For the matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$:

General Solution to a Linear System

If $\boldsymbol{b} \in \mathcal{C}(\mathbf{A})$
$$\boldsymbol{x}_g = \boldsymbol{x}_p + \boldsymbol{x}_n$$
where $\boldsymbol{x}_p \in \mathbb{R}^n$ is a particular solution obtained by backward substitution, and $\boldsymbol{x}_n \in \mathcal{N}(\mathbf{A})$.

Minimum Norm Solution

$\boldsymbol{x}_r \in \mathcal{C}(\mathbf{A}^\top)$ where $\boldsymbol{x}_r = \operatorname{proj}_{\mathcal{C}(\mathbf{A}^\top)}(\boldsymbol{x}_g)$.

Least Squares

For $\boldsymbol{b} \notin \mathcal{C}(\mathbf{A})$, find $\boldsymbol{x}$:
$$\boldsymbol{x} = \underset{\boldsymbol{x}^* \in \mathbb{R}^n}{\arg\min} \|\boldsymbol{b} - \mathbf{A}\boldsymbol{x}^*\|$$
so that $\boldsymbol{b} - \mathbf{A}\boldsymbol{x}$ is orthogonal to $\mathcal{C}(\mathbf{A})$.
$$\mathbf{A}^\top(\boldsymbol{b} - \mathbf{A}\boldsymbol{x}) = \mathbf{0}$$
$$\mathbf{A}^\top\mathbf{A}\boldsymbol{x} = \mathbf{A}^\top\boldsymbol{b}$$
$$\boldsymbol{x} = \left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\boldsymbol{b}.$$

Orthogonal Projection

$$\mathbf{P} = \mathbf{A}\left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top$$

$\mathbf{P}$ operates on $\boldsymbol{b}$:
$$\boldsymbol{b}_P = \mathbf{P}\boldsymbol{b} = \operatorname{proj}_{\mathcal{C}(\mathbf{A})}(\boldsymbol{b}) = \mathbf{A}\boldsymbol{x}$$
$$= \mathbf{A}\left(\mathbf{A}^\top\mathbf{A}\right)^{-1}\mathbf{A}^\top\boldsymbol{b}$$
$\mathbf{P}$ is idempotent, $\mathbf{P}^2 = \mathbf{P}$.

Dependent Columns

If $\operatorname{nullity}(\mathbf{A}) > 0$, the Normal Equations yields infinitely many solutions
$$\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n$$
because $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{A}^\top\mathbf{A})$.

Orthogonal Complement Projections

Given $\mathbf{P} = \operatorname{proj}_V$:
$$\mathbf{Q} = \operatorname{proj}_{V^\perp} = \mathbf{I} - \mathbf{P}$$
because
$$\boldsymbol{b} = \operatorname{proj}_V(\boldsymbol{b}) + \operatorname{proj}_{V^\perp}(\boldsymbol{b}) = \mathbf{P}\boldsymbol{b} + \mathbf{Q}\boldsymbol{b}.$$
Additionally,
$$(\mathbf{P}\boldsymbol{b})^\top\mathbf{Q}\boldsymbol{b} = 0$$
$$\mathbf{P}\mathbf{Q} = \mathbf{0}$$
where $\mathbf{0}$ is the zero matrix.

Orthogonal Matrices

Change of Basis

Given the basis $W = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ solve:
$$\boldsymbol{b} = c_1\boldsymbol{w}_1 + \cdots + c_n\boldsymbol{w}_n$$
$$\boldsymbol{b} = \mathbf{W}\boldsymbol{c}$$
$$(\boldsymbol{b})_W = \boldsymbol{c}.$$

Orthonormal Basis

Every basis vector is normalised and orthogonal to every other basis vector. For $Q = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$, $\boldsymbol{q}_i^\top\boldsymbol{q}_j = \delta_{ij}$, where
$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$
Solving a system:
$$\boldsymbol{b} = c_1\boldsymbol{q}_1 + \cdots + c_n\boldsymbol{q}_n$$
$$\boldsymbol{q}_i^\top\boldsymbol{b} = c_i$$
$$\mathbf{Q}^\top\boldsymbol{b} = \boldsymbol{c} = (\boldsymbol{b})_Q$$

