# Solving Linear Systems and Least Squares

Last modified: 2026-11-24 09:19:45 (PST)

## Solving linear systems and least squares

> **Source**
>
> This section is adapted from Zhou [1] and Zhou [2], used under the MIT License (see the license text in Section  in Subspaces and Rank). The first source states that every invertible matrix has an LU factorization; that is false without reordering rows (the $2 \times 2$ row swap $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has none), so this section states only how an LU factorization, when one exists, solves a system. The least squares results are proved here from the closest-point theorem (Theorem 3 in Design Matrices and Projections) rather than by differentiating, and the QR factorization is derived from Gram-Schmidt (Theorem 4 in Orthonormal Bases, Outer Products and Hilbert Spaces). These parts of the source are not part of this section:
>
> - the operation counts for the triangular solves, LU and QR
> - solving a system by inverting a matrix through its QR factorization, and the comparison of that with solving it directly
> - the pivoted factorization $\mathbf{PA} = \mathbf{LU}$, which Example 0.10 only touches

**Definition 0.1** (Upper triangular matrix)**.** A square matrix $\mathbf{U}$ is **upper triangular** [3] if every entry below the diagonal is 0: $u_{ij} = 0$ whenever $i > j$.

**Definition 0.2** (Lower triangular matrix)**.** A square matrix $\mathbf{L}$ is **lower triangular** if every entry above the diagonal is 0: $\ell_{ij} = 0$ whenever $i < j$.

**Definition 0.3** (Unit lower triangular matrix)**.** A lower triangular matrix (Definition 0.2) whose diagonal entries are all 1 is **unit lower triangular**.

**Example 0.1** (Triangular and not)**.**

- $\begin{bmatrix} 2 & 1 & -1 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 \end{bmatrix}$ is upper triangular: its entries below the diagonal, in positions $(2,1)$, $(3,1)$ and $(3,2)$, are 0.
- $\begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$ is unit lower triangular.
- A diagonal matrix (Definition 4 in Special Matrices, Quadratic Forms and Trace) is both upper and lower triangular.
- $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is neither: its $(2,1)$ entry is not 0, and neither is its $(1,2)$ entry.

**Theorem 0.1** (Solving a triangular system by substitution)**.** *Let* $\mathbf{U}$ *be an* $n \times n$ *upper triangular matrix (Definition 0.1) whose diagonal entries* $u_{11}, \dots, u_{nn}$ *are all nonzero. For every* $\underset{\sim}{b} \in \mathbb{R}^n$*, the system* $\mathbf{U}\underset{\sim}{x} = \underset{\sim}{b}$ *has exactly one solution, given for* $i = n, n-1, \dots, 1$ *in turn by*

$$x_i = \frac{1}{u_{ii}} \left( b_i - \sum_{j=i+1}^{n} u_{ij}\, x_j \right)$$

*(**back substitution**). Likewise, if* $\mathbf{L}$ *is lower triangular with nonzero diagonal,* $\mathbf{L}\underset{\sim}{y} = \underset{\sim}{b}$ *has exactly one solution, given for* $i = 1, \dots, n$ *by* $y_i = \frac{1}{\ell_{ii}} \left( b_i - \sum_{j=1}^{i-1} \ell_{ij}\, y_j \right)$ *(**forward substitution**).*

Proof

*Proof.* Row $i$ of $\mathbf{U}\underset{\sim}{x} = \underset{\sim}{b}$ reads

$$\begin{aligned} b_i &= \sum_{j=1}^{n} u_{ij}\, x_j && \text{(Definition 8 in Matrix Multiplication)} \\ &= \sum_{j=i}^{n} u_{ij}\, x_j && (u_{ij} = 0 \text{ for } j < i) \\ &= u_{ii}\, x_i + \sum_{j=i+1}^{n} u_{ij}\, x_j. && \text{(split off the } j = i \text{ term)} \end{aligned}$$

This equation is equivalent to each of

$$u_{ii}\,x_i = b_i - \sum_{j=i+1}^{n} u_{ij}\,x_j \qquad \text{(subtract the sum from both sides)}$$

$$x_i = \frac{1}{u_{ii}}\left(b_i - \sum_{j=i+1}^{n} u_{ij}\,x_j\right). \quad \text{(divide by } u_{ii} \neq 0\text{)}$$

Row $n$ involves only $x_n$, so it fixes $x_n$; once $x_{i+1}, \dots, x_n$ are fixed, row $i$ fixes $x_i$. So the rows, taken from the last to the first, hold exactly when $\underset{\sim}{x}$ is the vector the formula builds: there is one solution, and only one. For $\mathbf{L}$, row $i$ reads $\sum_{j=1}^{i-1} \ell_{ij}\,y_j + \ell_{ii}\,y_i = b_i$, and the same argument runs from the first row to the last. □

**Example 0.2** (Back substitution on a $3 \times 3$ system)**.** Solve $\begin{bmatrix} 2 & 1 & -1 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 \end{bmatrix} \underset{\sim}{x} = \begin{bmatrix} 8 \\ 1 \\ 1 \end{bmatrix}$:

1. $$\begin{aligned} x_3 &= \frac{1}{-1} \cdot 1 \\ &= -1; \end{aligned}$$

2. $$\begin{aligned} x_2 &= \frac{1}{1/2}\left(1 - \tfrac{1}{2} \cdot (-1)\right) \\ &= 2 \cdot \tfrac{3}{2} \\ &= 3; \end{aligned}$$

3. $$\begin{aligned} x_1 &= \frac{1}{2}(8 - 1 \cdot 3 - (-1)(-1)) \\ &= \frac{1}{2} \cdot 4 \\ &= 2. \end{aligned}$$

So $\underset{\sim}{x} = (2, 3, -1)$. Without a nonzero diagonal the method fails: $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \underset{\sim}{x} = (1, 1)$ has no solution, since its second row reads $0 = 1$.

