San José State University

Math 250: Mathematical Data Visualization

Generalized inverse and pseudoinverse

Dr. Guangliang Chen

Outline

- Matrix generalized inverse
- Pseudoinverse
- Applications

Recall

... that a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is **invertible** if there exists a square matrix $\mathbf{B}$ of the same size such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

In this case, $\mathbf{B}$ is called the matrix inverse of $\mathbf{A}$ and denoted as $\mathbf{B} = \mathbf{A}^{-1}$.

Remark. A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible if and only if

- **$\mathbf{A}$ has full rank (nonsingular)**, i.e., $\operatorname{rank}(\mathbf{A}) = n$
- **$\mathbf{A}$ has a nonzero determinant**: $\det(\mathbf{A}) \neq 0$

Remark. For any invertible matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and any vector $\mathbf{b} \in \mathbb{R}^n$, the linear system $\mathbf{Ax} = \mathbf{b}$ has a unique solution

$$\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}.$$

MATLAB command for solving a linear system $\mathbf{Ax} = \mathbf{b}$

$A\backslash b$ % recommended

$inv(A) * b$ % avoid (especially when $\mathbf{A}$ is large)

What about general matrices?

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a square but singular matrix ($m = n, \det(\mathbf{A}) = 0$), or a rectangular matrix ($m \neq n$).

We would like to address the following questions:

- **Is there some kind of inverse?**
- **Given a vector $\mathbf{b} \in \mathbb{R}^m$, does the linear system $\mathbf{Ax} = \mathbf{b}$ have a solution?**
  - If yes, is the solution unique?
  - Otherwise, is there some kind of approximate solution?

More motivation: Least squares

In many practical tasks, the least squares problem arises naturally:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 \qquad (\text{where } \mathbf{A}\in\mathbb{R}^{m\times n}, \mathbf{b}\in\mathbb{R}^m \text{ are fixed})$$

Theorem. *If* $\mathbf{A}$ *has full column rank (i.e.,* $\operatorname{rank}(\mathbf{A}) = n \le m$*), then the above problem has a unique solution, called* <u>*least squares solution*</u>*:*

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

Proof. First, we rewrite the objective function as

$$f(\mathbf{x}) = \mathbf{x}^T(\mathbf{A}^T\mathbf{A})\mathbf{x} - 2\mathbf{x}^T(\mathbf{A}^T\mathbf{b}) + \|\mathbf{b}\|^2 \longleftarrow \text{Convex , differentiable}$$

Next, we find all critical points by setting the gradient to zero:

$$\nabla f = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b} = 0.$$

Since $\mathrm{rank}(\mathbf{A}^T\mathbf{A}) = \mathrm{rank}(\mathbf{A}) = n$, the matrix $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ is non-singular, and consequently the above equation has one and only one solution

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}.$$

To show that it is a local minimizer (and thus also a global minimizer), we compute the Hessian of f and obtain that

$$\nabla^2 f = 2\mathbf{A}^T\mathbf{A},$$

which is a positive definite matrix (because $\mathbf{A}^T\mathbf{A}$ is nonsingular).

We want to better understand the following two matrices:

- $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ (pseudoinverse): The least squares solution is

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

- $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ (projection matrix): Closest approximation of $\mathbf{b}$ is

$$\mathbf{A}\mathbf{x}^* = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

Generalized inverse

Def 0.1. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be any matrix. We call the matrix $\mathbf{G} \in \mathbb{R}^{n\times m}$ a **generalized inverse** of $\mathbf{A}$ if it satisfies

$$\mathbf{AGA} = \mathbf{A}$$

Remark. If $\mathbf{A}$ is square and invertible, then (1) $\mathbf{A}^{-1}$ is a generalized inverse of $\mathbf{A}$ and (2) it is the only generalized inverse of $\mathbf{A}$:

$$\mathbf{G} = \mathbf{A}^{-1}(\mathbf{AGA})\mathbf{A}^{-1} = \mathbf{A}^{-1}(\mathbf{A})\mathbf{A}^{-1} = \mathbf{A}^{-1}$$

This thus justifies the term “generalized inverse”.

Remark. For a general matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, its generalized inverse always exists but might not be unique.

For example, let $\mathbf{A} = [1, 2] \in \mathbb{R}^{1\times 2}$. Its generalized inverse is a matrix $\mathbf{G} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{2\times 1}$ satisfying

$$[1, 2] = \mathbf{A} = \mathbf{AGA} = [1, 2] \begin{bmatrix} x \\ y \end{bmatrix} [1, 2] = (x + 2y) \cdot [1, 2].$$

This shows that any $\mathbf{G} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{2\times 1}$ with $x + 2y = 1$ is a generalized inverse of $\mathbf{A}$, e.g., $\mathbf{G} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\mathbf{G} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

The following theorem indicates a way to find the generalized inverse of any matrix.

