the linear least-squares problem

Theorem: *If the column vectors of $\boldsymbol{A}$ are linearly independent, then the linear least-squares problem $(\star)$ has a unique solution $\widehat{\boldsymbol{x}}$.*

Proof:
From the previous theorem, the unique solution $\widehat{\boldsymbol{x}}$ of the normal equations, i.e., $\boldsymbol{A}^\top \boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{b}$, is also a solution of the linear least-squares problem $(\star)$.

Next, we want to show that $\widehat{\boldsymbol{x}}$ is the only critical point of the function $f(\boldsymbol{x}) := \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2$, which implies that the linear least-squares problem $(\star)$ has this unique solution $\widehat{\boldsymbol{x}}$. Let $\nabla f(\boldsymbol{x}) = \mathbf{0}$. Then we have

$$\begin{aligned}
\mathbf{0} &= \nabla f(\boldsymbol{x}) = \nabla\big((\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\big) = \nabla\big((\boldsymbol{b}^\top - \boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})\big)\\
&= \nabla\big(\boldsymbol{b}^\top\boldsymbol{b} - \boldsymbol{b}^\top\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{b} + \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\big)\\
&= \nabla\big(\boldsymbol{b}^\top\boldsymbol{b} - 2\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{b} + \boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x}\big)\\
&= -2A^\top\boldsymbol{b} + 2A^\top A\boldsymbol{x}.
\end{aligned}$$

It leads to $\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top\boldsymbol{b}$. Since $\boldsymbol{A}^\top\boldsymbol{A}$ is nonsingular, $\widehat{\boldsymbol{x}}$ is the only critical point. Hence, the linear least-squares problem $(\star)$ has a unique solution $\widehat{\boldsymbol{x}}$.

Singular value decomposition

The linear least-squares problem can be solved using the SVD:

Singular value decomposition (SVD): *Let $A \in \mathbb{R}^{m\times n}$. Then we have*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top := \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m \end{bmatrix}_{m\times m} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_n \end{bmatrix}_{n\times n}^\top,$$

where $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *are orthogonal matrices,*

$$\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_{m\times m}, \quad \boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}_{n\times n},$$

$\boldsymbol{\Sigma} = diag(\sigma_1, \ldots, \sigma_r, 0 \ldots, 0) \in \mathbb{R}^{m\times n}$ *with*

$$\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}}$$

is a diagonal matrix of singular values, and $r = \operatorname{rank}(\boldsymbol{A})$.

Solving the linear least-squares problem using SVD

Let $A \in \mathbb{R}^{m \times n}$, $m > n$, and $b \in \mathbb{R}^m$. Assume that A has full column rank. Consider the overdetermined linear system $Ax = b$. By the SVD, we write

$$A = (U_1 \;\; U_2) \begin{bmatrix} S \\ 0 \end{bmatrix} V^\top,$$

where $U_1 \in \mathbb{R}^{m \times n}$. Then

$$\|r\|_2^2 = \|b - Ax\|_2^2 = \|b - U \begin{bmatrix} S \\ 0 \end{bmatrix} V^\top x\|_2^2.$$

Let $\widetilde{b} = U^\top b$, with $\widetilde{b}_i = U_i^\top b$, $i = 1, 2$, and $y = V^\top x$. Since the 2-norm is invariant under orthogonal transformations, we have

$$\|r\|_2^2 = \|U\Big(U^\top b - \begin{bmatrix} S \\ 0 \end{bmatrix} V^\top x\Big)\|_2^2 = \| \begin{bmatrix} \widetilde{b}_1 \\ \widetilde{b}_2 \end{bmatrix} - \begin{bmatrix} S \\ 0 \end{bmatrix} y\|_2^2.$$

Therefore,

$$\|r\|_2^2 = \|\widetilde{b}_1 - Sy\|_2^2 + \|\widetilde{b}_2\|_2^2.$$

Solving the linear least-squares problems using SVD (cont'd)

The minimum occurs when $\boldsymbol{y} = \boldsymbol{S}^{-1}\widetilde{\boldsymbol{b}}_1$. Thus, the least-squares solution is given by

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{y} = \boldsymbol{V}\boldsymbol{S}^{-1}\widetilde{\boldsymbol{b}}_1 = \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U}_1^\top \boldsymbol{b}.$$

Since

$$\boldsymbol{S}^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \cdots, \frac{1}{\sigma_n}\right), \quad (\mathrm{rank}(\boldsymbol{A}) = n)$$

the least-squares solution can be written

$$\boxed{\boldsymbol{x} = \sum_{i=1}^{n} \frac{\boldsymbol{u}_i^\top \boldsymbol{b}}{\sigma_i} \boldsymbol{v}_i.}$$

Note that the assumption that $\boldsymbol{A}$ has full column rank implies $\sigma_i > 0$ for $i = 1, 2, \cdots, n$.