**Theorem 0.2** (QR factorization)**.** *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix with* $\operatorname{rank}(\mathbf{A}) = n$*. Then*

$$\underset{m\times n}{\mathbf{A}} = \underset{m\times n}{\mathbf{Q}}\ \underset{n\times n}{\mathbf{R}},$$

*where the columns of* $\mathbf{Q}$ *are orthonormal (Definition 28 in Vectors) and* $\mathbf{R}$ *is upper*

*triangular (Definition 0.1) with positive diagonal entries.*

Proof

*Proof.* The $n$ columns $\underset{\sim}{a}_1, \dots, \underset{\sim}{a}_n$ of $\mathbf{A}$ are linearly independent (Definition 3 in Subspaces and Rank), so the Gram-Schmidt process (Definition 2 in Orthonormal Bases, Outer Products and Hilbert Spaces) runs all $n$ steps and gives orthonormal $\underset{\sim}{q}_1, \dots, \underset{\sim}{q}_n$ (Theorem 4 in Orthonormal Bases, Outer Products and Hilbert Spaces, parts 1 and 3); let $\mathbf{Q}$ have these columns. At step $i$, $\underset{\sim}{q}'_i \neq \underset{\sim}{0}$, and rearranging the orthogonalize step,

$$\begin{aligned} \underset{\sim}{a}_i &= \underset{\sim}{q}'_i + \sum_{j=1}^{i-1} (\underset{\sim}{q}_j \cdot \underset{\sim}{a}_i)\, \underset{\sim}{q}_j && \text{(add the sum to both sides of the orthogonalize step)} \\ &= \|\underset{\sim}{q}'_i\|\, \underset{\sim}{q}_i + \sum_{j=1}^{i-1} (\underset{\sim}{q}_j \cdot \underset{\sim}{a}_i)\, \underset{\sim}{q}_j. && \text{(normalize step: } \underset{\sim}{q}'_i = \|\underset{\sim}{q}'_i\|\, \underset{\sim}{q}_i) \end{aligned}$$

Let $\mathbf{R}$ be the $n \times n$ matrix with $r_{ji} = \underset{\sim}{q}_j \cdot \underset{\sim}{a}_i$ for $j < i$, $r_{ii} = \|\underset{\sim}{q}'_i\| > 0$, and $r_{ji} = 0$ for $j > i$; it is upper triangular with positive diagonal. Column $i$ of $\mathbf{QR}$ is $\mathbf{Q}$ times column $i$ of $\mathbf{R}$ (Definition 1 in Matrix Multiplication), and

$$\begin{aligned} \mathbf{Q}\,(r_{1i}, \dots, r_{ni}) &= \sum_{j=1}^{n} r_{ji}\, \underset{\sim}{q}_j && \text{(Theorem 3 in Subspaces and Rank)} \\ &= \sum_{j=1}^{i} r_{ji}\, \underset{\sim}{q}_j && (r_{ji} = 0 \text{ for } j > i) \\ &= r_{ii}\, \underset{\sim}{q}_i + \sum_{j=1}^{i-1} r_{ji}\, \underset{\sim}{q}_j && \text{(split off the } j = i \text{ term)} \\ &= \|\underset{\sim}{q}'_i\|\, \underset{\sim}{q}_i + \sum_{j=1}^{i-1} (\underset{\sim}{q}_j \cdot \underset{\sim}{a}_i)\, \underset{\sim}{q}_j && \text{(the entries of } \mathbf{R}) \\ &= \underset{\sim}{a}_i. && \text{(the display)} \end{aligned}$$

So $\mathbf{QR} = \mathbf{A}$. □

**Example 0.3** (A QR factorization from Gram-Schmidt)**.** For $\mathbf{A}$ with columns $\underset{\sim}{a}_1 = (1,1,0)$, $\underset{\sim}{a}_2 = (1,0,1)$, $\underset{\sim}{a}_3 = (0,1,1)$, Example 6 in Orthonormal Bases, Outer Products and Hilbert Spaces and Example 8 in Orthonormal Bases, Outer Products and Hilbert Spaces found $\underset{\sim}{q}_1 = \frac{1}{\sqrt{2}}(1,1,0)$, $\underset{\sim}{q}_2 = \frac{1}{\sqrt{6}}(1,-1,2)$, $\underset{\sim}{q}_3 = \frac{1}{\sqrt{3}}(-1,1,1)$, with $\|\underset{\sim}{q}'_1\| = \sqrt{2}$, $\|\underset{\sim}{q}'_2\| = \sqrt{3/2}$, $\|\underset{\sim}{q}'_3\| = 2/\sqrt{3}$,

$$\begin{aligned} \underset{\sim}{q}_1 \cdot \underset{\sim}{a}_2 &= \underset{\sim}{q}_1 \cdot \underset{\sim}{a}_3 \\ &= \tfrac{1}{\sqrt{2}} \end{aligned}$$

and $\underset{\sim}{q}_2 \cdot \underset{\sim}{a}_3 = \frac{1}{\sqrt{6}}$. So

$$\mathbf{R} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{3/2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2}{\sqrt{3}} \end{bmatrix}.$$