Theorem 0.1. Let $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \in \mathbb{R}^{m\times n}$ be a matrix of rank r, and $A_{11} \in \mathbb{R}^{r\times r}$. If A_{11} is invertible, then $\mathbf{G} = \begin{bmatrix} A_{11}^{-1} & O \\ O & O \end{bmatrix} \in \mathbb{R}^{n\times m}$ is a generalized inverse of $\mathbf{A}$.

Remark. Any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with rank r can be rearranged through row and column permutations to have the above partitioned form with an invertible $r \times r$ submatrix in the top-left corner. This theorem essentially establishes the existence of a generalized inverse for any matrix.

Example 0.1. Consider the following matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Since $\operatorname{rank}(\mathbf{A}) = 2$ and the top-left 2×2 block happens to be invertible, we can easily find a generalized inverse

$$\mathbf{G} = \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

You are asked to verify that $\mathbf{AGA} = \mathbf{A}$.

The generalized inverse can also be used to find a solution to a consistent linear system (i.e., a system with at least a solution).

Theorem 0.2. Consider the linear system $\mathbf{Ax} = \mathbf{b}$. Suppose $\mathbf{b} \in \mathrm{Col}(\mathbf{A})$ such that the system is consistent. Let $\mathbf{G}$ be a generalized inverse of $\mathbf{A}$, i.e., $\mathbf{AGA} = \mathbf{A}$. Then $\mathbf{x}^* = \mathbf{Gb}$ is a particular solution to the system.

Proof. Multiplying both sides of $\mathbf{Ax} = \mathbf{b}$ by $\mathbf{AG}$ gives that

$$(\mathbf{AG})\mathbf{b} = (\mathbf{AG})\mathbf{Ax} = (\mathbf{AGA})\mathbf{x} = \mathbf{Ax} = \mathbf{b}.$$

This shows that $\mathbf{x}^* = \mathbf{Gb}$ is a particular solution to the linear system. □

Example 0.2. Consider the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 15 \\ 24 \end{bmatrix}.$$

It is consistent because $\mathbf{x} = \mathbf{1}$ is a solution.

According to the theorem, a particular solution to the system is

$$\mathbf{x}^* = \mathbf{G}\mathbf{b} = \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 15 \\ 24 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$$

Projection matrices

Def 0.2. A square matrix $\mathbf{P}$ is called a **projection matrix** if it is idempotent, i.e., $\mathbf{P} = \mathbf{P}^2$.

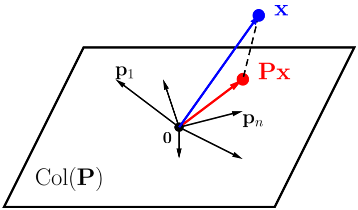
Remark. Let $\mathbf{P}$ be a projection matrix. Then

- $\mathbf{P}$ must be digonalizable;
- $\mathbf{P}$ have eigenvalues of 0 and/or 1. Moreover, the algebraic multiplicity of 1 is equal to the rank and trace of $\mathbf{P}$.

Remark. The following statements explain what a projection matrix does:

- A projection matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ projects any vector in $\mathbb{R}^n$ onto its column space. To see this, let $\mathbf{x} \in \mathbb{R}^n$. Then

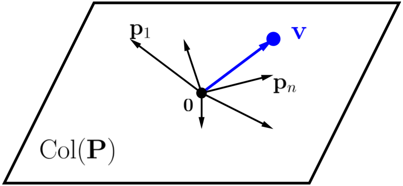
$$\mathbf{P}\mathbf{x} = [\mathbf{p}_1 \dots \mathbf{p}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum x_i \mathbf{p}_i \in \mathrm{Col}(\mathbf{P}).$$



- A projection matrix projects every vector already in its column space onto itself.

  To see this, let $\mathbf{v} \in \mathrm{Col}(\mathbf{P})$. Then there exists some $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{v} = \mathbf{P}\mathbf{x}$.

  It follows that $\mathbf{P}\mathbf{v} = \mathbf{P}(\mathbf{P}\mathbf{x}) = \mathbf{P}^2\mathbf{x} = \mathbf{P}\mathbf{x} = \mathbf{v}$.