Plane rotations

- A plane rotation matrix is *orthogonal:*

$$\boldsymbol{G} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \in \mathbb{R}^{2\times 2}, \quad c^2 + s^2 = 1.$$

- Multiplication of a nonzero vector $\boldsymbol{x} = (x_1, x_2)^\top$ by matrix $\boldsymbol{G}$ rotates the vector in a clockwise direction by an angle θ, where $c = \cos\theta$ and $s = \sin\theta$.

- Therefore, a plane rotation can be used to zero the 2nd element of a nonzero vector $\boldsymbol{x}$ by choosing $c = \dfrac{x_1}{\|\boldsymbol{x}\|_2}$ and $s = \dfrac{x_2}{\|\boldsymbol{x}\|_2}$:

$$\begin{bmatrix} \frac{x_1}{\|\boldsymbol{x}\|_2} & \frac{x_2}{\|\boldsymbol{x}\|_2} \\ -\frac{x_2}{\|\boldsymbol{x}\|_2} & \frac{x_1}{\|\boldsymbol{x}\|_2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \|\boldsymbol{x}\|_2 \\ 0 \end{bmatrix}, \quad \|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + x_2^2}.$$

Plane rotations (cont'd)

- *By embedding a two-dimensional rotation in a larger unit matrix, one can manipulate vectors of arbitrary dimension.*
- **Example:** Given a vector $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^4$, we transform it to $\kappa \boldsymbol{e}_1$ step by step as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c_1 & s_1 \\ 0 & 0 & -s_1 & c_1 \end{pmatrix} \begin{pmatrix} \times \\ \times \\ \times \\ \times \end{pmatrix} = \begin{pmatrix} \times \\ \times \\ * \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & c_2 & s_2 & 0 \\ 0 & -s_2 & c_2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \times \\ \times \\ \times \\ 0 \end{pmatrix} = \begin{pmatrix} \times \\ * \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} c_3 & s_3 & 0 & 0 \\ -s_3 & c_3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \times \\ \times \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \kappa \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where $\kappa = \|\boldsymbol{x}\|_2$. (why? three orthogonal transformations)

Householder transformations

- Let $\boldsymbol{v} \neq \mathbf{0}$ be an arbitrary vector, and define

$$\boldsymbol{P} := \boldsymbol{I} - \frac{2}{\boldsymbol{v}^\top \boldsymbol{v}} \boldsymbol{v}\boldsymbol{v}^\top := \boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^\top \quad \text{with } \boldsymbol{u} := \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2}.$$

  Then $\boldsymbol{P}^\top \boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}^\top = \boldsymbol{P}^2 = \boldsymbol{I}$, $\boldsymbol{P}$ *is symmetric and orthogonal and called a Householder transformation.*

- Let $\boldsymbol{x} \neq \boldsymbol{y}$ be given nonzero vectors with $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$.

  Can we determine a Householder transformation $\boldsymbol{P}$ *such that* $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{y}$*?*

  Solution: The equation $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{y}$ can be written as

$$\boldsymbol{P}\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} - \frac{2}{\boldsymbol{v}^\top \boldsymbol{v}} \boldsymbol{v}\boldsymbol{v}^\top \boldsymbol{x} = \boldsymbol{x} - \frac{2\boldsymbol{v}^\top \boldsymbol{x}}{\boldsymbol{v}^\top \boldsymbol{v}} \boldsymbol{v} = \boldsymbol{y},$$

  which is of the form $\beta \boldsymbol{v} = \boldsymbol{x} - \boldsymbol{y}$.

  Since $\boldsymbol{P}(\boldsymbol{x}/\beta) = \boldsymbol{y}/\beta$, without loss of generality, we may assume that $\beta = 1$. Then $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{y} \neq \mathbf{0}$.