As a check on the second column,

$$\begin{aligned} \tfrac{1}{\sqrt{2}}\,\underset{\sim}{q}_1 + \sqrt{3/2}\,\underset{\sim}{q}_2 &= \tfrac{1}{2}\,(1,1,0) + \tfrac{1}{2}\,(1,-1,2) \\ &= (1,0,1) \\ &= \underset{\sim}{a}_2. \end{aligned}$$

**Definition 0.4** (Least squares solution)**.** Let $\mathbf{A}$ be $m \times n$ and $\underset{\sim}{b} \in \mathbb{R}^m$. A **least squares solution** [4] of $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$ is a vector $\underset{\sim}{\hat{x}} \in \mathbb{R}^n$ such that

$$\|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{\hat{x}}\| \le \|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x}\| \quad \text{for every } \underset{\sim}{x} \in \mathbb{R}^n.$$

**Example 0.4** (Fitting a constant)**.** Let $\mathbf{A} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\underset{\sim}{b} = (1,3)$. The system $x\,(1,1) = (1,3)$ is inconsistent (Definition 3 in Generalized Inverses and Linear Systems): its first entry needs $x = 1$ and its second needs $x = 3$. For any number $x$,

$$\begin{aligned} \|\underset{\sim}{b} - \mathbf{A}x\|^2 &= (1-x)^2 + (3-x)^2 \\ &= 2\,(x-2)^2 + 2, \end{aligned}$$

which is smallest at $x = 2$, so $\hat{x} = 2$ is the least squares solution. The value $x = 1$ is not one: it gives $0 + 4 = 4 > 2$. The normal equations read $2x = 4$, and their solution is $x = 2$ too.

**Definition 0.5** (Normal equations)**.** For $\mathbf{A}$ and $\underset{\sim}{b}$ as in Definition 0.4, the equations $\mathbf{A}^\top\mathbf{A}\underset{\sim}{x} = \mathbf{A}^\top\underset{\sim}{b}$ are the **normal equations** [5].

**Theorem 0.3** (Least squares solutions solve the normal equations)**.** *Let* $\mathbf{A}$ *be* $m \times n$ *and* $\underset{\sim}{b} \in \mathbb{R}^m$.

1. $\underset{\sim}{\hat{x}}$ *is a least squares solution (Definition 0.4) exactly when it solves the normal equations* $\mathbf{A}^\top\mathbf{A}\underset{\sim}{\hat{x}} = \mathbf{A}^\top\underset{\sim}{b}$.

2. *The normal equations are consistent (Definition 3 in Generalized Inverses and Linear Systems).*
3. *Every least squares solution gives the same fitted vector $\mathbf{A}\hat{\underset{\sim}{x}}$: the orthogonal projection of $\underset{\sim}{b}$ onto $\mathcal{C}(\mathbf{A})$ (Definition 3 in Design Matrices and Projections).*
4. *The least squares solution is unique exactly when* $\operatorname{rank}(\mathbf{A}) = n$*, and then* $\hat{\underset{\sim}{x}} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\underset{\sim}{b}$.

Proof

*Proof.* Let $\underset{\sim}{u}$ be the orthogonal projection of $\underset{\sim}{b}$ onto $\mathcal{C}(\mathbf{A})$. The vectors $\mathbf{A}\underset{\sim}{x}$, as $\underset{\sim}{x}$ ranges over $\mathbb{R}^n$, are exactly the points of $\mathcal{C}(\mathbf{A})$ (Definition 1 in Column Space and Null Space).
**Least squares solutions are the solutions of $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{u}$.** If $\mathbf{A}\hat{\underset{\sim}{x}} = \underset{\sim}{u}$, then for every $\underset{\sim}{x}$, $\mathbf{A}\underset{\sim}{x} \in \mathcal{C}(\mathbf{A})$, so $\|\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}}\| = \|\underset{\sim}{b} - \underset{\sim}{u}\| \leq \|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x}\|$ (Theorem 3 in Design Matrices and Projections), and $\hat{\underset{\sim}{x}}$ is a least squares solution. Conversely, let $\hat{\underset{\sim}{x}}$ be a least squares solution. $\underset{\sim}{u} \in \mathcal{C}(\mathbf{A})$, so $\underset{\sim}{u} = \mathbf{A}\underset{\sim}{x}_0$ for some $\underset{\sim}{x}_0$ (Definition 1 in Column Space and Null Space), and

$$\begin{aligned}
\|\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}}\| &\leq \|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x}_0\| && \text{(Definition 0.4)}\\
&= \|\underset{\sim}{b} - \underset{\sim}{u}\| && (\underset{\sim}{u} = \mathbf{A}\underset{\sim}{x}_0)\\
&\leq \|\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}}\|, && \text{(Theorem 3 in Design Matrices and Projections, with } \underset{\sim}{w} = \mathbf{A}\hat{\underset{\sim}{x}} \in \mathcal{C}(\mathbf{A}))
\end{aligned}$$