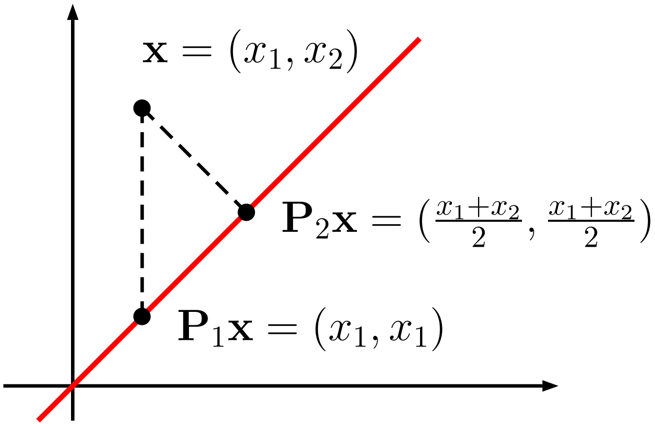
Example 0.3. Below are two projection matrices:

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{P}_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

They have the same column space, which is the 45 degree line through the origin in $\mathbb{R}^2$, and thus both project points in $\mathbb{R}^2$ onto the same line. However, the ways they project points are different: For any $\mathbf{x} \in \mathbb{R}^2$,

$$\mathbf{P}_1\mathbf{x} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1 \end{pmatrix}$$

$$\mathbf{P}_2\mathbf{x} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{x_1+x_2}{2} \\ \frac{x_1+x_2}{2} \end{pmatrix}$$


$\mathbf{x} = (x_1, x_2)$
$\mathbf{P}_2\mathbf{x} = (\frac{x_1+x_2}{2}, \frac{x_1+x_2}{2})$
$\mathbf{P}_1\mathbf{x} = (x_1, x_1)$

Theorem 0.3. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ with a generalized inverse $\mathbf{G} \in \mathbb{R}^{n\times m}$. Then $\mathbf{P} = \mathbf{AG} \in \mathbb{R}^{m\times m}$ is a projection matrix.

Proof. From $\mathbf{AGA} = \mathbf{A}$, we obtain

$$(\mathbf{AG})(\mathbf{AG}) = (\mathbf{AGA})\mathbf{G} = \mathbf{AG}.$$

This shows that $\mathbf{AG}$ is a projection matrix. □

Remark. Similarly, $\mathbf{GA} \in \mathbb{R}^{n\times n}$ is also a projection matrix

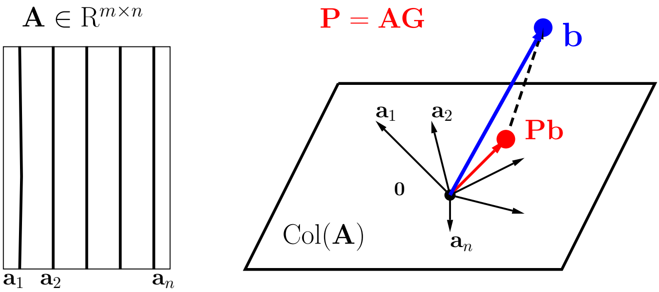
$$(\mathbf{GA})(\mathbf{GA}) = \mathbf{G}(\mathbf{AGA}) = \mathbf{GA}$$

Remark. $\mathbf{AG}$ and $\mathbf{A}$ must have the same column space. To see this,

(1) For any $\mathbf{y} \in \mathrm{Col}(\mathbf{AG})$, there exists some $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{y} = (\mathbf{AG})\mathbf{x}$. It follows that $\mathbf{y} = \mathbf{A}(\mathbf{Gx}) \in \mathrm{Col}(\mathbf{A})$. This shows that $\mathrm{Col}(\mathbf{AG}) \subseteq \mathrm{Col}(\mathbf{A})$.

(2) For any $\mathbf{y} \in \mathrm{Col}(\mathbf{A})$, there exists some $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathbf{Ax}$. Write $\mathbf{y} = (\mathbf{AGA})\mathbf{x} = (\mathbf{AG})(\mathbf{Ax})$. This shows that $\mathbf{y} \in \mathrm{Col}(\mathbf{AG})$. Thus, $\mathrm{Col}(\mathbf{A}) \subseteq \mathrm{Col}(\mathbf{AG})$.

Therefore, $\mathbf{AG}$ **is a projection matrix onto the column space of** $\mathbf{A}$.



Similarly, $\mathbf{GA}$ is a projection matrix onto the row space of $\mathbf{A}$.

Example 0.4. Consider the matrix $\mathbf{A}$ and its generalized inverse $\mathbf{G}$:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We have

$$\mathbf{AG} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix},$$

which represents a projection matrix onto the column space of $\mathbf{A}$.

Pseudoinverse

Briefly speaking, the matrix pseudoinverse is a generalized inverse with more constraints.

