

Projections and Linear Systems

Last modified: 2026-10-27 20:53:54 (PDT)

Design Matrix

Definition 0.1 (Design matrix, model matrix). In a regression model with n observations and p predictors, the **design matrix** [1] (or **model matrix**) $\mathbf{X}$ is the $n \times p$ matrix whose i -th row is the covariate vector $\tilde{x}_i^\top$ for observation i :

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix} \tilde{x}_1^\top \\ \tilde{x}_2^\top \\ \vdots \\ \tilde{x}_n^\top \end{bmatrix} \\ &= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}\end{aligned}$$

Remark (Products with the design matrix)

Remark 0.1 (Products with the design matrix). The product $\mathbf{X}\beta$ collects the values $\tilde{x}_i^\top \beta$ for all n observations into a single $n \times 1$ vector:

$$\mathbf{X}\beta = \begin{bmatrix} \tilde{x}_1^\top \beta \\ \vdots \\ \tilde{x}_n^\top \beta \end{bmatrix}$$

For example, take the rank-2 matrix from [Example 2 in Subspaces and Rank](#) as $\mathbf{X}$, with $n = 3$ observations and $p = 2$ columns: an intercept column and one covariate. With $\beta = (1, 2)$,

$$\begin{aligned}\mathbf{X}_{\sim}\beta &= \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}.\end{aligned}$$

The matrix $\mathbf{X}^\top \mathbf{X}$ is $p \times p$ and symmetric, since

$$\begin{aligned}(\mathbf{X}^\top \mathbf{X})^\top &= \mathbf{X}^\top (\mathbf{X}^\top)^\top \\ &= \mathbf{X}^\top \mathbf{X}\end{aligned}$$

([Theorem 17 in Matrices](#)). When $\mathbf{X}^\top \mathbf{X}$ is invertible, it appears in the least squares solution (Definition [0.10](#), defined later in this chapter) $\hat{\beta}_{\sim} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top y_{\sim}$.

Theorem 0.1 ($\mathbf{X}^\top \mathbf{X}$ is invertible when $\mathbf{X}$ has full column rank). *If $\mathbf{X}$ is an $n \times p$ matrix with $\text{rank}(\mathbf{X}) = p$ ([Definition 2 in Subspaces and Rank](#)), then the $p \times p$ matrix $\mathbf{X}^\top \mathbf{X}$ is invertible.*

i Proof

Proof. Let $\tilde{c}$ be a vector of length p with $\mathbf{X}^\top \mathbf{X} \tilde{c} = \tilde{0}$. Then

$$\begin{aligned}0 &= \tilde{c}^\top \mathbf{X}^\top \mathbf{X} \tilde{c} \quad (\text{multiply } \mathbf{X}^\top \mathbf{X} \tilde{c} = \tilde{0} \text{ on the left by } \tilde{c}^\top) \\ &= (\mathbf{X} \tilde{c})^\top (\mathbf{X} \tilde{c}) \quad (\text{transpose of a product}) \\ &= \|\mathbf{X} \tilde{c}\|^2 \quad (\text{definition of the Euclidean norm})\end{aligned}$$

so $\mathbf{X} \tilde{c} = \tilde{0}$. Because $\mathbf{X} \tilde{c} = c_1(\text{column } 1) + \cdots + c_p(\text{column } p)$ and the columns of $\mathbf{X}$ are linearly independent, $\tilde{c} = \tilde{0}$. So the only solution of $\mathbf{X}^\top \mathbf{X} \tilde{c} = \tilde{0}$ is $\tilde{c} = \tilde{0}$, which says that the columns of the square matrix $\mathbf{X}^\top \mathbf{X}$ are linearly independent, and a square matrix with linearly independent columns is invertible [2], Corollary 5.7, p. 143. $\square$

Example 0.1 (Inverting $\mathbf{X}^\top \mathbf{X}$). For the rank-2 matrix $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ from [Example 2 in Subspaces and Rank](#):

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \quad (\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix},$$

and multiplying the two out gives $\mathbf{I}_2$.

Definition 0.2 (Hat matrix). For an $n \times p$ design matrix $\mathbf{X}$ (Definition 0.1) with $\text{rank}(\mathbf{X}) = p$, the **hat matrix** [3] is the $n \times n$ matrix

$$\mathbf{H} \stackrel{\text{def}}{=} \underbrace{\mathbf{X}}_{n \times p} \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1}}_{p \times p} \underbrace{\mathbf{X}^\top}_{p \times n}$$

The inverse exists by Theorem 0.1.

Example 0.2 (The hat matrix of an intercept-only model). With $n = 2$ observations and only an intercept, $\mathbf{X} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (2×1 , rank 1), so $\mathbf{X}^\top \mathbf{X} = 2$ and

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \frac{1}{2} \cdot \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}. \end{aligned}$$

Then $\mathbf{H}\mathbf{y} = (\bar{y}, \bar{y})$, where $\bar{y} = (y_1 + y_2)/2$: the fitted values of an intercept-only model are the [sample mean](#).

Theorem 0.2 (Hat matrix is a projection matrix). *If $\mathbf{X}$ is an $n \times p$ design matrix with $\text{rank}(\mathbf{X}) = p$, then the hat matrix $\mathbf{H}$ (Definition 0.2) is an orthogonal projection matrix (Definition 8 in Special Matrices and Decompositions).*

Proof

Proof. We verify symmetry and idempotency. Both use that $\mathbf{X}^\top \mathbf{X}$ is symmetric,

$$\begin{aligned} (\mathbf{X}^\top \mathbf{X})^\top &= \mathbf{X}^\top (\mathbf{X}^\top)^\top \\ &= \mathbf{X}^\top \mathbf{X} \end{aligned}$$

(Theorem 17 in Matrices), so its inverse is symmetric too (Corollary 1 in Special Matrices and Decompositions).

Symmetry:

$$\begin{aligned} \mathbf{H}^\top &= (\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)^\top && \text{(definition of } \mathbf{H} \text{)} \\ &= (\mathbf{X}^\top)^\top \cdot ((\mathbf{X}^\top \mathbf{X})^{-1})^\top \cdot \mathbf{X}^\top && \text{(transpose of a product, twice)} \\ &= \mathbf{X} \cdot ((\mathbf{X}^\top \mathbf{X})^{-1})^\top \cdot \mathbf{X}^\top && \text{(transposing twice changes nothing)} \\ &= \mathbf{X} \cdot (\mathbf{X}^\top \mathbf{X})^{-1} \cdot \mathbf{X}^\top && \text{(the inverse of a symmetric matrix is symmetric)} \\ &= \mathbf{H} && \text{(definition of } \mathbf{H} \text{)} \end{aligned}$$

Idempotency:

$$\begin{aligned}
\mathbf{H}^2 &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \cdot \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top && \text{(definition of } \mathbf{H} \text{)} \\
&= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top && \text{(regroup; matrix multiplication is associative)} \\
&= \mathbf{X} \mathbf{I}_p (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top && \text{(definition of the inverse)} \\
&= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top && \text{(identity matrix)} \\
&= \mathbf{H} && \text{(definition of } \mathbf{H} \text{)}
\end{aligned}$$

□

Example 0.3 (The intercept-only hat matrix is a projection). For $\mathbf{H} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ from Example 0.2, $\mathbf{H}^\top = \mathbf{H}$, and

$$\begin{aligned}
\mathbf{H}^2 &= \begin{bmatrix} 0.5 \cdot 0.5 + 0.5 \cdot 0.5 & 0.5 \cdot 0.5 + 0.5 \cdot 0.5 \\ 0.5 \cdot 0.5 + 0.5 \cdot 0.5 & 0.5 \cdot 0.5 + 0.5 \cdot 0.5 \end{bmatrix} \\
&= \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \\
&= \mathbf{H},
\end{aligned}$$

as Theorem 0.2 says.

i Remark (Why it is called the hat matrix)

Remark 0.2 (Why it is called the hat matrix). The hat matrix gives the fitted values of the least squares solution (Definition 0.10, defined later in this chapter):

$$\begin{aligned}
\hat{\tilde{y}} &= \mathbf{X} \hat{\tilde{\beta}} \\
&= \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{y} \\
&= \mathbf{H} \tilde{y}.
\end{aligned}$$

Multiplying by $\mathbf{H}$ “puts a hat” on $\tilde{y}$, which is where the name comes from. In Example 0.2, $\mathbf{H}$ puts a hat on $\tilde{y}$ by replacing each y_i with the sample mean $\bar{y}$.

Orthogonal projection onto a subspace

Source

This section is adapted from the second half of Zhou [4], used under the MIT License (see the license text in [Section 2 in Subspaces and Rank](#)). It connects the orthogonal projection matrices of [Definition 8 in Special Matrices and Decompositions](#), defined by an algebraic property, to the geometric decomposition of [Theorem 8 in Direct Sums and Orthogonal Complements](#).

Definition 0.3 (Orthogonal projection onto a subspace). Let $\mathcal{S}$ be a subspace of $\mathbb{R}^p$ and $y \in \mathbb{R}^p$. By [Theorem 8 in Direct Sums and Orthogonal Complements](#), $y = \tilde{u} + \tilde{v}$ for exactly one $\tilde{u} \in \mathcal{S}$ and $\tilde{v} \in \mathcal{S}^\perp$. The vector $\tilde{u}$ is the **orthogonal projection** of y onto $\mathcal{S}$. It is the projection of $\tilde{y}$ onto $\mathcal{S}$ along $\mathcal{S}^\perp$ ([Definition 4 in Direct Sums and Orthogonal Complements](#)).

Example 0.4 (Projecting onto a line in $\mathbb{R}^3$). In [Example 15 in Direct Sums and Orthogonal Complements](#), $y = (3, 1, 2)$ splits as $(2, 2, 0) + (1, -1, 2)$, with $(2, 2, 0) \in \mathcal{S} = \text{span}\{(1, 1, 0)\}$ and $(1, -1, 2) \in \mathcal{S}^\perp$. So the orthogonal projection of $(3, 1, 2)$ onto $\mathcal{S}$ is $(2, 2, 0)$. A vector already in $\mathcal{S}$, such as $(5, 5, 0)$, is its own projection, since $(5, 5, 0) = (5, 5, 0) + 0$ and $0 \in \mathcal{S}^\perp$ ([Theorem 6 in Direct Sums and Orthogonal Complements](#), [Theorem 1 in Subspaces and Rank](#)); a vector of $\mathcal{S}^\perp$, such as $(1, -1, 2)$, projects to 0 , since $(1, -1, 2) = 0 + (1, -1, 2)$ and $0 \in \mathcal{S}$ ([Theorem 1 in Subspaces and Rank](#)).

Example 0.5 (A split that is not the orthogonal one). $(3, 1, 2) = (3, 3, 0) + (0, -2, 2)$ also writes $(3, 1, 2)$ as a vector of $\mathcal{S} = \text{span}\{(1, 1, 0)\}$ plus a remainder, but $(1, 1, 0) \cdot (0, -2, 2) = -2 \neq 0$, so the remainder is not in $\mathcal{S}^\perp$, and $(3, 3, 0)$ is not the orthogonal projection of $(3, 1, 2)$ onto $\mathcal{S}$.

Theorem 0.3 (The orthogonal projection is the closest point). *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^p$, let $y \in \mathbb{R}^p$, and let $\tilde{u}$ be the orthogonal projection of y onto $\mathcal{S}$ (Definition 0.3). Then for every $\tilde{w} \in \mathcal{S}$,*

$$\|y - \tilde{u}\| \leq \|y - \tilde{w}\|,$$

with equality only when $\tilde{w} = \tilde{u}$.

i Proof

Proof. By Definition 0.3, $y - \underline{u} \in \mathcal{S}^\perp$. Both $\underline{u}$ and $\underline{w}$ are in $\mathcal{S}$, so $\underline{u} - \underline{w} = \underline{u} + (-1)\underline{w} \in \mathcal{S}$ (Definition 4 in Subspaces and Rank), and therefore $(y - \underline{u}) \perp (\underline{u} - \underline{w})$ (Definition 5 in Direct Sums and Orthogonal Complements). Then

$$\begin{aligned}\|\underline{y} - \underline{w}\|^2 &= \|(\underline{y} - \underline{u}) + (\underline{u} - \underline{w})\|^2 && \text{(add and subtract } \underline{u}) \\ &= \|\underline{y} - \underline{u}\|^2 + \|\underline{u} - \underline{w}\|^2 && \text{(Pythagorean theorem, Theorem 2 in Inner Products and Orthogonality)} \\ &\geq \|\underline{y} - \underline{u}\|^2, && (\|\underline{u} - \underline{w}\|^2 \geq 0, \text{ Theorem 1 in Inner Products and Orthogonality})\end{aligned}$$

with equality exactly when $\|\underline{u} - \underline{w}\| = 0$, that is, when $\underline{w} = \underline{u}$ (Theorem 1 in Inner Products and Orthogonality, part 1). Taking nonnegative square roots preserves the inequality. $\square$

Example 0.6 (No point of the line is closer). In Example 0.4, $y = (3, 1, 2)$ has projection $\underline{u} = (2, 2, 0)$ onto $\mathcal{S} = \text{span}\{(1, 1, 0)\}$, at distance $\|(1, -1, 2)\| = \sqrt{6}$. Another point of $\mathcal{S}$, such as $\underline{w} = (3, 3, 0)$, is farther:

$$\begin{aligned}\|\underline{y} - \underline{w}\| &= \|(0, -2, 2)\| \\ &= \sqrt{8} \\ &> \sqrt{6},\end{aligned}$$

and the difference of squares is

$$\begin{aligned}\|\underline{u} - \underline{w}\|^2 &= \|(-1, -1, 0)\|^2 \\ &= 2 \\ &= 8 - 6,\end{aligned}$$

by the Pythagorean theorem (Theorem 2 in Inner Products and Orthogonality), since $\underline{y} - \underline{u} \in \mathcal{S}^\perp$ and $\underline{u} - \underline{w} \in \mathcal{S}$.