Orthogonal Matrices

$$\mathbf{Q}^\top = \mathbf{Q}^{-1} \iff \mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}.$$

Projection onto a Vector

$$\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{b}) = \boldsymbol{b}_P = \boldsymbol{a}\left(\boldsymbol{a}^\top\boldsymbol{a}\right)^{-1}\boldsymbol{a}^\top\boldsymbol{b}$$
$$= \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}\cdot\boldsymbol{b}$$
Using a unit vector $\boldsymbol{q}$:
$$\operatorname{proj}_{\boldsymbol{q}}(\boldsymbol{b}) = \boldsymbol{q}(\boldsymbol{q}\cdot\boldsymbol{b})$$

Gram-Schmidt Process

Converts the basis W that spans $\operatorname{span}(\mathbf{A})$ to an orthonormal basis Q
$$\boldsymbol{v}_1 = \boldsymbol{w}_1 \qquad \boldsymbol{q}_1 = \boldsymbol{v}_1/\|\boldsymbol{v}_1\|$$
$$\boldsymbol{v}_2 = \boldsymbol{w}_2 - \boldsymbol{q}_1(\boldsymbol{q}_1\cdot\boldsymbol{w}_2) \qquad \boldsymbol{q}_2 = \boldsymbol{v}_2/\|\boldsymbol{v}_2\|$$
$$\vdots \qquad \vdots$$
$$\boldsymbol{v}_i = \boldsymbol{w}_i - \sum_{j=1}^{i-1}\boldsymbol{q}_j(\boldsymbol{q}_j\cdot\boldsymbol{w}_i) \qquad \boldsymbol{q}_i = \boldsymbol{v}_i/\|\boldsymbol{v}_i\|.$$
V and Q span W and V is orthogonal.

QR Decomposition

$$\mathbf{A} = \mathbf{Q}\mathbf{R}.$$

$\mathbf{R}$ is upper-triangular. QR decomposition also finds the Least Squares solution.

Eigenvalues and Eigenvectors

$$(\lambda\mathbf{I} - \mathbf{A})\boldsymbol{v} = \mathbf{0} : \boldsymbol{v} \neq \mathbf{0}$$

Characteristic Polynomial

$$P(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}) = 0.$$

Eigen Decomposition

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{D} \iff \mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}.$$

$\mathbf{V} = [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_n]$ $\mathbf{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Algebraic Multiplicity $\mu_{\mathbf{A}}(\lambda_i)$

The multiplicity of λ_i in $P(\lambda)$, for $d \leq n$ distinct eigenvalues,
$$P(\lambda) = (\lambda - \lambda_1)^{\mu_{\mathbf{A}}(\lambda_1)}\cdots(\lambda - \lambda_d)^{\mu_{\mathbf{A}}(\lambda_d)}.$$
This requires
$$1 \leq \mu_{\mathbf{A}}(\lambda_i) \leq n$$
$$\mu_{\mathbf{A}} = \sum_{i=1}^{d}\mu_{\mathbf{A}}(\lambda_i) = n$$

Geometric Multiplicity $\gamma_{\mathbf{A}}(\lambda_i)$

The dimension of the eigenspace associated with λ_i.
$$\gamma_{\mathbf{A}}(\lambda_i) = \operatorname{nullity}(\lambda_i\mathbf{I} - \mathbf{A}).$$
Given $d \leq n$ distinct eigenvalues,
$$1 \leq \gamma_{\mathbf{A}}(\lambda_i) \leq \mu_{\mathbf{A}}(\lambda_i) \leq n$$
$$\gamma_{\mathbf{A}} = \sum_{i=1}^{d}\gamma_{\mathbf{A}}(\lambda_i)$$
so that
$$d \leq \gamma_{\mathbf{A}} \leq n.$$

Eigenvectors corresponding to distinct eigenvalues are linearly dependent.

Defective Matrix

$\mathbf{A}$ lacks a complete eigenbasis, $\exists\lambda_k : \gamma_{\mathbf{A}}(\lambda_k) < \mu_{\mathbf{A}}(\lambda_k)$.