Def 0.3. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. We call the matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ the **pseudoinverse** of $\mathbf{A}$ if it satisfies all four conditions below:

(1) $\mathbf{ABA} = \mathbf{A}$ $\longleftarrow$ $\mathbf{B}$ is a generalized inverse of $\mathbf{A}$

(2) $\mathbf{BAB} = \mathbf{B}$ $\longleftarrow$ $\mathbf{A}$ is a generalized inverse of $\mathbf{B}$

(3) $(\mathbf{AB})^T = \mathbf{AB}$ $\longleftarrow$ $\mathbf{AB}$ is symmetric

(4) $(\mathbf{BA})^T = \mathbf{BA}$ $\longleftarrow$ $\mathbf{BA}$ is symmetric

Remark.

- If $\mathbf{B}$ only satisfies (1), it is known as a generalized inverse of $\mathbf{A}$; if $\mathbf{B}$ only satisfies (1) and (2), it is called a **reflexive generalized inverse**.
- For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the pseudoinverse always exists and is unique. We denote the pseudoinverse of $\mathbf{A}$ as $\mathbf{A}^\dagger$.
- A pseudoinverse is sometimes called the **Moore–Penrose inverse**, after the pioneering works by E. H. Moore and Roger Penrose.
- The symmetric form of the definition implies $\mathbf{B} = \mathbf{A}^\dagger$ and $\mathbf{A} = \mathbf{B}^\dagger$, and thus, $\mathbf{A} = (\mathbf{A}^\dagger)^\dagger$.

Example 0.5. Consider $\mathbf{A} = [1, 2] \in \mathbb{R}^{1\times 2}$ again. We showed that any matrix $\mathbf{G} = (x, y)^T \in \mathbb{R}^{2\times 1}$ with $x + 2y = 1$ is a generalized inverse of $\mathbf{A}$:

$$[1,2] = \mathbf{A} = \mathbf{AGA} = [1,2]\begin{bmatrix} x \\ y \end{bmatrix}[1,2] = (x+2y)\cdot[1,2].$$

To find its pseudoinverse, we need to write down three more equations:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{G} = \mathbf{GAG} = \begin{bmatrix} x \\ y \end{bmatrix}[1,2]\begin{bmatrix} x \\ y \end{bmatrix} = (x+2y)\cdot\begin{bmatrix} x \\ y \end{bmatrix}$$

$$x + 2y = (\mathbf{AG})^T = \mathbf{AG} = [1,2]\begin{bmatrix} x \\ y \end{bmatrix} = x + 2y$$

$$\begin{bmatrix} x & y \\ 2x & 2y \end{bmatrix} = (\mathbf{GA})^T = \mathbf{GA} = \begin{bmatrix} x \\ y \end{bmatrix}[1,2] = \begin{bmatrix} x & 2x \\ y & 2y \end{bmatrix} \longrightarrow 2x = y$$

Solving the two equations together gives that

$$x = \frac{1}{5},\ y = \frac{2}{5}.$$

Thus, the pseudoinverse of $\mathbf{A}$ is

$$\mathbf{A}^{\dagger} = \begin{bmatrix} \frac{1}{5} \\ \frac{2}{5} \end{bmatrix}.$$

Example 0.6. Let $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. Verify that $\mathbf{A}^\dagger = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix}$.

By direct calculation,

$$\begin{aligned}
\mathbf{A}\mathbf{A}^\dagger &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} && \text{(symmetric)} \\
\mathbf{A}^\dagger\mathbf{A} &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} && \text{(symmetric)} \\
\mathbf{A}\mathbf{A}^\dagger\mathbf{A} &= \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} && = \mathbf{A} \\
\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} && = \mathbf{A}^\dagger
\end{aligned}$$

Example 0.7 (Cont'd). Consider the matrix again

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

which has the following generalized inverse (i.e., $\mathbf{AGA} = \mathbf{A}$):

$$\mathbf{G} = \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

It can be verified that $\mathbf{A}$ is also a generalized inverse of $\mathbf{G}$:

$$\mathbf{GAG} = \mathbf{G}$$

Thus, $\mathbf{G}$ is (at least) a reflexive generalized inverse of $\mathbf{A}$.

However, neither $\mathbf{AG}$ nor $\mathbf{GA}$ is symmetric:

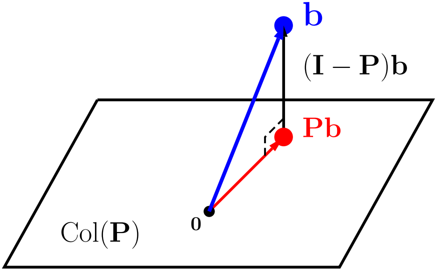
$$\mathbf{AG} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix}$$

$$\mathbf{GA} = \begin{bmatrix} -\frac{5}{3} & \frac{2}{3} & 0 \\ \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, $\mathbf{G}$ is not the pseudoinverse of $\mathbf{A}$.