Theorem 0.4 (Projecting with an orthonormal basis). *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^p$ with $\dim(\mathcal{S}) = r \geq 1$, and let $\mathbf{Q}$ be the $p \times r$ matrix whose columns $q_1, \dots, q_r$ are an orthonormal basis of $\mathcal{S}$ (Definition 10 in Inner Products and Orthogonality, Corollary 4 in Inner Products and Orthogonality). Then for every $y \in \mathbb{R}^p$, $\mathbf{Q}\mathbf{Q}^\top y$ is the orthogonal projection of y onto $\mathcal{S}$, and $\mathbf{Q}\mathbf{Q}^\top$ is an orthogonal projection matrix (Definition 8 in Special Matrices and Decompositions).*

i Proof

Proof. $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_r$. Write q_{ki} for entry k of $\tilde{q}_i$, which is entry (k, i) of $\mathbf{Q}$. Then

$$\begin{aligned} (\mathbf{Q}^\top \mathbf{Q})_{ij} &= \sum_{k=1}^p (\mathbf{Q}^\top)_{ik} q_{kj} \quad (\text{Definition 7 in Matrices}) \\ &= \sum_{k=1}^p q_{ki} q_{kj} \quad (\text{Definition 3 in Matrices}) \\ &= \tilde{q}_i \cdot \tilde{q}_j, \quad (\text{Definition 8 in Vectors}) \end{aligned}$$

which is 1 if $i = j$ and 0 otherwise (Definition 23 in Vectors), so $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_r$ (Definition 12 in Matrices).

The split. Let $u = \mathbf{Q}\mathbf{Q}^\top y$ and $v = y - u$, so $y = u + v$. $u = \mathbf{Q}(\mathbf{Q}^\top y)$ (Theorem 5 in Matrices) is in $\mathcal{C}(\tilde{\mathbf{Q}})$ (Definition 1 in Column Space, Null Space and Rank-Nullity), which is $\text{span}\{\tilde{q}_1, \dots, \tilde{q}_r\} = \mathcal{S}$ (Theorem 1 in Column Space, Null Space and Rank-Nullity). For v ,

$$\begin{aligned} \mathbf{Q}^\top v &= \mathbf{Q}^\top (y - \mathbf{Q}\mathbf{Q}^\top y) \quad (\text{substitute}) \\ &= \mathbf{Q}^\top y - \mathbf{Q}^\top (\mathbf{Q}\mathbf{Q}^\top y) \quad (\text{Theorem 4 in Subspaces and Rank, with coefficients 1 and } -1) \\ &= \mathbf{Q}^\top y - (\mathbf{Q}^\top \mathbf{Q}) \mathbf{Q}^\top y \quad (\text{Theorem 5 in Matrices}) \\ &= \mathbf{Q}^\top y - \mathbf{I}_r \mathbf{Q}^\top y \quad (\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_r, \text{ from the first part of the proof}) \\ &= \mathbf{Q}^\top y - \mathbf{Q}^\top y \quad (\text{Theorem 8 in Matrices}) \\ &= \mathbf{0}_r, \quad (\text{arithmetic}) \end{aligned}$$

so $v \in \mathcal{N}(\mathbf{Q}^\top)$, and

$$\begin{aligned} \mathcal{N}(\mathbf{Q}^\top) &= \mathcal{C}(\mathbf{Q})^\perp \\ &= \mathcal{S}^\perp \end{aligned}$$

(Theorem 7 in Direct Sums and Orthogonal Complements). By the uniqueness in Theorem 8 in Direct Sums and Orthogonal Complements, $\tilde{u}$ is the orthogonal projection of $\tilde{y}$ onto $\mathcal{S}$.

$\mathbf{Q}\mathbf{Q}^\top$ is symmetric and idempotent.

$$\begin{aligned} (\mathbf{Q}\mathbf{Q}^\top)^\top &= (\mathbf{Q}^\top)^\top \mathbf{Q}^\top \quad (\text{Theorem 17 in Matrices}) \\ &= \mathbf{Q}\mathbf{Q}^\top, \quad (\text{transposing twice, Definition 3 in Matrices}) \end{aligned}$$

and

$$\begin{aligned} (\mathbf{Q}\mathbf{Q}^\top)(\mathbf{Q}\mathbf{Q}^\top) &= \mathbf{Q}(\mathbf{Q}^\top \mathbf{Q}) \mathbf{Q}^\top \quad (\text{Theorem 5 in Matrices}) \\ &= \mathbf{Q}\mathbf{I}_r \mathbf{Q}^\top \quad (\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_r, \text{ from the first part of the proof}) \\ &= \mathbf{Q}\mathbf{Q}^\top. \quad (\text{Theorem 8 in Matrices}) \end{aligned}$$

□

Example 0.7 (The projection matrix onto a line). For $\mathcal{S} = \text{span}\{(1, 1, 0)\}$, the single vector $\tilde{q}_1 = \frac{1}{\sqrt{2}}(1, 1, 0)$ is an orthonormal basis: it spans $\mathcal{S}$, it is nonzero and so linearly independent, and

$$\begin{aligned}\|\tilde{q}_1\| &= \sqrt{\frac{1}{2} + \frac{1}{2} + 0} \\ &= 1.\end{aligned}$$

With $\mathbf{Q} = [\tilde{q}_1]$,

$$\begin{aligned}\mathbf{Q}\mathbf{Q}^\top &= \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.\end{aligned}$$

Applied to $y = (3, 1, 2)$, it gives $(\frac{3+1}{2}, \frac{3+1}{2}, 0) = (2, 2, 0)$, the projection found in Example 0.4.

Theorem 0.5 (A projection matrix projects onto its column space, and is the only one that does).

1. If $\mathbf{P}$ is a $p \times p$ orthogonal projection matrix ([Definition 8 in Special Matrices and Decompositions](#)), then for every $y \in \mathbb{R}^p$, $\mathbf{P}y$ is the orthogonal projection of y onto $\mathcal{C}(\mathbf{P})$ ([Definition 0.3](#)).
2. If $\mathbf{P}_1$ and $\mathbf{P}_2$ are $p \times p$ matrices such that, for every y , both $\mathbf{P}_1 y$ and $\mathbf{P}_2 y$ are the orthogonal projection of y onto the same subspace $\mathcal{S}$, then $\mathbf{P}_1 = \mathbf{P}_2$.

i Proof

Proof. Part 1. Write $y = \mathbf{P}y + (y - \mathbf{P}y)$. The first term is in $\mathcal{C}(\mathbf{P})$ ([Definition 1 in Column Space, Null Space and Rank-Nullity](#)). For the second,

$$\begin{aligned}
\mathbf{P}(\tilde{y} - \mathbf{P}\tilde{y}) &= \mathbf{P}\tilde{y} - \mathbf{P}(\mathbf{P}\tilde{y}) && \text{(Theorem 4 in Subspaces and Rank, with coefficients 1 and } -1) \\
&= \mathbf{P}\tilde{y} - (\mathbf{P}\mathbf{P})\tilde{y} && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{P}\tilde{y} - \mathbf{P}^2\tilde{y} && \text{(Definition 2 in Special Matrices and Decompositions)} \\
&= \mathbf{P}\tilde{y} - \mathbf{P}\tilde{y} && (\mathbf{P} \text{ is idempotent}) \\
&= \tilde{0}_p, && \text{(arithmetic)}
\end{aligned}$$

so $\tilde{y} - \mathbf{P}\tilde{y} \in \mathcal{N}(\mathbf{P})$. Since $\mathbf{P} = \mathbf{P}^\top$,

$$\begin{aligned}
\mathcal{N}(\mathbf{P}) &= \mathcal{N}(\mathbf{P}^\top) \\
&= \mathcal{C}(\mathbf{P})^\perp
\end{aligned}$$

(Theorem 7 in Direct Sums and Orthogonal Complements). By the uniqueness in Theorem 8 in Direct Sums and Orthogonal Complements, $\mathbf{P}\tilde{y}$ is the orthogonal projection of $\tilde{y}$ onto $\mathcal{C}(\mathbf{P})$.

Part 2. The orthogonal projection of a vector onto $\mathcal{S}$ is unique (Definition 0.3), so $\mathbf{P}_1\tilde{e}_j = \mathbf{P}_2\tilde{e}_j$ for each indicator vector $\tilde{e}_j$ (Definition 18 in Vectors). $\mathbf{P}_1\tilde{e}_j$ is column j of $\mathbf{P}_1$: by Theorem 3 in Subspaces and Rank it is the combination of the columns of $\mathbf{P}_1$ with coefficients the entries of $\tilde{e}_j$, which are 1 in place j and 0 elsewhere (Definition 18 in Vectors). Likewise for $\mathbf{P}_2$, so the two matrices have the same columns. $\square$

Example 0.8 (Two routes to the same projection matrix). The matrix $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ of Example 12 in Special Matrices and Decompositions is an orthogonal projection matrix, and its column space is $\text{span}\{(1, 0)\}$, the horizontal axis. By part 1, $\mathbf{P}(v_1, v_2) = (v_1, 0)$ is the orthogonal projection onto that axis: the remainder $(0, v_2)$ is orthogonal to $(1, 0)$. The vector $\tilde{q}_1 = (1, 0)$ has length 1 and spans the axis, so it is an orthonormal basis of it, and with $\mathbf{B} = [\tilde{q}_1]$,

$$\begin{aligned}
\mathbf{B}\mathbf{B}^\top &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \\
&= \mathbf{P},
\end{aligned}$$

as part 2 and Theorem 0.4 require.

The oblique projection $\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ of Example 12 in Special Matrices and Decompositions also maps every vector into the horizontal axis. It sends $(0, 1)$ to $(1, 0)$, but the remainder $(0, 1) - (1, 0) = (-1, 1)$ is not orthogonal to the axis, so $(1, 0)$ is not the orthogonal projection of $(0, 1)$ onto the axis, which is $\mathbf{P}(0, 1) = (0, 0)$.

Theorem 0.6 (The hat matrix projects onto the column space of the design matrix). *If $\mathbf{X}$ is an $n \times p$ design matrix with $\text{rank}(\mathbf{X}) = p$, then for every $y \in \mathbb{R}^n$, $\mathbf{H}y$ (Definition 0.2) is the orthogonal projection of y onto $\mathcal{C}(\mathbf{X})$. So $\mathbf{H} = \mathbf{Q}\mathbf{Q}^\top$ for any $n \times p$ matrix $\mathbf{Q}$ whose columns are an orthonormal basis of $\mathcal{C}(\mathbf{X})$.*

i Proof

Proof. $\mathbf{H}$ is an orthogonal projection matrix (Theorem 0.2), so $\mathbf{H}y$ is the orthogonal projection of y onto $\mathcal{C}(\mathbf{H})$ (Theorem 0.5, part 1). It remains to show $\mathcal{C}(\mathbf{H}) = \mathcal{C}(\mathbf{X})$.

$\mathcal{C}(\mathbf{H}) \subseteq \mathcal{C}(\mathbf{X})$. $\mathbf{H}\tilde{z} = \mathbf{X}((\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\tilde{z})$ (Theorem 5 in Matrices), which is in $\mathcal{C}(\mathbf{X})$ (Definition 1 in Column Space, Null Space and Rank-Nullity).