Matrix Similarity

$\mathbf{A}$ and $\mathbf{B}$ are similar if
$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$
They share $P(\lambda)$, ranks, determinants, trace, and eigenvalues (including μ and γ).

Symmetric Matrices

$\mathbf{A} \in \mathbb{R}^{n\times n}$ where $\mathbf{A}^\top = \mathbf{A}$.

1. $\mathbf{A}$ is always diagonalisable.
2. Eigenvalues and eigenvectors are always real.
3. The eigenspaces are orthogonal.

$$\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top$$

Skew-Symmetric Matrices

$\mathbf{A} \in \mathbb{R}^{n\times n}$ where $\mathbf{A}^\top = -\mathbf{A}$.

1. Eigenvalues are purely imaginary.

Definite Matrices

Symmetric matrices are also known as definite matrices, and can be classified into four categories.

1. Positive definite matrices: All eigenvalues are positive.
2. Positive semidefinite matrices: All eigenvalues are nonnegative.
3. Negative definite matrices: All eigenvalues are nonpositive.
4. Negative semidefinite matrices: All eigenvalues are negative.

Matrix Functions

Given a nondefective $\mathbf{A} \in \mathbb{R}^{n\times n}$:
$$f(\mathbf{A}) = \mathbf{V}f(\mathbf{D})\mathbf{V}^{-1}$$
$$= \mathbf{V}\operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n))\mathbf{V}^{-1}.$$
for an analytic function f.

Cayley-Hamilton Theorem

$$\forall\mathbf{A} \in \mathbb{R}^{n\times n} : P(\mathbf{A}) = \mathbf{0}$$

Singular Value Decomposition

The singular value decomposition (SVD) of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is given by
$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$
where $\mathbf{U} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix, $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is a diagonal matrix, and $\mathbf{V} \in \mathbb{R}^{m\times m}$ is an orthogonal matrix. $\mathbf{U}$ is known as the left-singular matrix and $\mathbf{V}$ is the right-singular matrix, corresponding to how these matrices are multiplied to $\boldsymbol{\Sigma}$.
$\boldsymbol{\Sigma}$ consists of the **singular values** σ_i of $\mathbf{A}$, which can be determined using the following process:
$$\mathbf{A}^\top\mathbf{A} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\right)^\top\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\right)$$
$$= \mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$
$$= \mathbf{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\mathbf{V}^\top$$

similarly,

$$\begin{aligned}\mathbf{A}\mathbf{A}^\top &= \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\right)\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\right)^\top \\ &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top \\ &= \mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\end{aligned}$$

In both instances, we form an orthogonal eigendecomposition where $\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ are the eigenvalues of $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$, respectively. But because the eigenvalues of $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ are equal, $\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$.

The singular values are non-negative constants and the entries of $\boldsymbol{\Sigma}$ are always non-increasing: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ where r is the rank of $\mathbf{A}$.

Orthonormal Bases for the Fundamental Subspaces

For $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(\mathbf{A}) = r \leq n$:

$$\mathbf{A}\mathbf{V} = \mathbf{U}\boldsymbol{\Sigma}$$

with

$$\mathbf{A}\mathbf{V} = \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_r & \boldsymbol{a}_{r+1} & \cdots & \boldsymbol{a}_n \\ | & & | & | & & | \end{bmatrix}\begin{bmatrix} | & & | & | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_r & \boldsymbol{v}_{r+1} & \cdots & \boldsymbol{v}_n \\ | & & | & | & & | \end{bmatrix}$$

$$\mathbf{U}\boldsymbol{\Sigma} = \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \cdots & \boldsymbol{u}_m \\ | & & | & | & & | \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & 0 & \\ & \ddots & \\ & 0 & \\ & & \end{bmatrix}\begin{array}{l} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} r \text{ rows} \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} n-r \text{ rows} \\ \left.\vphantom{\begin{matrix}a\\b\end{matrix}}\right\} m-n \text{ rows}\end{array}$$

where we have r singular values because $\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{A}^\top\mathbf{A}$. If we express each equation separately, then

$$\begin{aligned}\mathbf{A}\boldsymbol{v}_1 &= \sigma_1\boldsymbol{u}_1 \\ &\vdots \\ \mathbf{A}\boldsymbol{v}_r &= \sigma_2\boldsymbol{u}_r \\ \mathbf{A}\boldsymbol{v}_{r+1} &= \mathbf{0} \\ &\vdots \\ \mathbf{A}\boldsymbol{v}_n &= \mathbf{0}\end{aligned}$$

which shows us that $\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$ forms a basis for the null space of $\mathbf{A}$, requiring the remaining columns of $\mathbf{V}$ to form a basis for the row space of $\mathbf{A}$.