Orthogonal projection matrices

Def 0.4. A square matrix $\mathbf{P}$ is called a **orthogonal projection matrix** if it is both symmetric and idempotent, i.e., $\mathbf{P} = \mathbf{P}^T$ and $\mathbf{P} = \mathbf{P}^2$.



Let $\mathbf{P} \in \mathbb{R}^{n\times n}$ be any orthogonal projection matrix. Because it is still a projection matrix, it must project any vector $\mathbf{b} \in \mathbb{R}^n$ onto its column space, i.e., $\mathbf{Pb} \in \mathrm{Col}(\mathbf{P})$.

This leads to the following decomposition of $\mathbf{b}$:

$$\mathbf{b} = \mathbf{Pb} + (\mathbf{I} - \mathbf{P})\mathbf{b}.$$

Since $\mathbf{P} = \mathbf{P}^T$ by definition, we have

$$(\mathbf{Pb})^T(\mathbf{I} - \mathbf{P})\mathbf{b} = \mathbf{b}^T\mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{b} = \mathbf{b}^T(\mathbf{P} - \mathbf{P}^2)\mathbf{b} = 0.$$

This shows that the two components, $\mathbf{Pb}$ and $(\mathbf{I} - \mathbf{P})\mathbf{b}$, are orthogonal to each other.

Example 0.8. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ are orthogonal projection matrices, but $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 2 & 0 \end{bmatrix}$ is not (it is just a projection matrix).

Example 0.9. The centering matrix $\mathbf{C}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n$ is also an orthogonal projection matrix (see notes for details).

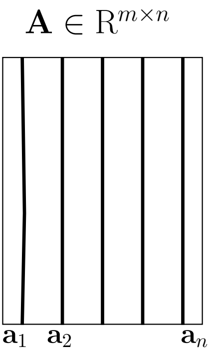
Theorem 0.4. For any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{A}\mathbf{A}^\dagger$ is an orthogonal projection matrix (onto the column space of $\mathbf{A}$).

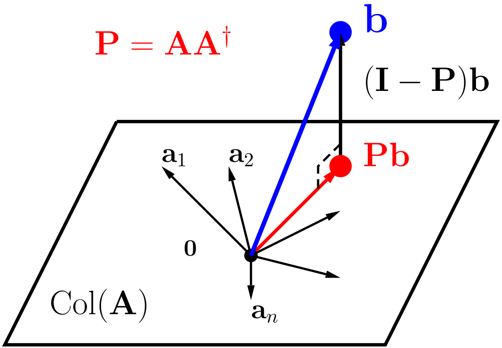
Proof. First, $\mathbf{A}^\dagger$ is still a generalized inverse. Thus, $\mathbf{A}\mathbf{A}^\dagger$ is a projection matrix (onto the column space of $\mathbf{A}$).

Secondly, since $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$, $\mathbf{A}\mathbf{A}^\dagger$ must be symmetric.

Therefore, by definition, $\mathbf{A}\mathbf{A}^\dagger$ is an orthogonal projection matrix. □

Remark. Similarly, $\mathbf{A}^\dagger\mathbf{A}$ is also an orthogonal projection matrix (onto the row space of $\mathbf{A}$).


$\mathbf{A} \in \mathrm{R}^{m\times n}$
$\mathbf{a}_1$
$\mathbf{a}_2$
$\mathbf{a}_n$


$\mathbf{P} = \mathbf{A}\mathbf{A}^\dagger$
$\mathbf{b}$
$(\mathbf{I} - \mathbf{P})\mathbf{b}$
$\mathbf{Pb}$
$\mathbf{a}_1$
$\mathbf{a}_2$
$\mathbf{0}$
$\mathrm{Col}(\mathbf{A})$
$\mathbf{a}_n$

Finding matrix pseudoinverse

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Our goal is to find $\mathbf{A}^\dagger$ (which exists and is unique).

We first consider the following two special settings:

- **$\mathbf{A}$ is a tall matrix with full column rank** (i.e., $\mathrm{rank}(\mathbf{A}) = n \le m$). Note that in this case, $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ is invertible.
- **$\mathbf{A}$ is a "diagonal" matrix** (i.e., $a_{ij} = 0$ whenever $i \neq j$).

Afterwards, we present how to find the pseudoinverse of a general matrix via its SVD.