$\mathcal{C}(\mathbf{X}) \subseteq \mathcal{C}(\mathbf{H})$. First,

$$\begin{aligned} \mathbf{H}\mathbf{X} &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{X} && \text{(Definition 0.2)} \\ &= \mathbf{X}((\mathbf{X}^\top\mathbf{X})^{-1}(\mathbf{X}^\top\mathbf{X})) && \text{(Theorem 5 in Matrices)} \\ &= \mathbf{X}\mathbf{I}_p && \text{(Definition 5 in Special Matrices and Decompositions)} \\ &= \mathbf{X}. && \text{(Theorem 8 in Matrices)} \end{aligned}$$

So for every $\tilde{b} \in \mathbb{R}^p$, $\mathbf{X}\tilde{b} = (\mathbf{H}\mathbf{X})\tilde{b}$, since $\mathbf{H}\mathbf{X} = \mathbf{X}$, and $(\mathbf{H}\mathbf{X})\tilde{b} = \mathbf{H}(\mathbf{X}\tilde{b})$ (Theorem 5 in Matrices), which is in $\mathcal{C}(\mathbf{H})$.

$\mathcal{C}(\mathbf{X})$ has dimension $p \geq 1$ (Theorem 6 in Column Space, Null Space and Rank-Nullity), so an orthonormal basis of it has p vectors (Definition 9 in Subspaces and Rank), and $\mathbf{Q}\mathbf{Q}^\top$ also gives the orthogonal projection onto $\mathcal{C}(\mathbf{X})$ (Theorem 0.4, with n in place of p and $r = p$). By Theorem 0.5, part 2, $\mathbf{H} = \mathbf{Q}\mathbf{Q}^\top$. $\square$

Example 0.9 (The intercept-only hat matrix, from an orthonormal basis). For the intercept-only design $\mathbf{X} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ of Example 0.2, $\mathcal{C}(\mathbf{X}) = \text{span}\{(1, 1)\}$, with orthonormal basis $\tilde{q}_1 = \frac{1}{\sqrt{2}}(1, 1)$, since

$$\begin{aligned} \|\tilde{q}_1\| &= \sqrt{\frac{1}{2} + \frac{1}{2}} \\ &= 1. \end{aligned}$$

With $\mathbf{Q} = [\tilde{q}_1]$,

$$\begin{aligned} \mathbf{Q}\mathbf{Q}^\top &= \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \end{aligned}$$

the hat matrix found in Example 0.2. The fitted values $(\bar{y}, \bar{y})$ are the closest point to y on the line of constant vectors (Theorem 0.3), and the residuals $(y_1 - \bar{y}, y_2 - \bar{y})$ are orthogonal to $(1, 1)$:

$$\begin{aligned}(1, 1) \cdot (y_1 - \bar{y}, y_2 - \bar{y}) &= y_1 + y_2 - 2\bar{y} \\ &= 0,\end{aligned}$$

since $2\bar{y} = y_1 + y_2$.

Generalized inverses and linear systems

i Source

This section is adapted from Zhou [5], used under the MIT License (see the license text in Section 2 in Subspaces and Rank). The source defers the existence of the Moore-Penrose inverse to the singular value decomposition; the proof here builds it from a rank factorization (Theorem 15 in Column Space, Null Space and Rank-Nullity) instead, and the section adds the projection $\mathbf{XGX}^\top$ for any generalized inverse $\mathbf{G}$ of $\mathbf{X}^\top\mathbf{X}$, which the source does not cover. Theorem 0.13 restores the source's formulas for the Moore-Penrose inverse of a matrix of full column rank or full row rank, and the fourth equivalent statement in Theorem 0.14 restores the source's rank test for consistency. These parts of the source are not part of this section:

- the QR forms of the full-rank formulas
- $\text{rank}(\mathbf{A}^-) \geq \text{rank}(\mathbf{A})$, $\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{AA}^-)$, $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{AA}^-) = \text{rank}(\mathbf{A}^-\mathbf{A})$, and $\text{rank}(\mathbf{A}^+) = \text{rank}(\mathbf{A})$
- the remarks that $\mathbf{AA}^-$ and $\mathbf{A}^-\mathbf{A}$ are projections, oblique in general and orthogonal for $\mathbf{A}^+$
- the reflexive generalized inverse (conditions 1 and 2 alone), and the generalized inverses of $\mathbf{A}^\top$ and $k\mathbf{A}$
- the operation counts for the triangular solves, LU and QR, and the comparison of the Cholesky factorization with LU

Theorem 0.7 (A scalar factor passes through a matrix product). *For an $m \times k$ matrix $\mathbf{A}$, a $k \times n$ matrix $\mathbf{B}$ and a number c ,*

$$\begin{aligned}\mathbf{A}(c\mathbf{B}) &= c(\mathbf{AB}) \\ &= (c\mathbf{A})\mathbf{B}.\end{aligned}$$

In particular, $\mathbf{A}(\mathbf{B} - \mathbf{C}) = \mathbf{AB} - \mathbf{AC}$ and $(\mathbf{B} - \mathbf{C})\mathbf{D} = \mathbf{BD} - \mathbf{CD}$ for matrices of compatible sizes, where $\mathbf{B} - \mathbf{C} = \mathbf{B} + (-1)\mathbf{C}$ (Theorem 4 in Matrices, Definition 6 in

Matrices).

i Proof

Proof. $\mathbf{A}(c\mathbf{B}) = c(\mathbf{AB})$. For each entry (i, j) ,

$$\begin{aligned} (\mathbf{A}(c\mathbf{B}))_{ij} &= \sum_{l=1}^k a_{il} (c\mathbf{B})_{lj} \quad (\text{Definition 7 in Matrices}) \\ &= \sum_{l=1}^k a_{il} (c b_{lj}) \quad (\text{Definition 6 in Matrices}) \\ &= \sum_{l=1}^k c (a_{il} b_{lj}) \quad (\text{reorder the product of numbers}) \\ &= c \sum_{l=1}^k a_{il} b_{lj} \quad (\text{factor } c \text{ out of the sum}) \\ &= c (\mathbf{AB})_{ij} \quad (\text{Definition 7 in Matrices}) \\ &= (c(\mathbf{AB}))_{ij}. \quad (\text{Definition 6 in Matrices}) \end{aligned}$$

$(c\mathbf{A})\mathbf{B} = c(\mathbf{AB})$. For each entry (i, j) ,

$$\begin{aligned} ((c\mathbf{A})\mathbf{B})_{ij} &= \sum_{l=1}^k (c a_{il}) b_{lj} \quad (\text{Definition 7 in Matrices, Definition 6 in Matrices}) \\ &= c \sum_{l=1}^k a_{il} b_{lj} \quad (\text{factor } c \text{ out of the sum}) \\ &= (c(\mathbf{AB}))_{ij}. \quad (\text{Definition 7 in Matrices, Definition 6 in Matrices}) \end{aligned}$$

Differences.

$$\begin{aligned} \mathbf{A}(\mathbf{B} - \mathbf{C}) &= \mathbf{A}(\mathbf{B} + (-1)\mathbf{C}) \quad (\text{the difference as a sum}) \\ &= \mathbf{AB} + \mathbf{A}((-1)\mathbf{C}) \quad (\text{Theorem 6 in Matrices}) \\ &= \mathbf{AB} + (-1)(\mathbf{AC}) \quad (\text{first part}) \\ &= \mathbf{AB} - \mathbf{AC}, \quad (\text{the difference as a sum}) \end{aligned}$$

and $(\mathbf{B} - \mathbf{C})\mathbf{D} = \mathbf{BD} - \mathbf{CD}$ the same way, using the second law in [Theorem 6 in Matrices](#) and the second part. A column vector is a matrix with one column, so all of this applies with $\mathbf{B}$, $\mathbf{C}$ or $\mathbf{D}$ a vector. $\square$

Example 0.10 (Pulling a scalar out of a product). With $\mathbf{A} = \begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$:

$$\begin{aligned}\mathbf{A} (2 \mathbf{B}) &= \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \end{bmatrix} \\ &= 22,\end{aligned}$$

and

$$\begin{aligned}2 (\mathbf{AB}) &= 2 (3 + 8) \\ &= 22.\end{aligned}$$

Definition 0.4 (Generalized inverse). A **generalized inverse** [6] of an $m \times n$ matrix $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{G}$ such that

$$\underset{m \times n}{\mathbf{A}} \underset{n \times m}{\mathbf{G}} \underset{m \times n}{\mathbf{A}} = \mathbf{A}.$$

Example 0.11 (A matrix with many generalized inverses). Let $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$. Then

$$\begin{aligned}\mathbf{AGA} &= \begin{bmatrix} g_{11} & g_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad (\text{Definition 7 in Matrices, for } \mathbf{AG}) \\ &= \begin{bmatrix} g_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad (\text{Definition 7 in Matrices})\end{aligned}$$

which equals $\mathbf{A}$ exactly when $g_{11} = 1$. So every matrix $\begin{bmatrix} 1 & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$ is a generalized inverse of $\mathbf{A}$: a generalized inverse need not be unique. The zero matrix is not one, since its $(1, 1)$ entry is 0.

Theorem 0.8 (An invertible matrix has only one generalized inverse). *If $\mathbf{A}$ is an invertible $p \times p$ matrix (Definition 6 in Special Matrices and Decompositions), its only generalized inverse is $\mathbf{A}^{-1}$.*

i Proof

Proof.

$$\begin{aligned}\mathbf{AA}^{-1}\mathbf{A} &= \mathbf{I}_p\mathbf{A} \\ &= \mathbf{A}\end{aligned}$$

(Definition 5 in Special Matrices and Decompositions, Theorem 8 in Matrices), so $\mathbf{A}^{-1}$ is

a generalized inverse. If $\mathbf{G}$ is any generalized inverse, then

$$\begin{aligned}
\mathbf{G} &= \mathbf{I}_p \mathbf{G} \mathbf{I}_p && \text{(Theorem 8 in Matrices)} \\
&= (\mathbf{A}^{-1} \mathbf{A}) \mathbf{G} (\mathbf{A} \mathbf{A}^{-1}) && \text{(Definition 5 in Special Matrices and Decompositions)} \\
&= \mathbf{A}^{-1} (\mathbf{A} \mathbf{G} \mathbf{A}) \mathbf{A}^{-1} && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{A}^{-1} \mathbf{A} \mathbf{A}^{-1} && \text{(Definition 0.4)} \\
&= \mathbf{I}_p \mathbf{A}^{-1} && \text{(Definition 5 in Special Matrices and Decompositions)} \\
&= \mathbf{A}^{-1}. && \text{(Theorem 8 in Matrices)}
\end{aligned}$$

□

Example 0.12 (The generalized inverse of an invertible matrix). For $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$ of Example 5 in Special Matrices and Decompositions, the only generalized inverse is $\mathbf{A}^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0 & 1 \end{bmatrix}$. The matrix of Example 0.11, by contrast, has infinitely many, so by Theorem 0.8 it is singular.

Theorem 0.9 (A generalized inverse solves every solvable system). *Let $\mathbf{A}$ be $m \times n$, let $\mathbf{G}$ be a generalized inverse of $\mathbf{A}$ (Definition 0.4), and let $\tilde{b} \in \mathbb{R}^m$. If $\mathbf{A}\tilde{x} = \tilde{b}$ has a solution, then $\mathbf{G}\tilde{b}$ is a solution.*

i Proof

Proof. Let $\tilde{x}_0$ be a solution, so $\mathbf{A}\tilde{x}_0 = \tilde{b}$. Then

$$\begin{aligned}
\mathbf{A}(\mathbf{G}\tilde{b}) &= \mathbf{A}\mathbf{G}(\mathbf{A}\tilde{x}_0) && \text{(substitute } \tilde{b} = \mathbf{A}\tilde{x}_0) \\
&= (\mathbf{A}\mathbf{G}\mathbf{A})\tilde{x}_0 && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{A}\tilde{x}_0 && \text{(Definition 0.4)} \\
&= \tilde{b}. && (\tilde{x}_0 \text{ is a solution})
\end{aligned}$$

□

Example 0.13 (Solving with a generalized inverse). With $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and its generalized inverse $\mathbf{G} = \begin{bmatrix} 1 & 5 \\ 7 & 0 \end{bmatrix}$ (Example 0.11), the system $\mathbf{A}\tilde{x} = (2, 0)$ has the solution $(2, 0)$, and $\mathbf{G}(2, 0) = (2, 14)$ is a solution too: $\mathbf{A}(2, 14) = (2, 0)$. The system $\mathbf{A}\tilde{x} = (2, 1)$ has

no solution, since the second entry of $\mathbf{A}\tilde{x}$ is always 0; there $\mathbf{G}(2,1) = (7,14)$ gives $\mathbf{A}(7,14) = (7,0) \neq (2,1)$.