so the last inequality is an equality, and the equality case of Theorem 3 in Design Matrices and Projections gives $\mathbf{A}\hat{\underset{\sim}{x}} = \underset{\sim}{u}$.
**Parts 1 and 3.** $\mathbf{A}\hat{\underset{\sim}{x}} = \underset{\sim}{u}$ holds exactly when $\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}} \in \mathcal{C}(\mathbf{A})^\perp$: if $\mathbf{A}\hat{\underset{\sim}{x}} = \underset{\sim}{u}$, then $\underset{\sim}{b} - \underset{\sim}{u} \in \mathcal{C}(\mathbf{A})^\perp$ (Definition 3 in Design Matrices and Projections); conversely, if $\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}} \in \mathcal{C}(\mathbf{A})^\perp$, then $\underset{\sim}{b} = \mathbf{A}\hat{\underset{\sim}{x}} + (\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}})$ is a split into $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{A})^\perp$, so $\mathbf{A}\hat{\underset{\sim}{x}} = \underset{\sim}{u}$ by the uniqueness in Theorem 4 in Orthogonal Complements and the Fundamental Theorem. And $\mathcal{C}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^\top)$ (Theorem 3 in Orthogonal Complements and the Fundamental Theorem), where

$$\begin{aligned}
\mathbf{A}^\top(\underset{\sim}{b} - \mathbf{A}\hat{\underset{\sim}{x}}) &= \mathbf{A}^\top\underset{\sim}{b} - \mathbf{A}^\top(\mathbf{A}\hat{\underset{\sim}{x}}) && \text{(Theorem 1 in Generalized Inverses and Linear Systems)}\\
&= \mathbf{A}^\top\underset{\sim}{b} - \mathbf{A}^\top\mathbf{A}\hat{\underset{\sim}{x}}, && \text{(Theorem 1 in Matrix Multiplication)}
\end{aligned}$$

which is $\underset{\sim}{0}_n$ exactly when the normal equations hold.
**Part 2.** The vector $\underset{\sim}{x}_0$ above satisfies $\mathbf{A}\underset{\sim}{x}_0 = \underset{\sim}{u}$, so it is a least squares solution, and it solves the normal equations by part 1.
**Part 4.** The least squares solutions are exactly the solutions of $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{u}$, a consistent system, so there is exactly one when $\operatorname{rank}(\mathbf{A}) = n$ and more than one otherwise (Corollary 2 in Generalized Inverses and Linear Systems). When $\operatorname{rank}(\mathbf{A}) = n$, $\mathbf{A}^\top\mathbf{A}$ is invertible (Theorem 1 in Design Matrices and Projections), and with $\mathbf{M} \overset{\text{def}}{=} \mathbf{A}^\top\mathbf{A}$,

$$
\begin{aligned}
\underset{\sim}{\hat{x}} &= \mathbf{I}_n \, \underset{\sim}{\hat{x}} && \text{(Theorem 4 in Matrix Multiplication)} \\
&= (\mathbf{M}^{-1}\mathbf{M}) \, \underset{\sim}{\hat{x}} && \text{(Definition 5 in Special Matrices, Quadratic Forms and Trace)} \\
&= \mathbf{M}^{-1} \, (\mathbf{M}\underset{\sim}{\hat{x}}) && \text{(Theorem 1 in Matrix Multiplication)} \\
&= \mathbf{M}^{-1} \, \mathbf{A}^\top \underset{\sim}{b}. && \text{(the normal equations)}
\end{aligned}
$$

□

**Example 0.5** (A least squares line)**.** Fit $y = x_1 + x_2 t$ to the points $(t, y) = (1,1), (2,2), (3,2)$: $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ (the $\mathbf{X}$ of Example 1 in Design Matrices and Projections) and $\underset{\sim}{b} = (1,2,2)$. $\mathbf{A}$ has rank 2 (Example 2 in Subspaces and Rank), so the solution is unique (part 4). $\mathbf{A}^\top\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$ has inverse $\frac{1}{6}\begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix}$ (Example 1 in Design Matrices and Projections), and

$$
\begin{aligned}
\mathbf{A}^\top \underset{\sim}{b} &= (1+2+2, \; 1+4+6) \\
&= (5, 11),
\end{aligned}
$$

so

$$
\begin{aligned}
\underset{\sim}{\hat{x}} &= \frac{1}{6}(14 \cdot 5 - 6 \cdot 11, \; -6 \cdot 5 + 3 \cdot 11) \\
&= \frac{1}{6}\,(4, 3) \\
&= \left(\tfrac{2}{3}, \tfrac{1}{2}\right).
\end{aligned}
$$

The fitted values are $\mathbf{A}\underset{\sim}{\hat{x}} = \left(\frac{7}{6}, \frac{5}{3}, \frac{13}{6}\right)$, and the residuals $\underset{\sim}{b} - \mathbf{A}\underset{\sim}{\hat{x}} = \left(-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}\right)$ are orthogonal to both columns of $\mathbf{A}$: $-\frac{1}{6} + \frac{1}{3} - \frac{1}{6} = 0$ and $-\frac{1}{6} + \frac{2}{3} - \frac{1}{2} = 0$.

**Example 0.6** (Many least squares solutions, one fitted vector)**.** With the rank-1 matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$ (Example 2 in Subspaces and Rank) and $\underset{\sim}{b} = (1,2,2)$, $\mathbf{A}^\top\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix}$ and $\mathbf{A}^\top\underset{\sim}{b} = (5, 10)$, so both normal equations say $3x_1 + 6x_2 = 5$, that is, $x_1 + 2x_2 = \frac{5}{3}$. There are infinitely many least squares solutions, such as $\left(\frac{5}{3}, 0\right)$ and $\left(0, \frac{5}{6}\right)$, as part 4 predicts for $\text{rank}(\mathbf{A}) = 1 < 2$; but every one gives the same fitted vector

$$
\begin{aligned}
\mathbf{A}\underset{\sim}{\hat{x}} &= (x_1 + 2x_2)\,(1,1,1) \\
&= \tfrac{5}{3}\,(1,1,1),
\end{aligned}
$$

as part 3 says.