This means that $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$ also forms a basis for the column space of $\mathbf{A}$, and hence, the remaining columns of $\mathbf{U}$ form a basis for the left-null space of $\mathbf{A}$.

To summarise:

$$\mathcal{C}\left(\mathbf{A}^\top\right) = \operatorname{span}\left(\{\boldsymbol{v}_{i\leq r}\}\right) \qquad \mathcal{C}\left(\mathbf{A}\right) = \operatorname{span}\left(\{\boldsymbol{u}_{i\leq r}\}\right)$$
$$\mathcal{N}\left(\mathbf{A}\right) = \operatorname{span}\left(\{\boldsymbol{v}_{r<i\leq n}\}\right) \quad \mathcal{N}\left(\mathbf{A}^\top\right) = \operatorname{span}\left(\{\boldsymbol{u}_{r<i\leq m}\}\right)$$

for $i \in \mathbb{N}$. Additionally, $\mathbf{V}$ forms an orthonormal basis for $\mathbb{R}^n$ while $\mathbf{U}$ forms an orthonormal basis for $\mathbb{R}^m$.

Singular Bases

Consider the bases V and U from the columns of the orthogonal matrices $\mathbf{V}$ and $\mathbf{U}$, respectively. Let $(\boldsymbol{x})_V = \boldsymbol{c}$ and $(\boldsymbol{b})_U = \boldsymbol{k}$, so that the system $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ becomes

$$\begin{aligned}\mathbf{A}\boldsymbol{x} &= \boldsymbol{b} \\ \mathbf{A}\mathbf{V}\boldsymbol{c} &= \mathbf{U}\boldsymbol{k} \\ \mathbf{U}^\top\mathbf{A}\mathbf{V}\boldsymbol{c} &= \boldsymbol{k} \\ \boldsymbol{\Sigma}\boldsymbol{c} &= \boldsymbol{k} \\ \boldsymbol{\Sigma}\left(\boldsymbol{x}\right)_V &= \left(\boldsymbol{b}\right)_U.\end{aligned}$$

Therefore, with respect to the orthogonal bases V and U, the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is represented by the matrix $\boldsymbol{\Sigma}$.

Reduced SVD

By ignoring the additional $m - n$ rows in $\boldsymbol{\Sigma}$, we can form the reduced SVD which removes the additional "0" rows of $\boldsymbol{\Sigma}$. This results in $\mathbf{U} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$.

$$\mathbf{U}\boldsymbol{\Sigma} = \begin{bmatrix} | & & | & | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \cdots & \boldsymbol{u}_n \\ | & & | & | & & | \end{bmatrix}\begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

The reduced SVD also removes the $m-n$ left-null space basis vectors $\{\boldsymbol{u}_{n<i\leq m}\}$ from $\mathbf{U}$.

Pseudoinverse

Using the orthogonal basis vectors obtained through the SVD of $\mathbf{A} \in \mathbb{R}^{m\times n}$, we can show that $\boldsymbol{v}_i \mapsto \sigma_i\boldsymbol{u}_i$ for all $i \leq r$. Therefore

$$\mathbf{A}\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i.$$

If we consider the inverse mapping $\boldsymbol{u}_i \mapsto \frac{1}{\sigma_i}\boldsymbol{v}_i$, then

$$\mathbf{A}^\dagger\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{v}_i$$

where $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$. To determine $\mathbf{A}^\dagger$, the above relationship must hold for all $i \leq r$. If we multiply the RHS by $\boldsymbol{u}_i^\top\boldsymbol{u}_i$