Definition 0.5 (Moore-Penrose inverse). A **Moore-Penrose inverse** [7] of an $m \times n$ matrix $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{G}$ satisfying all four conditions

1. $\mathbf{AGA} = \mathbf{A}$;
2. $\mathbf{GAG} = \mathbf{G}$;
3. $(\mathbf{AG})^\top = \mathbf{AG}$;
4. $(\mathbf{GA})^\top = \mathbf{GA}$.

Example 0.14 (Checking the four conditions). For $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, the matrix $\mathbf{G} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ satisfies all four:

$$\begin{aligned}\mathbf{AG} &= \mathbf{GA} \\ &= \mathbf{A}\end{aligned}$$

and $\mathbf{A}^2 = \mathbf{A}$, so conditions 1 and 2 read $\mathbf{A} = \mathbf{A}$, and conditions 3 and 4 hold because $\mathbf{A}$ is symmetric.

The generalized inverse $\mathbf{G}' = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ of $\mathbf{A}$ (Example 0.11) is not a Moore-Penrose inverse:

$\mathbf{AG}' = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, which is not symmetric, so condition 3 fails.

Theorem 0.10 (A Moore-Penrose inverse is a generalized inverse). *If $\mathbf{G}$ is a Moore-Penrose inverse of an $m \times n$ matrix $\mathbf{A}$ (Definition 0.5), then $\mathbf{G}$ is a generalized inverse of $\mathbf{A}$ (Definition 0.4).*

Proof

Proof. A generalized inverse of $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{G}$ with $\mathbf{AGA} = \mathbf{A}$. A Moore-Penrose inverse is an $n \times m$ matrix, and condition 1 of Definition 0.5 is $\mathbf{AGA} = \mathbf{A}$. $\square$

Example 0.15 (The Moore-Penrose inverse of a rank-one matrix). The 3×2 matrix $\mathbf{A}$ below has rank 1, because its second column is twice its first, so it has no ordinary inverse. The function `MASS::ginv()` computes its Moore-Penrose inverse $\mathbf{G}$. The code checks the four conditions of Definition 0.5, of which the first is the equation that makes $\mathbf{G}$ a generalized inverse.

```

A <- cbind(c(1, 2, 3), c(2, 4, 6))
G <- MASS::ginv(A)
conditions <- c(
  "1: AGA = A" = isTRUE(all.equal(A %*% G %*% A, A)),
  "2: GAG = G" = isTRUE(all.equal(G %*% A %*% G, G)),
  "3: AG symmetric" = isTRUE(all.equal(A %*% G, t(A %*% G))),
  "4: GA symmetric" = isTRUE(all.equal(G %*% A, t(G %*% A)))
)
conditions
#>      1: AGA = A      2: GAG = G 3: AG symmetric 4: GA symmetric
#>      TRUE          TRUE          TRUE          TRUE
round(G, 4)
#>      [,1]  [,2]  [,3]
#> [1,] 0.0143 0.0286 0.0429
#> [2,] 0.0286 0.0571 0.0857

```

All 4 of 4 conditions hold, so $\mathbf{G}$ is a Moore-Penrose inverse of $\mathbf{A}$. By Theorem 0.10 it is also a generalized inverse, which is condition 1: $\mathbf{AGA} = \mathbf{A}$ up to rounding error.

Theorem 0.11 (A matrix has at most one Moore-Penrose inverse). *If $\mathbf{G}_1$ and $\mathbf{G}_2$ are both Moore-Penrose inverses of an $m \times n$ matrix $\mathbf{A}$ (Definition 0.5), then $\mathbf{G}_1 = \mathbf{G}_2$.*

i Proof

Proof. $\mathbf{AG}_1 = \mathbf{AG}_2$.

$$\begin{aligned}
\mathbf{AG}_1 &= (\mathbf{AG}_1)^\top && \text{(condition 3 for } \mathbf{G}_1) \\
&= \mathbf{G}_1^\top \mathbf{A}^\top && \text{(Theorem 17 in Matrices)} \\
&= \mathbf{G}_1^\top (\mathbf{AG}_2 \mathbf{A})^\top && \text{(condition 1 for } \mathbf{G}_2) \\
&= \mathbf{G}_1^\top \mathbf{A}^\top \mathbf{G}_2^\top \mathbf{A}^\top && \text{(Theorem 17 in Matrices, twice)} \\
&= (\mathbf{G}_1^\top \mathbf{A}^\top) (\mathbf{G}_2^\top \mathbf{A}^\top) && \text{(Theorem 5 in Matrices)} \\
&= (\mathbf{AG}_1)^\top (\mathbf{AG}_2)^\top && \text{(Theorem 17 in Matrices, twice)} \\
&= (\mathbf{AG}_1)(\mathbf{AG}_2) && \text{(condition 3 for } \mathbf{G}_1 \text{ and } \mathbf{G}_2) \\
&= (\mathbf{AG}_1 \mathbf{A}) \mathbf{G}_2 && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{AG}_2. && \text{(condition 1 for } \mathbf{G}_1)
\end{aligned}$$

$$\mathbf{G}_1 \mathbf{A} = \mathbf{G}_2 \mathbf{A}.$$

$$\begin{aligned}
\mathbf{G}_1 \mathbf{A} &= (\mathbf{G}_1 \mathbf{A})^\top && \text{(condition 4 for } \mathbf{G}_1) \\
&= \mathbf{A}^\top \mathbf{G}_1^\top && \text{(Theorem 17 in Matrices)} \\
&= (\mathbf{A} \mathbf{G}_2 \mathbf{A})^\top \mathbf{G}_1^\top && \text{(condition 1 for } \mathbf{G}_2) \\
&= \mathbf{A}^\top \mathbf{G}_2^\top \mathbf{A}^\top \mathbf{G}_1^\top && \text{(Theorem 17 in Matrices, twice)} \\
&= (\mathbf{A}^\top \mathbf{G}_2^\top) (\mathbf{A}^\top \mathbf{G}_1^\top) && \text{(Theorem 5 in Matrices)} \\
&= (\mathbf{G}_2 \mathbf{A})^\top (\mathbf{G}_1 \mathbf{A})^\top && \text{(Theorem 17 in Matrices, twice)} \\
&= (\mathbf{G}_2 \mathbf{A})(\mathbf{G}_1 \mathbf{A}) && \text{(condition 4 for } \mathbf{G}_2 \text{ and } \mathbf{G}_1) \\
&= \mathbf{G}_2 (\mathbf{A} \mathbf{G}_1 \mathbf{A}) && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{G}_2 \mathbf{A}. && \text{(condition 1 for } \mathbf{G}_1)
\end{aligned}$$

Conclusion.

$$\begin{aligned}
\mathbf{G}_1 &= \mathbf{G}_1 \mathbf{A} \mathbf{G}_1 && \text{(condition 2 for } \mathbf{G}_1) \\
&= \mathbf{G}_1 (\mathbf{A} \mathbf{G}_1) && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{G}_1 (\mathbf{A} \mathbf{G}_2) && \text{(the first step)} \\
&= (\mathbf{G}_1 \mathbf{A}) \mathbf{G}_2 && \text{(Theorem 5 in Matrices)} \\
&= (\mathbf{G}_2 \mathbf{A}) \mathbf{G}_2 && \text{(the second step)} \\
&= \mathbf{G}_2 \mathbf{A} \mathbf{G}_2 && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{G}_2. && \text{(condition 2 for } \mathbf{G}_2)
\end{aligned}$$

□

Example 0.16 (Only one of the many generalized inverses qualifies). The matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ has infinitely many generalized inverses $\begin{bmatrix} 1 & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$ (Example 0.11), and $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is a Moore-Penrose inverse (Example 0.14). By Theorem 0.11 it is the only one, so every other choice of g_{12}, g_{21}, g_{22} breaks condition 2, 3 or 4. For instance, $g_{21} = 1$ (the rest 0) gives $\mathbf{G} \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, which is not symmetric.

Theorem 0.12 (Every matrix has a Moore-Penrose inverse). *Every $m \times n$ matrix $\mathbf{A}$ has a Moore-Penrose inverse (Definition 0.5), written $\mathbf{A}^+$; it is unique by Theorem 0.11.*

- If $\mathbf{A} = \mathbf{0}_{m \times n}$, then $\mathbf{A}^+ = \mathbf{0}_{n \times m}$.
- If $\text{rank}(\mathbf{A}) = r \geq 1$ and $\mathbf{A} = \mathbf{C} \mathbf{R}$ is a rank factorization (Definition 11 in Column Space, Null Space and Rank-Nullity), then

$$\underbrace{\mathbf{A}^+}_{n \times m} = \underbrace{\mathbf{R}^\top}_{n \times r} \underbrace{(\mathbf{R}\mathbf{R}^\top)^{-1}}_{r \times r} \underbrace{(\mathbf{C}^\top \mathbf{C})^{-1}}_{r \times r} \underbrace{\mathbf{C}^\top}_{r \times m}.$$

i Proof

Proof. The two cases cover every matrix. If $\mathbf{A} \neq \mathbf{0}_{m \times n}$, it has a nonzero column, which is linearly independent on its own ($c \tilde{v} = 0$ with $\tilde{v} \neq 0$ forces $c = 0$), so $\text{rank}(\mathbf{A}) \geq 1$ ([Definition 2 in Subspaces and Rank](#)).

The zero matrix. With $\mathbf{A} = \mathbf{0}_{m \times n}$ and $\mathbf{G} = \mathbf{0}_{n \times m}$, every product in conditions 1 to 4 is a zero matrix, and a zero matrix equals its own transpose, so all four conditions hold.

The two inverses exist. Now suppose $\text{rank}(\mathbf{A}) = r \geq 1$. A rank factorization exists ([Theorem 15 in Column Space, Null Space and Rank-Nullity](#)). $\mathbf{C}$ is $m \times r$ and $\mathbf{R}$ is $r \times n$. Both have rank r :

$$\begin{aligned} r &= \text{rank}(\mathbf{C}\mathbf{R}) \quad (\mathbf{A} = \mathbf{C}\mathbf{R}) \\ &\leq \text{rank}(\mathbf{C}) \quad (\text{Theorem 8 in Column Space, Null Space and Rank-Nullity}) \\ &\leq r, \quad (\text{Corollary 1 in Column Space, Null Space and Rank-Nullity, } \mathbf{C} \text{ has } r \text{ columns}) \end{aligned}$$

and, using $(\mathbf{C}\mathbf{R})^\top = \mathbf{R}^\top \mathbf{C}^\top$ ([Theorem 17 in Matrices](#)),

$$\begin{aligned} r &= \text{rank}(\mathbf{R}^\top \mathbf{C}^\top) \quad (\text{Theorem 9 in Column Space, Null Space and Rank-Nullity}) \\ &\leq \text{rank}(\mathbf{R}^\top) \quad (\text{Theorem 8 in Column Space, Null Space and Rank-Nullity}) \\ &\leq r. \quad (\text{Corollary 1 in Column Space, Null Space and Rank-Nullity, } \mathbf{R}^\top \text{ has } r \text{ columns}) \end{aligned}$$

So $\mathbf{C}$ and $\mathbf{R}^\top$ have full column rank r , and $\mathbf{C}^\top \mathbf{C}$ and $\mathbf{R}\mathbf{R}^\top = (\mathbf{R}^\top)^\top \mathbf{R}^\top$ are invertible ([Theorem 0.1](#)). They are symmetric:

$$\begin{aligned} (\mathbf{C}^\top \mathbf{C})^\top &= \mathbf{C}^\top (\mathbf{C}^\top)^\top \\ &= \mathbf{C}^\top \mathbf{C} \end{aligned}$$

([Theorem 17 in Matrices](#), [Definition 3 in Matrices](#)), and likewise for $\mathbf{R}\mathbf{R}^\top$. Write $\mathbf{S} \stackrel{\text{def}}{=} (\mathbf{R}\mathbf{R}^\top)^{-1}$ and $\mathbf{T} \stackrel{\text{def}}{=} (\mathbf{C}^\top \mathbf{C})^{-1}$, both symmetric ([Corollary 1 in Special Matrices and Decompositions](#)), and $\mathbf{G} \stackrel{\text{def}}{=} \mathbf{R}^\top \mathbf{S} \mathbf{T} \mathbf{C}^\top$.