**Theorem 0.4** (Solving least squares by QR)**.** *Let* $\mathbf{A}$ *be* $m \times n$ *with* $\operatorname{rank}(\mathbf{A}) = n$*, with QR factorization* $\mathbf{A} = \mathbf{QR}$ *(Theorem 0.2), and let* $\underset{\sim}{b} \in \mathbb{R}^m$*. The least squares solution (Definition 0.4) is the unique solution of*

$$\underset{n\times n}{\mathbf{R}}\ \underset{\sim}{\hat{x}} = \underset{n\times m}{\mathbf{Q}^\top}\ \underset{\sim}{b},$$

*which back substitution computes (Theorem 0.1). When* $m = n$*, it is the unique solution of* $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$*.*

Proof

*Proof.* $\mathbf{R}$ is upper triangular with positive diagonal, so $\mathbf{R}\underset{\sim}{x} = \mathbf{Q}^\top \underset{\sim}{b}$ has exactly one solution $\underset{\sim}{\hat{x}}$ (Theorem 0.1). $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_n$, as in the proof of Theorem 4 in Design Matrices and Projections. Then

$$\begin{aligned}
\mathbf{A}^\top\mathbf{A}\underset{\sim}{\hat{x}} &= (\mathbf{QR})^\top(\mathbf{QR})\,\underset{\sim}{\hat{x}} && \text{(substitute } \mathbf{A} = \mathbf{QR}\text{)}\\
&= \mathbf{R}^\top\mathbf{Q}^\top(\mathbf{QR})\,\underset{\sim}{\hat{x}} && \text{(Theorem 6 in Matrix Multiplication)}\\
&= \mathbf{R}^\top(\mathbf{Q}^\top\mathbf{Q})(\mathbf{R}\underset{\sim}{\hat{x}}) && \text{(Theorem 1 in Matrix Multiplication)}\\
&= \mathbf{R}^\top\mathbf{I}_n(\mathbf{R}\underset{\sim}{\hat{x}}) && (\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_n)\\
&= \mathbf{R}^\top(\mathbf{R}\underset{\sim}{\hat{x}}) && \text{(Theorem 4 in Matrix Multiplication)}\\
&= \mathbf{R}^\top(\mathbf{Q}^\top\underset{\sim}{b}) && (\mathbf{R}\underset{\sim}{\hat{x}} = \mathbf{Q}^\top\underset{\sim}{b})\\
&= (\mathbf{R}^\top\mathbf{Q}^\top)\,\underset{\sim}{b} && \text{(Theorem 1 in Matrix Multiplication)}\\
&= (\mathbf{QR})^\top\,\underset{\sim}{b} && \text{(Theorem 6 in Matrix Multiplication)}\\
&= \mathbf{A}^\top\underset{\sim}{b}, && (\mathbf{A} = \mathbf{QR})
\end{aligned}$$

so $\underset{\sim}{\hat{x}}$ solves the normal equations and is the least squares solution (Theorem 0.3), which is unique because $\operatorname{rank}(\mathbf{A}) = n$. When $m = n$,

$$\begin{aligned}
\operatorname{rank}(\mathbf{A}) &= n\\
&= m,
\end{aligned}$$

so $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$ is consistent for every $\underset{\sim}{b}$ (Corollary 2 in Generalized Inverses and Linear Systems, part 1), so its solution leaves residual $\underset{\sim}{0}$, is a least squares solution, and is therefore $\underset{\sim}{\hat{x}}$. □

**Example 0.7** (The least squares line by QR)**.** For $\mathbf{A}$ and $\underset{\sim}{b}$ of Example 0.5, Gram-Schmidt gives $\underset{\sim}{q}_1 = \frac{1}{\sqrt{3}}(1,1,1)$; then

$$\begin{aligned} \underset{\sim}{q}_1 \cdot (1,2,3) &= \tfrac{6}{\sqrt{3}} \\ &= 2\sqrt{3}, \end{aligned}$$

$$\begin{aligned} \underset{\sim}{q}_2' &= (1,2,3) - 2\,(1,1,1) \\ &= (-1,0,1) \end{aligned}$$

with norm $\sqrt{2}$, and $\underset{\sim}{q}_2 = \frac{1}{\sqrt{2}}\,(-1,0,1)$. So

$$\begin{aligned} \mathbf{R} &= \begin{bmatrix} \sqrt{3} & 2\sqrt{3} \\ 0 & \sqrt{2} \end{bmatrix}, \\ \mathbf{Q}^\top \underset{\sim}{b} &= \left(\tfrac{1+2+2}{\sqrt{3}},\ \tfrac{-1+0+2}{\sqrt{2}}\right) \\ &= \left(\tfrac{5}{\sqrt{3}}, \tfrac{1}{\sqrt{2}}\right). \end{aligned}$$

Back substitution gives

$$\begin{aligned} x_2 &= \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} \\ &= \tfrac{1}{2} \end{aligned}$$

and

$$\begin{aligned} x_1 &= \tfrac{1}{\sqrt{3}}\left(\tfrac{5}{\sqrt{3}} - 2\sqrt{3} \cdot \tfrac{1}{2}\right) \\ &= \tfrac{5}{3} - 1 \\ &= \tfrac{2}{3}, \end{aligned}$$

the solution found in Example 0.5.