Theorem 0.5. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be any tall matrix with full column rank (i.e., $\mathrm{rank}(\mathbf{A}) = n \leq m$). Then the pseudoinverse of $\mathbf{A}$ is

$$\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T.$$

Proof. It suffices to verify the four conditions for being a pseudoinverse:

$$\begin{aligned}
\mathbf{A}\mathbf{A}^\dagger\mathbf{A} &= \mathbf{A}\cdot(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\cdot\mathbf{A} = \mathbf{A} \\
\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\cdot\mathbf{A}\cdot(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{A}^\dagger \\
\mathbf{A}\mathbf{A}^\dagger &= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T && \text{(symmetric)} \\
\mathbf{A}^\dagger\mathbf{A} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\cdot\mathbf{A} = \mathbf{I}_n && \text{(symmetric)}
\end{aligned}$$

Therefore, $\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudoinverse of $\mathbf{A}$. □

Remark. The theorem implies that for any tall matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with full column rank (i.e., $\operatorname{rank}(\mathbf{A}) = n \le m$), the following is an **orthogonal projection** matrix (onto the column space of $\mathbf{A}$):

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T.$$

Example 0.10. Find the pseudoinverse of $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Solution: Observe that this matrix has full column rank (i.e., 2).

Since

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

we have

$$\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \end{pmatrix}$$

It follows that the orthogonal projection matrix onto $\mathrm{Col}(\mathbf{A})$ is

$$\mathbf{A}\mathbf{A}^{\dagger} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \frac{1}{3} \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

For instance, the orthogonal projection of $\mathbf{1}$ onto $\mathrm{Col}(\mathbf{A})$ is

$$\mathbf{A}\mathbf{A}^{\dagger}\mathbf{1} = \frac{1}{3} \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix} = \mathbf{A} \cdot \underbrace{\frac{1}{3} \begin{pmatrix} 4 \\ 2 \end{pmatrix}}_{\mathbf{A}^{\dagger}\mathbf{1}}$$

Remark. Let $\mathbf{U} \in \mathbb{R}^{m\times n}$ be a tall matrix with orthonormal columns (e.g., an orthonormal basis matrix). Then it has full column rank, and

$$\mathbf{U}^T\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} [\mathbf{u}_1 \ldots \mathbf{u}_n] = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} = \mathbf{I}_n$$

It follows that

- $\mathbf{U}^\dagger = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T = \mathbf{U}^T$ (pseudoinverse), and
- $\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}\mathbf{U}^T$ (orthogonal projection matrix).

Example 0.11. Let

$$\mathbf{A} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

which has orthonormal columns. Therefore, the pseudoinverse of $\mathbf{A}$ is $\mathbf{A}^\dagger = \mathbf{A}^T$ and the orthogonal projection matrix is

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}\mathbf{A}^T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Theorem 0.6. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a diagonal matrix, i.e., all of its entries are zero except some of those along its diagonal. Then the pseudoinverse of $\mathbf{A}$ is another diagonal matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ such that

$$b_{ii} = \begin{cases} \frac{1}{a_{ii}}, & \text{if } a_{ii} \neq 0 \\ 0, & \text{if } a_{ii} = 0 \end{cases}$$

Proof. We verify this result using an example. Let

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \qquad \text{and} \quad \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{bmatrix}.$$

Then

$$\mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{BA} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

both of which are symmetric. Furthermore,

$$\mathbf{ABA} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} = \mathbf{A}$$

$$\mathbf{BAB} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{3} \\ 0 & 0 \end{bmatrix} = \mathbf{B}.$$

Thus, $\mathbf{B}$ is the pseudoinverse of $\mathbf{A}$. □

Theorem 0.7. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be an arbitrary matrix. Suppose its full SVD is $\mathbf{A} = \mathbf{U\Sigma V}^T$. Then the pseudoinverse of $\mathbf{A}$ is

$$\mathbf{A}^\dagger = \mathbf{V\Sigma}^\dagger\mathbf{U}^T$$

Proof We verify the four conditions directly:

$$\begin{aligned}
\mathbf{AA}^\dagger\mathbf{A} &= \mathbf{U\Sigma V}^T \cdot \mathbf{V\Sigma}^\dagger\mathbf{U}^T \cdot \mathbf{U\Sigma V}^T = \mathbf{U\Sigma\Sigma}^\dagger\mathbf{\Sigma V}^T = \mathbf{U\Sigma V}^T = \mathbf{A} \\
\mathbf{A}^\dagger\mathbf{AA}^\dagger &= \mathbf{V\Sigma}^\dagger\mathbf{U}^T \cdot \mathbf{U\Sigma V}^T \cdot \mathbf{V\Sigma}^\dagger\mathbf{U}^T = \mathbf{V\Sigma}^\dagger\mathbf{\Sigma\Sigma}^\dagger\mathbf{U}^T = \mathbf{V\Sigma}^\dagger\mathbf{U}^T = \mathbf{A}^\dagger \\
\mathbf{AA}^\dagger &= \mathbf{U\Sigma V}^T \cdot \mathbf{V\Sigma}^\dagger\mathbf{U}^T = \mathbf{U\Sigma\Sigma}^\dagger\mathbf{U}^T \qquad \text{(symmetric)} \\
\mathbf{A}^\dagger\mathbf{A} &= \mathbf{V\Sigma}^\dagger\mathbf{U}^T \cdot \mathbf{U\Sigma V}^T = \mathbf{V\Sigma}^\dagger\mathbf{\Sigma V}^T \qquad \text{(symmetric)}
\end{aligned}$$