Two products.

$$\begin{aligned} \mathbf{A}\mathbf{G} &= \mathbf{C}\mathbf{R}\mathbf{R}^\top \mathbf{S} \mathbf{T} \mathbf{C}^\top \quad (\text{substitute } \mathbf{A} \text{ and } \mathbf{G}) \\ &= \mathbf{C}(\mathbf{R}\mathbf{R}^\top) \mathbf{S} \mathbf{T} \mathbf{C}^\top \quad (\text{Theorem 5 in Matrices}) \\ &= \mathbf{C}\mathbf{I}_r \mathbf{T} \mathbf{C}^\top \quad (\text{Definition 5 in Special Matrices and Decompositions}) \\ &= \mathbf{C}\mathbf{T}\mathbf{C}^\top, \quad (\text{Theorem 8 in Matrices}) \end{aligned}$$

$$\begin{aligned}
\mathbf{GA} &= \mathbf{R}^\top \mathbf{S} \mathbf{T} \mathbf{C}^\top \mathbf{C} \mathbf{R} && \text{(substitute } \mathbf{G} \text{ and } \mathbf{A}) \\
&= \mathbf{R}^\top \mathbf{S} \mathbf{T} (\mathbf{C}^\top \mathbf{C}) \mathbf{R} && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{R}^\top \mathbf{S} \mathbf{I}_r \mathbf{R} && \text{(Definition 5 in Special Matrices and Decompositions)} \\
&= \mathbf{R}^\top \mathbf{S} \mathbf{R}. && \text{(Theorem 8 in Matrices)}
\end{aligned}$$

Conditions 3 and 4.

$$\begin{aligned}
(\mathbf{AG})^\top &= (\mathbf{CTC}^\top)^\top && \text{(first product)} \\
&= (\mathbf{C} (\mathbf{TC}^\top))^\top && \text{(Theorem 5 in Matrices)} \\
&= (\mathbf{TC}^\top)^\top \mathbf{C}^\top && \text{(Theorem 17 in Matrices)} \\
&= (\mathbf{C}^\top)^\top \mathbf{T}^\top \mathbf{C}^\top && \text{(Theorem 17 in Matrices)} \\
&= \mathbf{C} \mathbf{T}^\top \mathbf{C}^\top && \text{(Definition 3 in Matrices)} \\
&= \mathbf{CTC}^\top, && \text{(\mathbf{T} is symmetric)}
\end{aligned}$$

$$\begin{aligned}
(\mathbf{GA})^\top &= (\mathbf{R}^\top \mathbf{S} \mathbf{R})^\top && \text{(second product)} \\
&= (\mathbf{R}^\top (\mathbf{S} \mathbf{R}))^\top && \text{(Theorem 5 in Matrices)} \\
&= (\mathbf{S} \mathbf{R})^\top (\mathbf{R}^\top)^\top && \text{(Theorem 17 in Matrices)} \\
&= \mathbf{R}^\top \mathbf{S}^\top (\mathbf{R}^\top)^\top && \text{(Theorem 17 in Matrices)} \\
&= \mathbf{R}^\top \mathbf{S}^\top \mathbf{R} && \text{(Definition 3 in Matrices)} \\
&= \mathbf{R}^\top \mathbf{S} \mathbf{R}. && \text{(\mathbf{S} is symmetric)}
\end{aligned}$$

Condition 1.

$$\begin{aligned}
\mathbf{AGA} &= \mathbf{CTC}^\top \mathbf{C} \mathbf{R} && \text{(first product)} \\
&= \mathbf{C} \mathbf{T} (\mathbf{C}^\top \mathbf{C}) \mathbf{R} && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{C} \mathbf{I}_r \mathbf{R} && \text{(Definition 5 in Special Matrices and Decompositions)} \\
&= \mathbf{A}. && \text{(Theorem 8 in Matrices)}
\end{aligned}$$

Condition 2.

$$\begin{aligned}
\mathbf{GAG} &= \mathbf{R}^\top \mathbf{S} \mathbf{R} \mathbf{R}^\top \mathbf{S} \mathbf{T} \mathbf{C}^\top && \text{(second product)} \\
&= \mathbf{R}^\top \mathbf{S} (\mathbf{R} \mathbf{R}^\top) \mathbf{S} \mathbf{T} \mathbf{C}^\top && \text{(Theorem 5 in Matrices)} \\
&= \mathbf{R}^\top \mathbf{I}_r \mathbf{S} \mathbf{T} \mathbf{C}^\top && \text{(Definition 5 in Special Matrices and Decompositions)} \\
&= \mathbf{G}. && \text{(Theorem 8 in Matrices)}
\end{aligned}$$

□

Example 0.17 (The Moore-Penrose inverse of a rank-one matrix). For $\mathbf{A}$ in [Example 1 in Column Space, Null Space and Rank-Nullity](#), the rank factorization $\mathbf{C} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, $\mathbf{R} = \begin{bmatrix} 1 & -2 & -2 \end{bmatrix}$ ([Example 25 in Column Space, Null Space and Rank-Nullity](#)) gives

$$\begin{aligned}\mathbf{C}^\top \mathbf{C} &= 1 + 9 \\ &= 10\end{aligned}$$

and

$$\begin{aligned}\mathbf{R}\mathbf{R}^\top &= 1 + 4 + 4 \\ &= 9,\end{aligned}$$

so

$$\begin{aligned}\mathbf{A}^+ &= \frac{1}{9 \cdot 10} \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix} \begin{bmatrix} 1 & 3 \end{bmatrix} \\ &= \frac{1}{90} \begin{bmatrix} 1 & 3 \\ -2 & -6 \\ -2 & -6 \end{bmatrix}.\end{aligned}$$

As a check on condition 1,

$$\begin{aligned}\mathbf{A}\mathbf{A}^+ &= \mathbf{C}\mathbf{T}\mathbf{C}^\top \\ &= \frac{1}{10} \begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix},\end{aligned}$$

and its first row times $\mathbf{A}$ is $\frac{1}{10}(1 \cdot (1, -2, -2) + 3 \cdot (3, -6, -6)) = (1, -2, -2)$, the first row of $\mathbf{A}$; the second row is 3 times the first, like the second row of $\mathbf{A}$.

Theorem 0.13 (The Moore-Penrose inverse of a matrix of full column rank or full row rank). *Let $\mathbf{A}$ be an $m \times n$ [matrix](#).*

1. *If $\mathbf{A}$ has full column rank ([Definition 3 in Subspaces and Rank](#)), then $\mathbf{A}^\top \mathbf{A}$ is [invertible](#) and $\mathbf{A}^+ = ((\mathbf{A}^\top \mathbf{A}))^{-1} \mathbf{A}^\top$.*
2. *If $\text{rank}(\mathbf{A}) = m$, then $\mathbf{A}\mathbf{A}^\top$ is invertible and $\mathbf{A}^+ = \mathbf{A}^\top ((\mathbf{A}\mathbf{A}^\top))^{-1}$.*

i Proof

Proof. By Theorem [0.11](#), it is enough to show that the [matrix](#) $\mathbf{G}$ named in each part satisfies the four conditions of Definition [0.5](#).

Part 1. $\mathbf{A}^\top \mathbf{A}$ is [invertible](#) by Theorem 0.1. Let $\mathbf{G} = ((\mathbf{A}^\top \mathbf{A}))^{-1} \mathbf{A}^\top$. Then

$$\begin{aligned} \mathbf{GA} &= ((\mathbf{A}^\top \mathbf{A}))^{-1} \mathbf{A}^\top \mathbf{A} \quad (\text{definition of } \mathbf{G}) \\ &= \mathbf{I}_n. \quad (\text{Definition 5 in Special Matrices and Decompositions}) \end{aligned}$$

- Condition 4: $(\mathbf{GA})^\top = \mathbf{I}_n^\top = \mathbf{I}_n = \mathbf{GA}$.
- Condition 1: $\mathbf{AGA} = \mathbf{AI}_n = \mathbf{A}$.
- Condition 2: $\mathbf{GAG} = \mathbf{I}_n \mathbf{G} = \mathbf{G}$.
- Condition 3: $(\mathbf{AG})^\top = \mathbf{A} \left(((\mathbf{A}^\top \mathbf{A}))^{-1} \right)^\top \mathbf{A}^\top = \mathbf{A} \left((\mathbf{A}^\top \mathbf{A})^\top \right)^{-1} \mathbf{A}^\top = \mathbf{A} \left((\mathbf{A}^\top \mathbf{A}) \right)^{-1} \mathbf{A}^\top = \mathbf{AG}$, by Theorem 17 in Matrices, Theorem 2 in Special Matrices and Decompositions, and because $\mathbf{A}^\top \mathbf{A}$ is [symmetric](#).

Part 2. $\text{rank}(\mathbf{AA}^\top) = \text{rank}(\mathbf{A}) = m$ (Theorem 9 in Column Space, Null Space and Rank-Nullity), and $\mathbf{AA}^\top$ is $m \times m$, so it is invertible (Theorem 1 in Diagonalization and Cholesky). Let $\mathbf{G} = \mathbf{A}^\top ((\mathbf{AA}^\top))^{-1}$. Then $\mathbf{AG} = \mathbf{AA}^\top ((\mathbf{AA}^\top))^{-1} = \mathbf{I}_m$ (Definition 5 in Special Matrices and Decompositions).

- Condition 3: $(\mathbf{AG})^\top = \mathbf{I}_m^\top = \mathbf{I}_m = \mathbf{AG}$.
- Condition 1: $\mathbf{AGA} = \mathbf{I}_m \mathbf{A} = \mathbf{A}$.
- Condition 2: $\mathbf{GAG} = \mathbf{GI}_m = \mathbf{G}$.
- Condition 4: $(\mathbf{GA})^\top = \mathbf{A}^\top \left(((\mathbf{AA}^\top))^{-1} \right)^\top \mathbf{A} = \mathbf{A}^\top ((\mathbf{AA}^\top))^{-1} \mathbf{A} = \mathbf{GA}$, by Theorem 17 in Matrices, Theorem 2 in Special Matrices and Decompositions, and because $\mathbf{AA}^\top$ is [symmetric](#).

□

Example 0.18 (Moore-Penrose inverses of a column and a row). The 2×1 [matrix](#) $\mathbf{A} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ has [full column rank](#), and $\mathbf{A}^\top \mathbf{A} = [2]$, so

$$\begin{aligned} \mathbf{A}^+ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \end{aligned}$$

Then $\mathbf{A}^+ \mathbf{A} = [1]$, and $\mathbf{AA}^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ is [symmetric](#).

The 1×2 matrix $\mathbf{B} = \begin{bmatrix} 1 & 1 \end{bmatrix}$ has [rank](#) 1, which is its count of rows, and $\mathbf{BB}^\top = [2]$, so

$$\begin{aligned}\mathbf{B}^+ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{2} \\ &= \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.\end{aligned}$$

Definition 0.6 (Consistent linear system). A linear system (Definition 14 in Matrices) $\mathbf{A}\tilde{x} = \tilde{b}$, with $\mathbf{A}$ an $m \times n$ matrix and $\tilde{b} \in \mathbb{R}^m$, is **consistent** [8] if it has at least one solution $\tilde{x} \in \mathbb{R}^n$, and **inconsistent** otherwise.

Example 0.19 (A consistent system and an inconsistent one). With $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{A}(x_1, x_2) = (x_1, 0)$ (Definition 13 in Matrices).

- $\mathbf{A}\tilde{x} = (2, 0)$ is consistent: $\tilde{x} = (2, 0)$ is a solution.
- $\mathbf{A}\tilde{x} = (2, 1)$ is inconsistent: the second entry of $\mathbf{A}\tilde{x}$ is always 0, never 1.

Theorem 0.14 (When a linear system has a solution). Let $\mathbf{A}$ be $m \times n$, let $\tilde{b} \in \mathbb{R}^m$, and let $\mathbf{G}$ be any generalized inverse of $\mathbf{A}$ (Definition 0.4). The following statements are equivalent:

1. $\mathbf{A}\tilde{x} = \tilde{b}$ is consistent (Definition 0.6);
2. $\tilde{b} \in \mathcal{C}(\mathbf{A})$ (Definition 1 in Column Space, Null Space and Rank-Nullity);
3. $\mathbf{A}\mathbf{G}\tilde{b} = \tilde{b}$;
4. $\text{rank}(\begin{bmatrix} \mathbf{A} & \tilde{b} \end{bmatrix}) = \text{rank}(\mathbf{A})$, where $\begin{bmatrix} \mathbf{A} & \tilde{b} \end{bmatrix}$ is the $m \times (n+1)$ matrix whose last column is $\tilde{b}$.

i Proof

Proof. **1 and 2 are equivalent.** Statement 1 says $\tilde{b} = \mathbf{A}\tilde{x}$ for some $\tilde{x}$ (Definition 0.6), and $\mathcal{C}(\mathbf{A})$ is the set of vectors $\mathbf{A}\tilde{x}$ (Definition 1 in Column Space, Null Space and Rank-Nullity), so $\tilde{b}$ is in it exactly when $\tilde{b} = \mathbf{A}\tilde{x}$ for some $\tilde{x}$.