**Theorem 0.5** (The Moore-Penrose inverse gives a least squares solution)**.** *Let* $\mathbf{A}$ *be* $m \times n$ *and* $\underset{\sim}{b} \in \mathbb{R}^m$*. Then* $\mathbf{A}^+\underset{\sim}{b}$ *is a least squares solution of* $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$*, and a vector* $\underset{\sim}{x} \in \mathbb{R}^n$ *is a least squares solution if and only if*

$$\mathbf{A}\underset{\sim}{x} = \mathbf{A}\mathbf{A}^+\underset{\sim}{b}.$$

Proof

*Proof.* Let $\mathbf{P} = \mathbf{A}\mathbf{A}^+$. By conditions 3 and 1 of Definition 2 in Generalized Inverses and Linear Systems, $\mathbf{P}^\top = \mathbf{P}$ and $\mathbf{P}\mathbf{A} = \mathbf{A}$. Hence

$$\begin{aligned}
\mathbf{A}^\top(\mathbf{I}-\mathbf{P}) &= \mathbf{A}^\top - \mathbf{A}^\top\mathbf{P} && \text{(distributing)}\\
&= \mathbf{A}^\top - \mathbf{A}^\top\mathbf{P}^\top && (\mathbf{P}^\top = \mathbf{P})\\
&= \mathbf{A}^\top - (\mathbf{P}\mathbf{A})^\top && \text{(Theorem 6 in Matrix Multiplication)}\\
&= \mathbf{A}^\top - \mathbf{A}^\top && (\mathbf{P}\mathbf{A} = \mathbf{A})\\
&= \mathbf{0}.
\end{aligned}$$

For every $\underset{\sim}{x} \in \mathbb{R}^n$, write $\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x} = \underset{\sim}{u} + \underset{\sim}{v}$ with

$$\underset{\sim}{u} = \underset{\sim}{b} - \mathbf{P}\underset{\sim}{b} \qquad \text{and} \qquad \underset{\sim}{v} = \mathbf{P}\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x} = \mathbf{A}(\mathbf{A}^+\underset{\sim}{b} - \underset{\sim}{x}).$$

Their inner product is

$$\begin{aligned}
\underset{\sim}{v}\cdot\underset{\sim}{u} &= (\mathbf{A}^+\underset{\sim}{b} - \underset{\sim}{x})^\top \mathbf{A}^\top(\mathbf{I}-\mathbf{P})\underset{\sim}{b} && \text{(Theorem 6 in Matrix Multiplication)}\\
&= 0, && (\mathbf{A}^\top(\mathbf{I}-\mathbf{P}) = \mathbf{0})
\end{aligned}$$

so $\|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x}\|^2 = \|\underset{\sim}{u}\|^2 + \|\underset{\sim}{v}\|^2$. That is,

$$\|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x}\|^2 = \|\underset{\sim}{b} - \mathbf{P}\underset{\sim}{b}\|^2 + \|\mathbf{A}(\mathbf{A}^+\underset{\sim}{b} - \underset{\sim}{x})\|^2.$$

The first summand does not involve $\underset{\sim}{x}$, and the second is nonnegative, with a zero at $\underset{\sim}{x} = \mathbf{A}^+\underset{\sim}{b}$. So $\mathbf{A}^+\underset{\sim}{b}$ minimizes $\|\underset{\sim}{b} - \mathbf{A}\underset{\sim}{x}\|$, and the smallest norm is $\|\underset{\sim}{b} - \mathbf{P}\underset{\sim}{b}\|$. A vector $\underset{\sim}{x}$ attains that norm if and only if $\mathbf{A}(\mathbf{A}^+\underset{\sim}{b} - \underset{\sim}{x}) = \underset{\sim}{0}$, which is $\mathbf{A}\underset{\sim}{x} = \mathbf{A}\mathbf{A}^+\underset{\sim}{b}$. □

**Theorem 0.6** (The Moore-Penrose inverse gives the least squares solution of smallest norm)**.** *Let* $\mathbf{A}$ *be* $m \times n$ *and* $\underset{\sim}{b} \in \mathbb{R}^m$*. If* $\underset{\sim}{x}$ *is a least squares solution (Definition 0.4) of* $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$*, then*

$$\|\underset{\sim}{x}\|^2 = \|\mathbf{A}^+\underset{\sim}{b}\|^2 + \|\underset{\sim}{x} - \mathbf{A}^+\underset{\sim}{b}\|^2.$$

*So* $\|\underset{\sim}{x}\| \geq \|\mathbf{A}^+\underset{\sim}{b}\|$*, with equality only if* $\underset{\sim}{x} = \mathbf{A}^+\underset{\sim}{b}$*.*