Remark. The formula for $\mathbf{A}^\dagger$ is also in (full) SVD form:

$$\mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^T$$

It can be simplified to the compact SVD form

$$\mathbf{A}^\dagger = \mathbf{V}_r\mathbf{\Sigma}_r^{-1}\mathbf{U}_r^T$$

Thus, it suffices to find the compact SVD of $\mathbf{A}$ and use it to find $\mathbf{A}^\dagger$.

This simplified formula is computationally more efficient, as it avoids computing the redundant left/right singular vectors.

Example 0.12. Consider again the matrix (with compact SVD)

$$\underbrace{\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\mathbf{U}_2} \cdot \underbrace{\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix}}_{\mathbf{\Sigma}_2} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T}_{\mathbf{V}_2^T}$$

By the last theorem,

$$\mathbf{A}^\dagger = \underbrace{\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}}_{\mathbf{V}_2} \cdot \underbrace{\begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{pmatrix}}_{\mathbf{\Sigma}_2^{-1}} \cdot \underbrace{\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T}_{\mathbf{U}_2^T} = \frac{1}{3}\begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \end{pmatrix}$$

Let $\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T$ be the compact SVD of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$. We already know that the columns of $\mathbf{U}_r$ form an orthonormal basis for $\mathrm{Col}(\mathbf{A})$, and thus $\mathrm{Col}(\mathbf{U}_r) = \mathrm{Col}(\mathbf{A})$. Intuitively, the orthogonal projection matrices onto them must be the same, i.e.,

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}_r \mathbf{U}_r^T.$$

Consequently, we could just use the matrix $\mathbf{U}_r$ (which has orthonormal columns) to compute the orthogonal projection matrix.

This idea can be verified as follows:

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}_r \mathbf{\Sigma}_r \underbrace{\mathbf{V}_r^T \cdot \mathbf{V}_r}_{\mathbf{I}_r} \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^T = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^T = \mathbf{U}_r \mathbf{U}_r^T.$$

Example 0.13. In the preceding example, we have already obtained the compact SVD of the matrix $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Thus, we could compute the orthogonal projection matrix onto the column space of $\mathbf{A}$ as follows:

$$\mathbf{U}_2\mathbf{U}_2^T = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

It is the same with that obtained in Example 0.10.

Example 0.14. Find the pseudoinverse of $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$.

Observe that $\mathrm{rank}(\mathbf{A}) = 1$. Thus, we can obtain its compact SVD easily:

$$\mathbf{A} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \cdot \sqrt{2} \cdot \begin{pmatrix} 1 & 0 \end{pmatrix}$$

It follows that the orthogonal projection matrix is

$$\mathbf{A}\mathbf{A}^{\dagger} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

How to find $\mathbf{A}^{\dagger}$ by using the compact SVD?

MATLAB function for computing pseudoinverse

pinv Pseudoinverse.

$X = pinv(A)$ produces a matrix X of the same dimensions as A' so that $A * X * A = A, X * A * X = X$ and $A * X$ and $X * A$ are Hermitian. The computation is based on $SVD(A)$ and any singular values less than a tolerance are treated as zero.

$pinv(A, TOL)$ treats all singular values of A that are less than TOL as zero. By default, $TOL = max(size(A)) * eps(norm(A))$.

Applications of matrix pseudoinverse

- Linear least squares
- Minimum-norm solution to a consistent linear system

Linear least squares

Consider a system of linear equations $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A} \in \mathbb{R}^{m\times n}$ (not necessarily of full column rank) and $\mathbf{b} \in \mathbb{R}^m$.

In general, a vector $\mathbf{x}$ that solves the system may not exist, or if one does exist, it may not be unique.

In either case, we seek a least squares solution instead by solving the following general least squares problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|$$

This problem always has a solution, as the next slide shows.