1 implies 3. By Theorem 0.9, $\mathbf{G}\tilde{b}$ is a solution, so $\mathbf{A}(\mathbf{G}\tilde{b}) = \tilde{b}$, which is statement 3 (Theorem 5 in Matrices).

3 implies 1. If $\mathbf{A}\mathbf{G}\tilde{b} = \tilde{b}$, then $\tilde{x} = \mathbf{G}\tilde{b}$ is a solution, since $\mathbf{A}(\mathbf{G}\tilde{b}) = (\mathbf{A}\mathbf{G})\tilde{b}$ (Theorem 5 in Matrices).

2 and 4 are equivalent. Every column of $\mathbf{A}$ is a column of $\begin{bmatrix} \mathbf{A} & \tilde{b} \end{bmatrix}$, so $\mathcal{C}(\mathbf{A}) \subseteq \mathcal{C}(\begin{bmatrix} \mathbf{A} & \tilde{b} \end{bmatrix})$.

- If $\tilde{b} \in \mathcal{C}(\mathbf{A})$, every combination of the columns of $\begin{bmatrix} \mathbf{A} & \tilde{b} \end{bmatrix}$ is a combination of

the columns of $\mathbf{A}$ (Definition 1 in Column Space, Null Space and Rank-Nullity), so the two column spaces (Definition 1 in Column Space, Null Space and Rank-Nullity) coincide and their ranks agree (Theorem 6 in Column Space, Null Space and Rank-Nullity).

- If the ranks agree, the two column spaces have the same dimension (Theorem 6 in Column Space, Null Space and Rank-Nullity), so they coincide (Theorem 11 in Subspaces and Rank), and $\tilde{b} \in \mathcal{C}([\mathbf{A} \ \tilde{b}]) = \mathcal{C}(\mathbf{A})$.

□

Example 0.20 (Testing consistency with a generalized inverse). For $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and the generalized inverse $\mathbf{G} = \begin{bmatrix} 1 & 5 \\ 7 & 0 \end{bmatrix}$, $\mathbf{AG} = \begin{bmatrix} 1 & 5 \\ 0 & 0 \end{bmatrix}$.

- $\tilde{b} = (2, 0)$:

$$\begin{aligned} \mathbf{AG}\tilde{b} &= (2, 0) \\ &= \tilde{b}, \end{aligned}$$

so the system is consistent.

- $\tilde{b} = (2, 1)$: $\mathbf{AG}\tilde{b} = (7, 0) \neq \tilde{b}$, so it is not, as Example 0.13 found directly.

Item 4 agrees. $\mathbf{A}$ has rank 1. For $\tilde{b} = (2, 0)$, $[\mathbf{A} \ \tilde{b}] = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ also has rank 1.

For $\tilde{b} = (2, 1)$, $[\mathbf{A} \ \tilde{b}] = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ has rank 2, because $(1, 0)$ and $(2, 1)$ are linearly independent.

Theorem 0.15 (All solutions of a linear system). *Let $\mathbf{A}$ be $m \times n$ with a generalized inverse $\mathbf{G}$ (Definition 0.4).*

1. $\mathcal{N}(\mathbf{A}) = \mathcal{C}(\mathbf{I}_n - \mathbf{GA})$: $\tilde{x}$ solves $\mathbf{A}\tilde{x} = \mathbf{0}_m$ exactly when $\tilde{x} = (\mathbf{I}_n - \mathbf{GA})\tilde{q}$ for some $\tilde{q} \in \mathbb{R}^n$.
2. If $\mathbf{A}\tilde{b} = \tilde{b}$ is consistent (Definition 0.6), then $\tilde{x}$ is a solution exactly when

$$\tilde{x} = \mathbf{G}\tilde{b} + (\mathbf{I}_n - \mathbf{GA})\tilde{q} \quad \text{for some } \tilde{q} \in \mathbb{R}^n :$$

one particular solution plus a vector of $\mathcal{N}(\mathbf{A})$.

i Proof

Proof. Part 1. First,

$$\begin{aligned}
 \mathbf{A}(\mathbf{I}_n - \mathbf{GA}) &= \mathbf{A}\mathbf{I}_n - \mathbf{A}(\mathbf{GA}) && \text{(Theorem 0.7)} \\
 &= \mathbf{A}\mathbf{I}_n - \mathbf{AGA} && \text{(Theorem 5 in Matrices)} \\
 &= \mathbf{A} - \mathbf{AGA} && \text{(Theorem 8 in Matrices)} \\
 &= \mathbf{A} - \mathbf{A} && \text{(Definition 0.4)} \\
 &= \mathbf{0}_{m \times n}, && \text{(arithmetic)}
 \end{aligned}$$

so $\mathbf{A}(\mathbf{I}_n - \mathbf{GA})q = \mathbf{0}_m$ for every q (Theorem 5 in Matrices), and $\mathcal{C}(\mathbf{I}_n - \mathbf{GA}) \subseteq \mathcal{N}(\mathbf{A})$. Conversely, if $\mathbf{A}\tilde{x} = \mathbf{0}_m$, then

$$\begin{aligned}
 (\mathbf{I}_n - \mathbf{GA})\tilde{x} &= \mathbf{I}_n\tilde{x} - (\mathbf{GA})\tilde{x} && \text{(Theorem 0.7, with } \tilde{x} \text{ as the last factor)} \\
 &= \tilde{x} - (\mathbf{GA})\tilde{x} && \text{(Theorem 8 in Matrices)} \\
 &= \tilde{x} - \mathbf{G}(\mathbf{A}\tilde{x}) && \text{(Theorem 5 in Matrices)} \\
 &= \tilde{x} - \mathbf{G}\mathbf{0}_m && (\mathbf{A}\tilde{x} = \mathbf{0}_m) \\
 &= \tilde{x}, && (\mathbf{G}\mathbf{0}_m = \mathbf{0}_n)
 \end{aligned}$$

so $\tilde{x}$ is in $\mathcal{C}(\mathbf{I}_n - \mathbf{GA})$, with $q = \tilde{x}$.

Part 2. $\mathbf{G}\tilde{b}$ is a solution (Theorem 0.9). For any $\tilde{x}$,

$$\begin{aligned}
 \mathbf{A}(\tilde{x} - \mathbf{G}\tilde{b}) &= \mathbf{A}\tilde{x} - \mathbf{A}(\mathbf{G}\tilde{b}) && \text{(Theorem 4 in Subspaces and Rank, with coefficients 1 and -1)} \\
 &= \mathbf{A}\tilde{x} - \mathbf{AG}\tilde{b} && \text{(Theorem 5 in Matrices)} \\
 &= \mathbf{A}\tilde{x} - \tilde{b}, && \text{(Theorem 0.14, statement 3)}
 \end{aligned}$$

so $\tilde{x}$ solves $\mathbf{A}\tilde{x} = \tilde{b}$ exactly when $\tilde{x} - \mathbf{G}\tilde{b} \in \mathcal{N}(\mathbf{A})$, that is, by part 1, exactly when $\tilde{x} - \mathbf{G}\tilde{b} = (\mathbf{I}_n - \mathbf{GA})q$ for some q . $\square$

Example 0.21 (All solutions of a consistent system). For $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{G} = \begin{bmatrix} 1 & 5 \\ 7 & 0 \end{bmatrix}$ and

$\tilde{b} = (2, 0)$: $\mathbf{G}\tilde{b} = (2, 14)$, $\mathbf{GA} = \begin{bmatrix} 1 & 0 \\ 7 & 0 \end{bmatrix}$, and $\mathbf{I}_2 - \mathbf{GA} = \begin{bmatrix} 0 & 0 \\ -7 & 1 \end{bmatrix}$, so

$$\tilde{x} = (2, 14) + (0, -7q_1 + q_2), \quad q_1, q_2 \in \mathbb{R}.$$

Since $-7q_1 + q_2$ can be any number, the solutions are exactly the vectors $(2, t)$, $t \in \mathbb{R}$, which is what

$$\begin{aligned}
 \mathbf{A}(x_1, x_2) &= (x_1, 0) \\
 &= (2, 0)
 \end{aligned}$$

says directly.

Corollary 0.1 (Solvable for every right-hand side, and solvable uniquely). *Let $\mathbf{A}$ be $m \times n$.*

1. $\mathbf{A}\tilde{x} = \tilde{b}$ is consistent for every $\tilde{b} \in \mathbb{R}^m$ exactly when $\text{rank}(\mathbf{A}) = m$.
2. A consistent system $\mathbf{A}\tilde{x} = \tilde{b}$ has exactly one solution exactly when $\text{rank}(\mathbf{A}) = n$.

Proof

Proof. Part 1. By Theorem 0.14, the system is consistent for every $\tilde{b}$ exactly when $\mathcal{C}(\mathbf{A}) = \mathbb{R}^m$. $\mathcal{C}(\mathbf{A})$ is a subspace of $\mathbb{R}^m$ (Theorem 1 in Column Space, Null Space and Rank-Nullity) with dimension $\text{rank}(\mathbf{A})$ (Theorem 6 in Column Space, Null Space and Rank-Nullity), and $\dim(\mathbb{R}^m) = m$ (Example 16 in Subspaces and Rank). If the dimension is m , then $\mathcal{C}(\mathbf{A}) = \mathbb{R}^m$ (Theorem 11 in Subspaces and Rank); if $\mathcal{C}(\mathbf{A}) = \mathbb{R}^m$, the dimension is m .

Part 2. By Theorem 0.15, the solutions are a particular solution plus the vectors of $\mathcal{N}(\mathbf{A})$, so there is exactly one solution exactly when $\mathcal{N}(\mathbf{A}) = \{\tilde{0}_n\}$, that is, when $\text{nullity}(\mathbf{A}) = 0$ (a subspace of dimension 0 has the empty list as a basis, whose span is $\{\tilde{0}\}$, by Definition 9 in Subspaces and Rank and Definition 5 in Subspaces and Rank; and $\{\tilde{0}\}$ has dimension 0, by Example 16 in Subspaces and Rank). By Theorem 7 in Column Space, Null Space and Rank-Nullity, that is when $\text{rank}(\mathbf{A}) = n$. $\square$

Example 0.22 (Three systems).

- The 2×3 matrix $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ of Example 15 in Column Space, Null Space and Rank-Nullity has rank $2 = m$, so every system with it is consistent, but $2 < 3 = n$, so the solutions are never unique.
- The 3×2 matrix $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ of Example 0.1 has rank $2 = n$, so a consistent system has one solution, but $2 < 3 = m$, so some systems are inconsistent: $(1, 0, 0) \notin \mathcal{C}(\mathbf{X})$, because $a(1, 1, 1) + b(1, 2, 3)$ has equal differences between consecutive entries, and $(1, 0, 0)$ does not.
- An invertible $p \times p$ matrix $\mathbf{A}$ has rank p : if $\mathbf{A}\tilde{x} = \tilde{0}$ then

$$\begin{aligned} \tilde{x} &= \mathbf{A}^{-1}\mathbf{A}\tilde{x} \\ &= \tilde{0}, \end{aligned}$$

so its nullity is 0 and its rank is p (Theorem 7 in Column Space, Null Space and Rank-Nullity). So every system with it is consistent (part 1, with $p = m$) and has exactly one solution (part 2, with $p = n$), namely $\mathbf{A}^{-1}\tilde{b}$, since $\mathbf{A}(\mathbf{A}^{-1}\tilde{b}) = \tilde{b}$.