Proof

*Proof.* Let $\underset{\sim}{d} = \underset{\sim}{x} - \mathbf{A}^+\underset{\sim}{b}$. By Theorem 0.5, $\mathbf{A}\underset{\sim}{x} = \mathbf{A}\mathbf{A}^+\underset{\sim}{b}$, so $\mathbf{A}\underset{\sim}{d} = \underset{\sim}{0}$.
Let $\underset{\sim}{w} = (\mathbf{A}^+)^\top\mathbf{A}^+\underset{\sim}{b}$. Then

$$\begin{aligned}
\mathbf{A}^+\underset{\sim}{b} &= \mathbf{A}^+\mathbf{A}\mathbf{A}^+\underset{\sim}{b} && \text{(condition 2 of Definition 2 in Generalized Inverses and Linear Systems)}\\
&= (\mathbf{A}^+\mathbf{A})^\top\mathbf{A}^+\underset{\sim}{b} && \text{(condition 4 of Definition 2 in Generalized Inverses and Linear Systems)}\\
&= \mathbf{A}^\top(\mathbf{A}^+)^\top\mathbf{A}^+\underset{\sim}{b} && \text{(Theorem 6 in Matrix Multiplication)}\\
&= \mathbf{A}^\top\underset{\sim}{w}. && \text{(definition of } \underset{\sim}{w})
\end{aligned}$$

Therefore

$$\begin{aligned} \mathbf{A}^{+}\underset{\sim}{b}\cdot\underset{\sim}{d} &= \mathbf{A}^{\top}\underset{\sim}{w}\cdot\underset{\sim}{d} && (\mathbf{A}^{+}\underset{\sim}{b}=\mathbf{A}^{\top}\underset{\sim}{w}) \\ &= \underset{\sim}{w}\cdot\mathbf{A}\underset{\sim}{d} && (\text{Theorem 6 in Matrix Multiplication}) \\ &= 0. && (\mathbf{A}\underset{\sim}{d}=\underset{\sim}{0}) \end{aligned}$$

Expanding $\|\underset{\sim}{x}\|^2 = \|\mathbf{A}^{+}\underset{\sim}{b}+\underset{\sim}{d}\|^2$ then gives

$$\begin{aligned} \|\underset{\sim}{x}\|^2 &= \|\mathbf{A}^{+}\underset{\sim}{b}\|^2 + 2\,\mathbf{A}^{+}\underset{\sim}{b}\cdot\underset{\sim}{d} + \|\underset{\sim}{d}\|^2 && (\text{expanding}) \\ &= \|\mathbf{A}^{+}\underset{\sim}{b}\|^2 + \|\underset{\sim}{d}\|^2. && (\mathbf{A}^{+}\underset{\sim}{b}\cdot\underset{\sim}{d}=0) \end{aligned}$$

Equality in $\|\underset{\sim}{x}\| \geq \|\mathbf{A}^{+}\underset{\sim}{b}\|$ holds only if $\|\underset{\sim}{d}\|=0$, that is, $\underset{\sim}{x}=\mathbf{A}^{+}\underset{\sim}{b}$. □

**Example 0.8** (A rank-deficient least squares problem)**.** Let $\mathbf{A}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$ and $\underset{\sim}{b}=(2,3)$. By Example 5 in Generalized Inverses and Linear Systems, $\mathbf{A}^{+}=\mathbf{A}$, so $\mathbf{A}^{+}\underset{\sim}{b}=(2,0)$ and $\mathbf{A}\mathbf{A}^{+}\underset{\sim}{b}=(2,0)$.
By Theorem 0.5, $\underset{\sim}{x}=(x_1,x_2)$ is a least squares solution if and only if $\mathbf{A}\underset{\sim}{x}=(x_1,0)=(2,0)$, that is, $x_1=2$ with $x_2$ arbitrary. Each of them leaves the same residual $\underset{\sim}{b}-\mathbf{A}\underset{\sim}{x}=(0,3)$, of norm 3.
For each such $\underset{\sim}{x}$,

$$\|\underset{\sim}{x}\|^2 = 4+x_2^2 = \|\mathbf{A}^{+}\underset{\sim}{b}\|^2 + \|\underset{\sim}{x}-\mathbf{A}^{+}\underset{\sim}{b}\|^2,$$

as Theorem 0.6 states. The smallest is 4, at $x_2=0$, which is $\mathbf{A}^{+}\underset{\sim}{b}=(2,0)$.

**Definition 0.6** (LU factorization)**.** An **LU factorization** [6] of an $n\times n$ matrix $\mathbf{A}$ is a product $\mathbf{A}=\mathbf{L}\mathbf{U}$ with $\mathbf{L}$ unit lower triangular (Definition 0.3) and $\mathbf{U}$ upper triangular (Definition 0.1).

**Example 0.9** (An LU factorization of a $3\times 3$ matrix)**.** Let $\mathbf{A}=\begin{bmatrix}2&1&-1\\-3&-1&2\\-2&1&2\end{bmatrix}$, $\mathbf{L}=\begin{bmatrix}1&0&0\\-\frac{3}{2}&1&0\\-1&4&1\end{bmatrix}$ (unit lower triangular) and $\mathbf{U}=\begin{bmatrix}2&1&-1\\0&\frac{1}{2}&\frac{1}{2}\\0&0&-1\end{bmatrix}$ (upper triangular).
By Definition 1 in Matrix Multiplication, entry $(i,j)$ of $\mathbf{L}\mathbf{U}$ is $\sum_k \ell_{ik}\,u_{kj}$, so row $i$ of $\mathbf{L}\mathbf{U}$ is $\sum_k \ell_{ik}$ times row $k$ of $\mathbf{U}$. Row 1 is $1\cdot(2,1,-1)=(2,1,-1)$; row 2 is $-\frac{3}{2}(2,1,-1)+(0,\frac{1}{2},\frac{1}{2})=(-3,-1,2)$, and row 3 is $-(2,1,-1)+4\,(0,\frac{1}{2},\frac{1}{2})+(0,0,-1)=(-2,1,2)$. These are the rows of $\mathbf{A}$, so $\mathbf{L}\mathbf{U}=\mathbf{A}$ is an LU factorization. The below-

diagonal entries of $\mathbf{L}$ are the negatives of the multiples of earlier rows that Gaussian elimination adds to reduce $\mathbf{A}$ to $\mathbf{U}$.