Householder transformations (cont'd)

- Since $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{y}$ and $\boldsymbol{x}^\top \boldsymbol{x} = \|\boldsymbol{x}\|_2^2 = \|\boldsymbol{y}\|_2^2 = \boldsymbol{y}^\top \boldsymbol{y}$, we obtain

  $$\boldsymbol{v}^\top \boldsymbol{v} = \boldsymbol{x}^\top \boldsymbol{x} - 2\boldsymbol{x}^\top \boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{y} = 2(\boldsymbol{x}^\top \boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{y}),$$

  and further

  $$\boldsymbol{v}^\top \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{x} - \boldsymbol{y}^\top \boldsymbol{x} = \frac{1}{2}\boldsymbol{v}^\top \boldsymbol{v}.$$

  Therefore, we have

  $$\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x} - \frac{2\boldsymbol{v}^\top \boldsymbol{x}}{\boldsymbol{v}^\top \boldsymbol{v}}\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{v} = \boldsymbol{y}.$$

- Now, we choose $\boldsymbol{y} = \kappa \boldsymbol{e}_1$, where $\kappa = \pm\|\boldsymbol{x}\|_2$. Let $\boldsymbol{v} = \boldsymbol{x} - \kappa \boldsymbol{e}_1$. We have the Householder transformation,

  $$\boldsymbol{P} := \boldsymbol{I} - \frac{2}{\boldsymbol{v}^\top \boldsymbol{v}}\boldsymbol{v}\boldsymbol{v}^\top := \boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^\top \quad \text{with } \boldsymbol{u} := \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2}.$$

  Then $\boldsymbol{P}\boldsymbol{x} = \kappa \boldsymbol{e}_1$, which zeros the components in the vector $\boldsymbol{x}$, except the first component.

QR decomposition

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m \geq n$. By a sequence of Householder transformations, we can transform

$$\boldsymbol{A} \to \boldsymbol{Q}^\top \boldsymbol{A} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}, \qquad \left(\Leftrightarrow \boldsymbol{A} = \boldsymbol{Q} \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} \right)$$

where $\boldsymbol{R} \in \mathbb{R}^{n \times n}$ is upper triangular and $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$ is orthogonal.

For example, let $\boldsymbol{A} \in \mathbb{R}^{5 \times 4}$.

$$H_1 A = H_1 \begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix} = \begin{pmatrix} + & + & + & + \\ 0 & + & + & + \\ 0 & + & + & + \\ 0 & + & + & + \\ 0 & + & + & + \end{pmatrix} \cdots\cdots$$

After the 4th step we have computed the upper triangular matrix $\boldsymbol{R}$. The sequence of transformations can be summarized as

$$\boldsymbol{Q}^\top \boldsymbol{A} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{Q}^\top = \boldsymbol{H}_4 \boldsymbol{H}_3 \boldsymbol{H}_2 \boldsymbol{H}_1.$$

QR decomposition (cont'd)

Assume that $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m \geq n$. Then the Householder matrices $\boldsymbol{H}_i$ have the following structure:

$$
\begin{aligned}
\boldsymbol{H}_1 &= \boldsymbol{I} - 2\boldsymbol{u}_1\boldsymbol{u}_1^\top, \quad \boldsymbol{u}_1 \in \mathbb{R}^m, \\
\boldsymbol{H}_2 &= \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{P}_2 \end{bmatrix}, \quad \boldsymbol{P}_2 = \boldsymbol{I} - 2\boldsymbol{u}_2\boldsymbol{u}_2^\top, \quad \boldsymbol{u}_2 \in \mathbb{R}^{m-1}, \\
\boldsymbol{H}_3 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \boldsymbol{P}_3 \end{bmatrix}, \quad \boldsymbol{P}_3 = \boldsymbol{I} - 2\boldsymbol{u}_3\boldsymbol{u}_3^\top, \quad \boldsymbol{u}_3 \in \mathbb{R}^{m-2}, \\
& \cdots
\end{aligned}
$$

QR decomposition Theorem

QR **decomposition:** *Any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m \geq n$, can be transformed to the form:*

$$\boldsymbol{A} = \boldsymbol{Q}\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix},$$

where $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ is orthogonal and $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ is upper triangular. If the columns of $\boldsymbol{A}$ are linearly independent, then $\boldsymbol{R}$ is nonsingular.