Theorem 0.16 (The projection onto a column space, with any generalized inverse). *Let $\mathbf{X}$ be an $n \times p$ matrix of any rank, and let $\mathbf{G}$ be any generalized inverse of $\mathbf{X}^\top \mathbf{X}$ (Definition 0.4). Then for every $y \in \mathbb{R}^n$, $\mathbf{X}\mathbf{G}\mathbf{X}^\top y$ is the orthogonal projection of y onto $\mathcal{C}(\mathbf{X})$ (Definition 0.3). So the matrix $\mathbf{X}\mathbf{G}\mathbf{X}^\top$ is the same for every choice of $\mathbf{G}$, and when $\text{rank}(\mathbf{X}) = p$ it is the hat matrix $\mathbf{H}$ (Definition 0.2).*

i Proof

Proof. $\mathbf{X}^\top y$ is in $\mathcal{C}(\mathbf{X}^\top \mathbf{X})$. $\mathcal{C}(\mathbf{X}^\top \mathbf{X}) \subseteq \mathcal{C}(\mathbf{X}^\top)$, since $(\mathbf{X}^\top \mathbf{X})\tilde{z} = \mathbf{X}^\top(\mathbf{X}\tilde{z})$ (Theorem 5 in Matrices). Both are subspaces of $\mathbb{R}^p$ (Theorem 1 in Column Space, Null Space and Rank-Nullity) with the same dimension: $\text{rank}(\mathbf{X}^\top \mathbf{X}) = \text{rank}(\mathbf{X}^\top)$ (Theorem 9 in Column Space, Null Space and Rank-Nullity, Theorem 6 in Column Space, Null Space and Rank-Nullity). So $\mathcal{C}(\mathbf{X}^\top \mathbf{X}) = \mathcal{C}(\mathbf{X}^\top)$ (Theorem 11 in Subspaces and Rank), and since $\mathbf{X}^\top y \in \mathcal{C}(\mathbf{X}^\top)$ (Definition 1 in Column Space, Null Space and Rank-Nullity), $\mathbf{X}^\top y = \mathbf{X}^\top \mathbf{X}\tilde{z}$ for some $\tilde{z} \in \mathbb{R}^p$.

The split. Let

$$\begin{aligned}\tilde{u} &= \mathbf{X}\mathbf{G}\mathbf{X}^\top y \\ &= \mathbf{X}(\mathbf{G}\mathbf{X}^\top \tilde{y})\end{aligned}$$

(Theorem 5 in Matrices), which is in $\mathcal{C}(\mathbf{X})$ (Definition 1 in Column Space, Null Space and Rank-Nullity). For $y - \tilde{u}$,

$$\begin{aligned}\mathbf{X}^\top(y - \tilde{u}) &= \mathbf{X}^\top y - \mathbf{X}^\top \tilde{u} && \text{(Theorem 4 in Subspaces and Rank, with coefficients 1 and } -1\text{)} \\ &= \mathbf{X}^\top \tilde{y} - \mathbf{X}^\top (\mathbf{X}\mathbf{G}\mathbf{X}^\top y) && \text{(substitute } \tilde{u}\text{)} \\ &= \mathbf{X}^\top \tilde{y} - (\mathbf{X}^\top \mathbf{X})\mathbf{G}(\mathbf{X}^\top y) && \text{(Theorem 5 in Matrices)} \\ &= \mathbf{X}^\top \tilde{y} - (\mathbf{X}^\top \mathbf{X})\mathbf{G}(\mathbf{X}^\top \tilde{X})\tilde{z} && \text{(first step)} \\ &= \mathbf{X}^\top \tilde{y} - ((\mathbf{X}^\top \mathbf{X})\mathbf{G}(\mathbf{X}^\top \mathbf{X}))\tilde{z} && \text{(Theorem 5 in Matrices)} \\ &= \mathbf{X}^\top \tilde{y} - (\mathbf{X}^\top \mathbf{X})\tilde{z} && \text{(Definition 0.4)} \\ &= \mathbf{X}^\top \tilde{y} - \mathbf{X}^\top y && \text{(first step)} \\ &= \mathbf{0}_p, && \text{(arithmetic)}\end{aligned}$$

so $y - \tilde{u} \in \mathcal{N}(\mathbf{X}^\top) = \mathcal{C}(\mathbf{X})^\perp$ (Theorem 7 in Direct Sums and Orthogonal Complements). By the uniqueness in Theorem 8 in Direct Sums and Orthogonal Complements, $\tilde{u}$ is the orthogonal projection of y onto $\mathcal{C}(\mathbf{X})$.

Same matrix. Every choice of $\mathbf{G}$ gives a matrix that projects every y onto $\mathcal{C}(\mathbf{X})$, so all choices give the same matrix (Theorem 0.5, part 2). When $\text{rank}(\mathbf{X}) = \tilde{p}$, $(\mathbf{X}^\top \mathbf{X})^{-1}$ is one choice (Theorem 0.1, Theorem 0.8), and it gives $\mathbf{H}$. $\square$

Example 0.23 (A design matrix with a repeated column). Let $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, an intercept column entered twice, so $\text{rank}(\mathbf{X}) = 1 < 2$ and $\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ is not invertible: its rank is $\text{rank}(\mathbf{X}) = 1 < 2$ (Theorem 9 in Column Space, Null Space and Rank-Nullity), while an invertible 2×2 matrix has rank 2 (Example 0.22). The matrix $\mathbf{G} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}$ is a generalized inverse of it:

$$\begin{aligned} (\mathbf{X}^\top \mathbf{X}) \mathbf{G} (\mathbf{X}^\top \mathbf{X}) &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad (\text{Definition 7 in Matrices, for the first product}) \\ &= \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}. \quad (\text{Definition 7 in Matrices}) \end{aligned}$$

Then

$$\begin{aligned} \mathbf{X} \mathbf{G} \mathbf{X}^\top &= \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (\text{Definition 7 in Matrices, for } \mathbf{X} \mathbf{G}) \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad (\text{Definition 7 in Matrices}) \end{aligned}$$

the hat matrix of the intercept-only model (Example 0.2). Another generalized inverse, $\mathbf{G}' = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$, which is one because $(\mathbf{X}^\top \mathbf{X}) \mathbf{G}' = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$ and that times $\mathbf{X}^\top \mathbf{X}$ is again $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$, gives $\mathbf{X} \mathbf{G}' = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$ and the same $\mathbf{X} \mathbf{G}' \mathbf{X}^\top = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, as Theorem 0.16 says it must. Both project onto $\mathcal{C}(\mathbf{X}) = \text{span}\{(1, 1)\}$, the column space of the intercept-only design.

Solving linear systems and least squares

Source

This section is adapted from Zhou [5] and Zhou [9], used under the MIT License (see the license text in [Section 2 in Subspaces and Rank](#)). The first source states that every invertible matrix has an LU factorization; that is false without reordering rows (the 2×2 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 0.3](#)) rather than by differentiating, and the QR factorization is derived from Gram-Schmidt ([Theorem 10 in Inner Products and Orthogonality](#)). 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 0.32](#) only touches

Definition 0.7 (Upper triangular matrix). A square matrix $\mathbf{U}$ is **upper triangular** [10] if every entry below the diagonal is 0: $u_{ij} = 0$ whenever $i > j$.

Definition 0.8 (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 0.9 (Unit lower triangular matrix). A lower triangular matrix ([Definition 0.8](#)) whose diagonal entries are all 1 is **unit lower triangular**.

Example 0.24 (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 and Decompositions](#)) 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 0.17 (Solving a triangular system by substitution). *Let $\mathbf{U}$ be an $n \times n$ upper triangular matrix (Definition 0.7) whose diagonal entries $u_{11}, \dots, u_{nn}$ are all nonzero. For every $\tilde{b} \in \mathbb{R}^n$, the system $\mathbf{U}\tilde{x} = \tilde{b}$ 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, $\mathbf{L}\tilde{y} = \tilde{b}$ 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).

i Proof

Proof. Row i of $\mathbf{U}\tilde{x} = \tilde{b}$ reads

$$\begin{aligned} b_i &= \sum_{j=1}^n u_{ij} x_j && \text{(Definition 13 in Matrices)} \\ &= \sum_{j=i}^n u_{ij} x_j && (u_{ij} = 0 \text{ for } j < i) \\ &= u_{ii} x_i + \sum_{j=i+1}^n u_{ij} x_j. && \text{(split off the } j = i \text{ term)} \end{aligned}$$

This equation is equivalent to each of

$$\begin{aligned} u_{ii} x_i &= b_i - \sum_{j=i+1}^n u_{ij} x_j && \text{(subtract the sum from both sides)} \\ x_i &= \frac{1}{u_{ii}} \left(b_i - \sum_{j=i+1}^n u_{ij} x_j \right). && \text{(divide by } u_{ii} \neq 0) \end{aligned}$$

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 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. $\square$

Example 0.25 (Back substitution on a 3×3 system). Solve $\begin{bmatrix} 2 & 1 & -1 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 \end{bmatrix} \tilde{x} = \begin{bmatrix} 8 \\ 1 \\ 1 \end{bmatrix}$:

1.

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

2.

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

3.

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

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

Theorem 0.18 (QR factorization). *Let $\mathbf{A}$ be an $m \times n$ matrix with $\text{rank}(\mathbf{A}) = n$. 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 23 in Vectors](#)) and $\mathbf{R}$ is upper triangular ([Definition 0.7](#)) with positive diagonal entries.

i Proof

Proof. The n columns $\tilde{a}_1, \dots, \tilde{a}_n$ of $\mathbf{A}$ are linearly independent ([Definition 3 in Subspaces and Rank](#)), so the Gram-Schmidt process ([Definition 11 in Inner Products and Orthogonality](#)) runs all n steps and gives orthonormal $\tilde{q}_1, \dots, \tilde{q}_n$ ([Theorem 10 in Inner Products and Orthogonality](#), parts 1 and 3); let $\mathbf{Q}$ have these columns. At step i , $\tilde{q}'_i \neq 0$, and rearranging the orthogonalize step,

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

Let $\mathbf{R}$ be the $n \times n$ matrix with $r_{ji} = \tilde{q}_j \cdot \tilde{a}_i$ for $j < i$, $r_{ii} = \|\tilde{q}'_i\| > 0$, 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 7 in Matrices](#)), and

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

So $\mathbf{QR} = \mathbf{A}$. □

Example 0.26 (A QR factorization from Gram-Schmidt). For $\mathbf{A}$ with columns $\tilde{a}_1 = (1, 1, 0)$, $\tilde{a}_2 = (1, 0, 1)$, $\tilde{a}_3 = (0, 1, 1)$, [Example 22 in Inner Products and Orthogonality](#) and [Example 24 in Inner Products and Orthogonality](#) found $\tilde{q}_1 = \frac{1}{\sqrt{2}}(1, 1, 0)$, $\tilde{q}_2 = \frac{1}{\sqrt{6}}(1, -1, 2)$, $\tilde{q}_3 = \frac{1}{\sqrt{3}}(-1, 1, 1)$, with $\|\tilde{q}'_1\| = \sqrt{2}$, $\|\tilde{q}'_2\| = \sqrt{3/2}$, $\|\tilde{q}'_3\| = 2/\sqrt{3}$,

$$\begin{aligned}
\tilde{q}_1 \cdot \tilde{a}_2 &= \tilde{q}_1 \cdot \tilde{a}_3 \\
&= \frac{1}{\sqrt{2}}
\end{aligned}$$

and $\tilde{q}_2 \cdot \tilde{a}_3 = \frac{1}{\sqrt{6}}$. 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,

$$\begin{aligned}
\frac{1}{\sqrt{2}} q_1 + \sqrt{3/2} q_2 &= \frac{1}{2} (1, 1, 0) + \frac{1}{2} (1, -1, 2) \\
&= (1, 0, 1) \\
&= a_2.
\end{aligned}$$

Definition 0.10 (Least squares solution). Let $\mathbf{A}$ be $m \times n$ and $\tilde{b} \in \mathbb{R}^m$. A **least squares solution** [11] of $\mathbf{A}\tilde{x} = \tilde{b}$ is a vector $\hat{\tilde{x}} \in \mathbb{R}^n$ such that

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

Example 0.27 (Fitting a constant). Let $\mathbf{A} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\tilde{b} = (1, 3)$. The system $x(1, 1) = (1, 3)$ is inconsistent (Definition 0.6): its first entry needs $x = 1$ and its second needs $x = 3$. For any number x ,

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

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 0.11 (Normal equations). For $\mathbf{A}$ and $\tilde{b}$ as in Definition 0.10, the equations $\mathbf{A}^\top \mathbf{A}\tilde{x} = \mathbf{A}^\top \tilde{b}$ are the **normal equations** [12].

Theorem 0.19 (Least squares solutions solve the normal equations). Let $\mathbf{A}$ be $m \times n$ and $\tilde{b} \in \mathbb{R}^m$.