**Example 0.10** (An invertible matrix with no LU factorization)**.** $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is invertible: $\mathbf{P}^2 = \mathbf{I}_2$, so $\mathbf{P}^{-1} = \mathbf{P}$ (Definition 5 in Special Matrices, Quadratic Forms and Trace); but it has no LU factorization. If $\mathbf{P} = \mathbf{LU}$, the $(1,1)$ entry gives $0 = 1 \cdot u_{11}$, so $u_{11} = 0$, and then the $(2,1)$ entry gives

$$\begin{aligned} 1 &= \ell_{21}\, u_{11} \\ &= 0, \end{aligned}$$

which is impossible. Swapping the two rows first removes the obstacle: the swapped matrix is $\mathbf{I}_2 = \mathbf{I}_2\, \mathbf{I}_2$.

**Theorem 0.7** (Solving a system with an LU factorization)**.** *If $\mathbf{A} = \mathbf{LU}$ is an LU factorization (Definition 0.6) and the diagonal entries of $\mathbf{U}$ are all nonzero, then for every $\underset{\sim}{b} \in \mathbb{R}^n$ the system $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$ has exactly one solution: solve $\mathbf{L}\underset{\sim}{y} = \underset{\sim}{b}$ by forward substitution, then $\mathbf{U}\underset{\sim}{x} = \underset{\sim}{y}$ by back substitution (Theorem 0.1).*

Proof

*Proof.* $\mathbf{L}$ has diagonal entries 1 and $\mathbf{U}$ has nonzero ones, so each of the two triangular systems has exactly one solution (Theorem 0.1). If $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are those solutions, then

$$\begin{aligned} \mathbf{A}\underset{\sim}{x} &= \mathbf{L}\,(\mathbf{U}\underset{\sim}{x}) \\ &= \mathbf{L}\underset{\sim}{y} \\ &= \underset{\sim}{b} \end{aligned}$$

(Theorem 1 in Matrix Multiplication), so $\underset{\sim}{x}$ is a solution. Conversely, if $\mathbf{A}\underset{\sim}{x} = \underset{\sim}{b}$, then $\underset{\sim}{y} \overset{\text{def}}{=} \mathbf{U}\underset{\sim}{x}$ satisfies

$$\begin{aligned} \mathbf{L}\underset{\sim}{y} &= \mathbf{L}\,(\mathbf{U}\underset{\sim}{x}) \\ &= (\mathbf{LU})\,\underset{\sim}{x} \\ &= \mathbf{A}\underset{\sim}{x} \\ &= \underset{\sim}{b} \end{aligned}$$

(Theorem 1 in Matrix Multiplication), so $\underset{\sim}{y}$ is the unique solution of that system, and $\underset{\sim}{x}$ is then the unique solution of $\mathbf{U}\underset{\sim}{x} = \underset{\sim}{y}$. □

**Example 0.11** (Two triangular solves)**.** With $\mathbf{L}$ and $\mathbf{U}$ of Example 0.9, solve $\mathbf{A}\underset{\sim}{x} = (8, -11, -3)$. Forward substitution on $\mathbf{L}\underset{\sim}{y} = (8, -11, -3)$ gives $y_1 = 8$,

$$\begin{aligned} y_2 &= -11 + \tfrac{3}{2} \cdot 8 \\ &= 1, \end{aligned}$$

$$\begin{aligned} y_3 &= -3 + 8 - 4 \cdot 1 \\ &= 1. \end{aligned}$$

Back substitution on $\mathbf{U}\underset{\sim}{x} = (8, 1, 1)$ is Example 0.2, which gives $\underset{\sim}{x} = (2, 3, -1)$. As a check,

$$\begin{aligned} \mathbf{A}\,(2, 3, -1) &= (4 + 3 + 1,\ -6 - 3 - 2,\ -4 + 3 - 2) \\ &= (8, -11, -3). \end{aligned}$$


[1] H. Zhou, "Linear equations and matrix inverses." Lecture notes for Biostat 216, Mathematical Methods for Biostatistics, University of California, Los Angeles, 2024. Accessed: Oct. 05, 2026. [Online]. Available: https://ucla-biostat-216.github.io/2024fall/slides/07-matinv/07-matinv.html

[2] H. Zhou, "Least squares." Lecture notes for Biostat 216, Mathematical Methods for Biostatistics, University of California, Los Angeles, 2024. Accessed: Oct. 05, 2026. [Online]. Available: https://ucla-biostat-216.github.io/2024fall/slides/08-ls/08-ls.html

[3] Wikipedia contributors, "Triangular matrix — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Triangular_matrix

[4] Wikipedia contributors, "Least squares — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Least_squares

[5] Wikipedia contributors, "Normal equations — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/Normal_equations

[6] Wikipedia contributors, "LU decomposition — Wikipedia, the free encyclopedia." 2026. Available: https://en.wikipedia.org/wiki/LU_decomposition