$$\mathbf{A}^\dagger\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top\boldsymbol{u}_i$$

we can show that $\mathbf{A}^\dagger$ takes the form $\frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top$. Using the Kronecker Delta definition of orthonormal vectors, we know that the product of two orthonormal basis vectors is 1 for $i = j$ and 0 for $i \neq j$. Therefore by taking the sum of all $\frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top$, we have

$$\begin{aligned}\mathbf{A}^\dagger\boldsymbol{u}_1 &= \left(\frac{1}{\sigma_1}\boldsymbol{v}_1\boldsymbol{u}_1^\top + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\boldsymbol{u}_r^\top\right)\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{v}_1\cancelto{1}{\boldsymbol{u}_1^\top\boldsymbol{u}_1} + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\cancelto{0}{\boldsymbol{u}_r^\top\boldsymbol{u}_1} = \frac{1}{\sigma_1}\boldsymbol{v}_1 \\ &\vdots \\ \mathbf{A}^\dagger\boldsymbol{u}_r &= \left(\frac{1}{\sigma_1}\boldsymbol{v}_1\boldsymbol{u}_1^\top + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\boldsymbol{u}_r^\top\right)\boldsymbol{u}_r = \frac{1}{\sigma_1}\boldsymbol{v}_1\cancelto{0}{\boldsymbol{u}_1^\top\boldsymbol{u}_r} + \cdots + \frac{1}{\sigma_r}\boldsymbol{v}_r\cancelto{1}{\boldsymbol{u}_r^\top\boldsymbol{u}_r} = \frac{1}{\sigma_r}\boldsymbol{v}_r\end{aligned}$$

Therefore the pseudoinverse is given by

$$\mathbf{A}^\dagger = \sum_{i=1}^r \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^\top$$

which is equivalent to

$$\mathbf{A}^\dagger = \mathbf{V}\begin{bmatrix} \frac{1}{\sigma_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{\sigma_r} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\mathbf{U}^\top$$

If we consider the SVD of $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \mathbf{U}_\Sigma\boldsymbol{\Sigma}_\Sigma\mathbf{V}_\Sigma^\top.$$

we have $\mathbf{U}_\Sigma = \mathbf{I}_m$, $\boldsymbol{\Sigma}_\Sigma = \boldsymbol{\Sigma}$, and $\mathbf{V}_\Sigma = \mathbf{I}_n$. Then the pseudoinverse of this matrix is then

$$\boldsymbol{\Sigma}^\dagger = \sum_{i=1}^r \frac{1}{\sigma_i}\boldsymbol{v}_{\Sigma,\,i}\boldsymbol{u}_{\Sigma,\,i}^\top$$

$$= \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_r} \\ 0 & & \\ & \ddots & \\ & & 0 \end{bmatrix}\begin{array}{l} \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} r \text{ rows} \\ \left.\vphantom{\begin{matrix}a\\b\\c\end{matrix}}\right\} n-r \text{ rows}\end{array} \quad 0$$

so that the pseudoinverse of $\mathbf{A}$ can be determined using

$$\mathbf{A}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^\dagger\mathbf{U}^\top.$$

Note that the pseudoinverse can also be obtained using the reduced SVD or by using the first r columns of $\mathbf{U}$ and $\mathbf{V}$ with the $r \times r$ submatrix of $\boldsymbol{\Sigma}$.

Truncated SVD

By expanding the SVD of $\mathbf{A}$, we can express it as the sum of rank-1 matrices:

$$\mathbf{A} = \sum_{i=1}^n \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^\top.$$

However if $r < n$, then

$$\mathbf{A} = \sum_{i=1}^r \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^\top$$

as $\sigma_{r<i\leq n} = 0$. As the singular values are ordered from largest to smallest, if we wished to approximate $\mathbf{A}$ by a matrix of lower rank, we can truncate this sum even further at $i = \nu$ for $\nu < r$, to generate a rank-ν approximation of $\mathbf{A}$:

$$\tilde{\mathbf{A}} = \sum_{i=1}^\nu \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^\top.$$

This allows us to represent the matrix $\mathbf{A}$ using only ν singular values and 2ν singular vectors ($\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ for $i = 1\ldots\nu$). This approximate decomposition is known as the truncated SVD.