1. $\hat{\tilde{x}}$ is a least squares solution (Definition 0.10) exactly when it solves the normal equations $\mathbf{A}^\top \mathbf{A}\hat{\tilde{x}} = \mathbf{A}^\top \tilde{b}$.
2. The normal equations are consistent (Definition 0.6).
3. Every least squares solution gives the same fitted vector $\mathbf{A}\hat{\tilde{x}}$: the orthogonal projection of $\tilde{b}$ onto $\mathcal{C}(\mathbf{A})$ (Definition 0.3).
4. The least squares solution is unique exactly when $\text{rank}(\mathbf{A}) = n$, and then $\hat{\tilde{x}} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \tilde{b}$.

i Proof

Proof. Let $\tilde{u}$ be the orthogonal projection of $\tilde{b}$ onto $\mathcal{C}(\mathbf{A})$. The vectors $\mathbf{A}\tilde{x}$, as $\tilde{x}$ ranges over $\mathbb{R}^n$, are exactly the points of $\mathcal{C}(\mathbf{A})$ (Definition 1 in Column Space, Null Space and Rank-Nullity). **Least squares solutions are the solutions of $\mathbf{A}\tilde{x} = \tilde{u}$.** If $\mathbf{A}\tilde{x} = \tilde{u}$, then for every $\tilde{x}$, $\mathbf{A}\tilde{x} \in \mathcal{C}(\mathbf{A})$, so $\|\tilde{b} - \mathbf{A}\tilde{x}\| = \|\tilde{b} - \tilde{u}\| \leq \|\tilde{b} - \mathbf{A}\tilde{x}\|$ (Theorem 0.3), and $\tilde{x}$ is a least squares solution. Conversely, let $\tilde{x}$ be a least squares solution. $\tilde{u} \in \mathcal{C}(\mathbf{A})$, so $\tilde{u} = \mathbf{A}\tilde{x}_0$ for some $\tilde{x}_0$ (Definition 1 in Column Space, Null Space and Rank-Nullity), and

$$\begin{aligned} \|\tilde{b} - \mathbf{A}\tilde{x}\| &\leq \|\tilde{b} - \mathbf{A}\tilde{x}_0\| \quad (\text{Definition 0.10}) \\ &= \|\tilde{b} - \tilde{u}\| \quad (\tilde{u} = \mathbf{A}\tilde{x}_0) \\ &\leq \|\tilde{b} - \mathbf{A}\tilde{x}\|, \quad (\text{Theorem 0.3, with } \tilde{w} = \mathbf{A}\tilde{x} \in \mathcal{C}(\mathbf{A})) \end{aligned}$$

so the last inequality is an equality, and the equality case of Theorem 0.3 gives $\mathbf{A}\tilde{x} = \tilde{u}$. **Parts 1 and 3.** $\mathbf{A}\tilde{x} = \tilde{u}$ holds exactly when $\tilde{b} - \mathbf{A}\tilde{x} \in \mathcal{C}(\mathbf{A})^\perp$: if $\mathbf{A}\tilde{x} = \tilde{u}$, then $\tilde{b} - \tilde{u} \in \mathcal{C}(\mathbf{A})^\perp$ (Definition 0.3); conversely, if $\tilde{b} - \mathbf{A}\tilde{x} \in \mathcal{C}(\mathbf{A})^\perp$, then $\tilde{b} = \mathbf{A}\tilde{x} + (\tilde{b} - \mathbf{A}\tilde{x})$ is a split into $\mathcal{C}(\mathbf{A})$ and $\mathcal{C}(\mathbf{A})^\perp$, so $\mathbf{A}\tilde{x} = \tilde{u}$ by the uniqueness in Theorem 8 in Direct Sums and Orthogonal Complements. And $\mathcal{C}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^\top)$ (Theorem 7 in Direct Sums and Orthogonal Complements), where

$$\begin{aligned} \mathbf{A}^\top (\tilde{b} - \mathbf{A}\tilde{x}) &= \mathbf{A}^\top \tilde{b} - \mathbf{A}^\top (\mathbf{A}\tilde{x}) \quad (\text{Theorem 0.7}) \\ &= \mathbf{A}^\top \tilde{b} - \mathbf{A}^\top \mathbf{A}\tilde{x}, \quad (\text{Theorem 5 in Matrices}) \end{aligned}$$

which is 0_n exactly when the normal equations hold.

Part 2. The vector $\tilde{x}_0$ above satisfies $\mathbf{A}\tilde{x}_0 = \tilde{u}$, 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 $\mathbf{A}\tilde{x} = \tilde{u}$, a consistent system, so there is exactly one when $\text{rank}(\mathbf{A}) = n$ and more than one otherwise (Corollary 0.1). When $\text{rank}(\mathbf{A}) = n$, $\mathbf{A}^\top \mathbf{A}$ is invertible (Theorem 0.1), and with $\mathbf{M} \stackrel{\text{def}}{=} \mathbf{A}^\top \mathbf{A}$,

$$\begin{aligned} \tilde{x} &= \mathbf{I}_n \tilde{x} \quad (\text{Theorem 8 in Matrices}) \\ &= (\mathbf{M}^{-1} \mathbf{M}) \tilde{x} \quad (\text{Definition 5 in Special Matrices and Decompositions}) \\ &= \mathbf{M}^{-1} (\mathbf{M}\tilde{x}) \quad (\text{Theorem 5 in Matrices}) \\ &= \mathbf{M}^{-1} \mathbf{A}^\top \tilde{b}. \quad (\text{the normal equations}) \end{aligned}$$

□

Example 0.28 (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 0.1) and $\tilde{b} = (1, 2, 2)$. $\mathbf{A}$ 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 0.1](#)), and

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

so

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

The fitted values are $\mathbf{A}\tilde{\mathbf{x}} = \left(\frac{7}{6}, \frac{5}{3}, \frac{13}{6}\right)$, and the residuals $\tilde{\mathbf{b}} - \mathbf{A}\tilde{\mathbf{x}} = \left(-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}\right)$ 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 0.29 (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 $\tilde{\mathbf{b}} = (1, 2, 2)$, $\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix}$ and

$\mathbf{A}^\top \tilde{\mathbf{b}} = (5, 10)$, 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 $(\frac{5}{3}, 0)$ and $(0, \frac{5}{6})$, as part 4 predicts for $\text{rank}(\mathbf{A}) = 1 < 2$; but every one gives the same fitted vector

$$\begin{aligned} \mathbf{A}\tilde{\mathbf{x}} &= (x_1 + 2x_2) (1, 1, 1) \\ &= \frac{5}{3} (1, 1, 1), \end{aligned}$$

as part 3 says.

Theorem 0.20 (Solving least squares by QR). *Let $\mathbf{A}$ be $m \times n$ with $\text{rank}(\mathbf{A}) = n$, with QR factorization $\mathbf{A} = \mathbf{Q}\mathbf{R}$ ([Theorem 0.18](#)), and let $\tilde{\mathbf{b}} \in \mathbb{R}^m$. The least squares solution ([Definition 0.10](#)) is the unique solution of*

$$\underbrace{\mathbf{R}}_{n \times n} \tilde{\mathbf{x}} = \underbrace{\mathbf{Q}^\top}_{n \times m} \tilde{\mathbf{b}},$$

which back substitution computes ([Theorem 0.17](#)). When $m = n$, it is the unique solution of $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$.

i Proof

Proof. $\mathbf{R}$ is upper triangular with positive diagonal, so $\mathbf{R}\tilde{x} = \mathbf{Q}^\top \tilde{b}$ has exactly one solution $\tilde{x}$ (Theorem 0.17). $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_n$, as in the proof of Theorem 0.4. Then

$$\begin{aligned}
 \mathbf{A}^\top \mathbf{A} \tilde{x} &= (\mathbf{QR})^\top (\mathbf{QR}) \tilde{x} && \text{(substitute } \mathbf{A} = \mathbf{QR}) \\
 &= \mathbf{R}^\top \mathbf{Q}^\top (\mathbf{QR}) \tilde{x} && \text{(Theorem 17 in Matrices)} \\
 &= \mathbf{R}^\top (\mathbf{Q}^\top \mathbf{Q}) (\mathbf{R} \tilde{x}) && \text{(Theorem 5 in Matrices)} \\
 &= \mathbf{R}^\top \mathbf{I}_n (\mathbf{R} \tilde{x}) && (\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_n) \\
 &= \mathbf{R}^\top (\mathbf{R} \tilde{x}) && \text{(Theorem 8 in Matrices)} \\
 &= \mathbf{R}^\top (\mathbf{Q}^\top \tilde{b}) && (\mathbf{R} \tilde{x} = \mathbf{Q}^\top \tilde{b}) \\
 &= (\mathbf{R}^\top \mathbf{Q}^\top) \tilde{b} && \text{(Theorem 5 in Matrices)} \\
 &= (\mathbf{QR})^\top \tilde{b} && \text{(Theorem 17 in Matrices)} \\
 &= \mathbf{A}^\top \tilde{b}, && (\mathbf{A} = \mathbf{QR})
 \end{aligned}$$

so $\tilde{x}$ solves the normal equations and is the least squares solution (Theorem 0.19), which is unique because $\text{rank}(\mathbf{A}) = n$. When $m = n$,

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

so $\mathbf{A}\tilde{x} = \tilde{b}$ is consistent for every $\tilde{b}$ (Corollary 0.1, part 1), so its solution leaves residual 0, is a least squares solution, and is therefore $\tilde{x}$. $\square$

Example 0.30 (The least squares line by QR). For $\mathbf{A}$ and $\tilde{b}$ of Example 0.28, Gram-Schmidt gives $q_1 = \frac{1}{\sqrt{3}}(1, 1, 1)$; then

$$\begin{aligned}
 q_1 \cdot (1, 2, 3) &= \frac{6}{\sqrt{3}} \\
 &= 2\sqrt{3},
 \end{aligned}$$

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

with norm $\sqrt{2}$, and $q_2 = \frac{1}{\sqrt{2}}(-1, 0, 1)$. So

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

Back substitution gives

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

and

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

the solution found in Example 0.28.

Definition 0.12 (LU factorization). An **LU factorization** [13] of an $n \times n$ matrix $\mathbf{A}$ is a product $\mathbf{A} = \mathbf{L}\mathbf{U}$ with $\mathbf{L}$ unit lower triangular (Definition 0.9) and $\mathbf{U}$ upper triangular (Definition 0.7).

Example 0.31 (An LU factorization of a 3×3 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 7 in Matrices, 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 0.32 (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 and Decompositions); but it has no LU factorization. If $\mathbf{P} = \mathbf{L}\mathbf{U}$, 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 0.21 (Solving a system with an LU factorization). *If $\mathbf{A} = \mathbf{LU}$ is an LU factorization (Definition 0.12) and the diagonal entries of $\mathbf{U}$ are all nonzero, then for every $\tilde{b} \in \mathbb{R}^n$ the system $\mathbf{A}\tilde{x} = \tilde{b}$ has exactly one solution: solve $\mathbf{L}\tilde{y} = \tilde{b}$ by forward substitution, then $\mathbf{U}\tilde{x} = \tilde{y}$ by back substitution (Theorem 0.17).*

i 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 0.17). If $\tilde{y}$ and $\tilde{x}$ are those solutions, then

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

(Theorem 5 in Matrices), so $\tilde{x}$ is a solution. Conversely, if $\mathbf{A}\tilde{x} = \tilde{b}$, then $\tilde{y} \stackrel{\text{def}}{=} \mathbf{U}\tilde{x}$ satisfies

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

(Theorem 5 in Matrices), so $\tilde{y}$ is the unique solution of that system, and $\tilde{x}$ is then the unique solution of $\mathbf{U}\tilde{x} = \tilde{y}$. $\square$

Example 0.33 (Two triangular solves). With $\mathbf{L}$ and $\mathbf{U}$ of Example 0.31, solve $\mathbf{A}\tilde{x} = (8, -11, -3)$. Forward substitution on $\mathbf{L}\tilde{y} = (8, -11, -3)$ gives $y_1 = 8$,

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

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

Back substitution on $\mathbf{U}\tilde{x} = (8, 1, 1)$ is Example 0.25, which gives $\tilde{x} = (2, 3, -1)$. 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] Wikipedia contributors, “Design matrix — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Design_matrix

- [2] S. Banerjee and A. Roy, *Linear algebra and matrix analysis for statistics*, vol. 181. Crc Press Boca Raton, 2014. Available: <https://www.routledge.com/Linear-Algebra-and-Matrix-Analysis-for-Statistics/Banerjee-Roy/p/book/9781420095388>
- [3] Wikipedia contributors, “Projection matrix — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Projection_matrix
- [4] H. Zhou, “Orthogonal projections.” 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/06-orthproj/06-orthproj.html>
- [5] 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>
- [6] Wikipedia contributors, “Generalized inverse — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Generalized_inverse
- [7] Wikipedia contributors, “Moore–penrose inverse — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse
- [8] Wikipedia contributors, “Consistent and inconsistent equations — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Consistent_and_inconsistent_equations
- [9] 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>
- [10] Wikipedia contributors, “Triangular matrix — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Triangular_matrix
- [11] Wikipedia contributors, “Least squares — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Least_squares
- [12] Wikipedia contributors, “Normal equations — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Normal_equations
- [13] Wikipedia contributors, “LU decomposition — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/LU_decomposition