Proof:

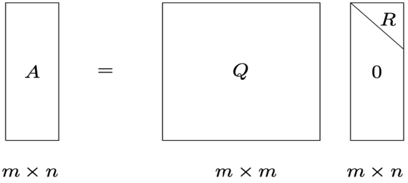
- The constructive procedure outlined in the preceding can easily be adapted. If the vector to be transformed already fulfills the requirement, then $\boldsymbol{P}_i$ is chosen to equal the identity matrix $\boldsymbol{I}$.
- Since

  $$\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix} = \boldsymbol{Q}^\top \boldsymbol{A},$$

  $\boldsymbol{Q}^\top$ is nonsingular, and the columns of $\boldsymbol{A}$ are linearly independent, the columns of the left-hand side matrix are linearly independent, and hence, the $n \times n$ upper triangular matrix $\boldsymbol{R}$ is nonsingular.

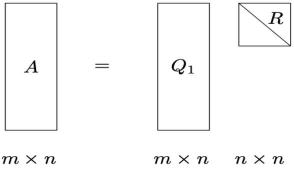
Thin QR decomposition

We illustrate the QR decomposition symbolically:



By partitioning $\mathbf{Q} = (\mathbf{Q}_1 \;\; \mathbf{Q}_2)$*, where* $\mathbf{Q}_1 \in \mathbb{R}^{m\times n}$*, we have the thin QR decomposition:*

$$\boldsymbol{A} = (\mathbf{Q}_1 \;\; \mathbf{Q}_2)\begin{bmatrix} \boldsymbol{R} \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1\boldsymbol{R}.$$



Solving the linear least-squares problem using QR

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m > n$, and $\boldsymbol{b} \in \mathbb{R}^m$. We can solve the following linear least-squares problem

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

using QR decomposition of $\boldsymbol{A}$. Recalling $\boldsymbol{r} := \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$, we have

$$\begin{aligned}
\|\boldsymbol{r}\|_2^2 &= \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 = \|\boldsymbol{b} - \boldsymbol{Q}\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{x}\|_2^2 \\
&= \|\boldsymbol{Q}\Big(\boldsymbol{Q}^\top\boldsymbol{b} - \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{x}\Big)\|_2^2 \\
&= \|\boldsymbol{Q}^\top\boldsymbol{b} - \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{0} \end{bmatrix}\boldsymbol{x}\|_2^2,
\end{aligned}$$

where we use the fact that the Euclidean vector norm is invariant under orthogonal transformation:

$$\|\boldsymbol{Q}\boldsymbol{y}\|_2^2 = (\boldsymbol{Q}\boldsymbol{y})^\top\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{y} = \|\boldsymbol{y}\|_2^2.$$

Solving the linear least-squares problem using QR (cont'd)

Now let $\boldsymbol{Q} = (\boldsymbol{Q}_1 \ \ \boldsymbol{Q}_2)$, where $\boldsymbol{Q}_1 \in \mathbb{R}^{m\times n}$, and denote

$$\boldsymbol{Q}^\top \boldsymbol{b} = \begin{bmatrix} \boldsymbol{Q}_1^\top \boldsymbol{b} \\ \boldsymbol{Q}_2^\top \boldsymbol{b} \end{bmatrix} =: \begin{bmatrix} \widetilde{\boldsymbol{b}}_1 \\ \widetilde{\boldsymbol{b}}_2 \end{bmatrix}.$$

Then we can write

$$\|\boldsymbol{r}\|_2^2 = \| \begin{bmatrix} \widetilde{\boldsymbol{b}}_1 \\ \widetilde{\boldsymbol{b}}_2 \end{bmatrix} - \begin{bmatrix} \boldsymbol{R}\boldsymbol{x} \\ \boldsymbol{0} \end{bmatrix} \|_2^2 = \|\widetilde{\boldsymbol{b}}_1 - \boldsymbol{R}\boldsymbol{x}\|_2^2 + \|\widetilde{\boldsymbol{b}}_2\|_2^2.$$

Since the columns of $\boldsymbol{A}$ *are linearly independent,* $\boldsymbol{R}$ *is nonsingular* and we can solve

$$\boldsymbol{R}\boldsymbol{x} = \widetilde{\boldsymbol{b}}_1,$$

and the solution is given by

$$\boldsymbol{x} = \boldsymbol{R}^{-1}\boldsymbol{Q}_1^\top \boldsymbol{b},$$

which is the unique minimizer of $\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$. (cf. page 5)