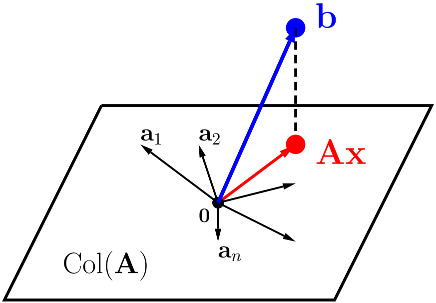
Theorem 0.8. A minimizer of the general least squares problem is

$$\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b}.$$

Proof. Since $\mathbf{Ax} \in \mathrm{Col}(\mathbf{A})$, the optimal $\mathbf{x}$ should be such that

$$\mathbf{Ax} = (\mathbf{AA}^\dagger)\mathbf{b}$$

Obviously, $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b}$ solves this equation and thus is a solution of the least squares problem (but it might not be the only solution). □

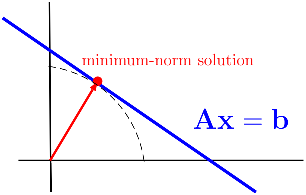


Remark. If $\mathbf{A}$ has full column rank (i.e., $\mathrm{rank}(\mathbf{A}) = n \le m$), then the least squares solution is unique: $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.

Minimum-norm solution to a consistent linear system

For under-determined systems $\mathbf{A}\mathbf{x} = \mathbf{b}$, the pseudoinverse may be used to construct the solution with minimum Euclidean norm among all solutions.

Theorem 0.9. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. If the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has solutions, then $\mathbf{x}^* = \mathbf{A}^\dagger\mathbf{b}$ is an exact solution and has the smallest possible norm, i.e., $\|\mathbf{x}^*\| \leq \|\mathbf{x}\|$ for all solutions $\mathbf{x}$.



Proof. First, since $\mathbf{A}^\dagger$ is a generalized inverse, it must be a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. To show that it has the smallest possible norm, for any solution $\mathbf{x} \in \mathbb{R}^n$, consider its orthogonal decomposition via $\mathbf{A}^\dagger\mathbf{A} \in \mathbb{R}^{n\times n}$:

$$\mathbf{x} = (\mathbf{A}^\dagger\mathbf{A})\mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} + (\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{x}$$

It follows that

$$\|\mathbf{x}\|^2 = \|\mathbf{A}^\dagger\mathbf{b}\|^2 + \|(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A})\mathbf{x}\|^2 \geq \|\mathbf{A}^\dagger\mathbf{b}\|^2$$

This shows that $\|\mathbf{x}\| \geq \|\mathbf{A}^\dagger\mathbf{b}\|$. □

Summary

- **Generalized inverse** $\mathbf{G} \in \mathbb{R}^{n\times m}$ for a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$:
  - *Definition*: $\mathbf{AGA} = \mathbf{A}$
  - *Existence*: $\mathbf{G}$ always exists but might not be unique
  - *Computing*: $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \longrightarrow \mathbf{G} = \begin{bmatrix} A_{11}^{-1} & O \\ O & O \end{bmatrix}$, if $A_{11} \in \mathbb{R}^{r\times r}, r = \mathrm{rank}(\mathbf{A})$ is invertible.
  - *Property*: $\mathbf{AG}$ is a projection matrix onto $\mathrm{Col}(\mathbf{A})$
  - *Application*: $\mathbf{x} = \mathbf{Gb}$ is a solution to $\mathbf{Ax} = \mathbf{b}$ (if consistent)

- **Pseudoinverse** $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ for a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$:
  - *Definition*: $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}^\dagger$, and $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}$, and both $\mathbf{A}\mathbf{A}^\dagger, \mathbf{A}^\dagger\mathbf{A}$ are symmetric
  - *Existence*: $\mathbf{A}^\dagger$ always exists and is unique
  - *Computing*:
    * If $\mathbf{A}$ has full column rank: $\mathbf{A}^\dagger = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$
    * If $\mathbf{A}$ is "diagonal": $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ is also "diagonal" with reciprocals of nonzero diagonals of $\mathbf{A}$

    * In general: $\mathbf{A}^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^T$ (using compact SVD $\mathbf{A} = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T$)

- *Property*: $\mathbf{A}\mathbf{A}^\dagger$ is an orthogonal projection matrix onto $\mathrm{Col}(\mathbf{A})$, and $\mathbf{A}\mathbf{A}^\dagger = \mathbf{U}_r \mathbf{U}_r^T$
- *Application*: For any $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$, the vector $\mathbf{A}^\dagger \mathbf{b}$ solves the least squares problem

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

    * If $\mathbf{A}$ has full column rank, then the solution is unique.
    * If $\mathbf{A}\mathbf{x} = \mathbf{b}$ has exact solutions, then $\mathbf{A}^\dagger \mathbf{b}$ is the minimum-norm solution.

Next time: Matrix norm and low-rank approximation

Read the book chapter on the topic.