Principal Component Analysis

To summarise, the SVD of $\mathbf{A}$ has three variations in which the dimensions of the decomposition matrices change.

- Full SVD:
$$\underset{m\times n}{\mathbf{A}} = \underset{m\times m}{\mathbf{U}}\ \underset{m\times n}{\boldsymbol{\Sigma}}\ \underset{n\times n}{\mathbf{V}}^\top$$
- Reduced SVD:
$$\underset{m\times n}{\mathbf{A}} = \underset{m\times n}{\mathbf{U}}\ \underset{n\times n}{\boldsymbol{\Sigma}}\ \underset{n\times n}{\mathbf{V}}^\top$$
- Truncated SVD:
$$\underset{m\times n}{\mathbf{A}} = \underset{m\times \nu}{\mathbf{U}}\ \underset{\nu\times \nu}{\boldsymbol{\Sigma}}\ \underset{n\times \nu}{\mathbf{V}}^\top$$

Consider the ith column of $\mathbf{A}$ in the truncated SVD,

$$\tilde{\mathbf{A}}_{:,\,i} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_{i,\,:}^\top$$

If we let $\boldsymbol{x}_i = \mathbf{\Sigma}\mathbf{V}_{i,\,:}^\top$, then

$$\tilde{\mathbf{A}}_{:,\,i} = \mathbf{U}\boldsymbol{x}_i$$

so that $\boldsymbol{x}_i \in \mathbb{R}^\nu$ is the coordinate vector of $\tilde{\mathbf{A}}_{:,\,i}$ with respect to the basis of left singular vectors $\mathbf{U}_{:,\,i\leq\nu}$. In statistics or machine learning, the columns of $\mathbf{U}$ are called "features" as they are the most important singular vectors used to construct $\mathbf{A}$. The vector $\boldsymbol{x}_i$ is then the coordinate vector of $\mathbf{A}$'s projection onto the "feature space".

In data analysis, this process is referred to as Principal Componenet Analysis (PCA) where the matrix $\mathbf{A}$ represents a set of observations in which each column contains a particular explanatory variable that one might be interested in. The model shown above can be used to explain the observations from each variable in that dataset.

General Vector Spaces

Vector Space Axioms

A set V of objects are called "vectors" if the following additive and multiplicative axioms are satisfied $\forall \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $\forall k, m \in \mathbb{R}$.

Axiom	Meaning
Closure under vector addition	$\boldsymbol{u} + \boldsymbol{v} \in V$
Commutativity of vector addition	$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
Associativity of vector addition	$\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
Identity element of vector addition	$\exists \mathbf{0} \in V : \boldsymbol{u} + \mathbf{0} = \mathbf{0} + \boldsymbol{u} = \boldsymbol{u}$
Inverse elements of vector addition	$\exists (-\boldsymbol{u}) \in V : \boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$

Table 1: Additive axioms.

Axiom	Meaning
Closure under scalar multiplication	$k\boldsymbol{u} \in V$
Distributivity of scalar multiplication with vector addition	$k(\boldsymbol{u} + \boldsymbol{v}) = k\boldsymbol{u} + k\boldsymbol{v}$
Distributivity of scalar multiplication with scalar addition	$(k + m)\boldsymbol{u} = k\boldsymbol{u} + m\boldsymbol{u}$
Associativity of scalar multiplication	$k(m\boldsymbol{u}) = (km)\boldsymbol{u}$
Identity element of scalar multiplication	$1\boldsymbol{u} = \boldsymbol{u}$

Table 2: Multiplicative axioms.

If these axioms are satisfied, the objects called "vectors" and the operators "addition" (+) and "scalar multiplication" (denoted by juxtaposition) form a general vector space V. It follows that the operations of addition and scalar multiplication need not resemble those in $\mathbb{R}^n$ to satisfy the 10 axioms described above. Rather any set with two operations can form a vector space if they satisfy the 10 axioms above.

Examples of Vector Spaces

Aside from the familiar vector space $\mathbb{R}^n$, we can also consider the following spaces which satisfy the 10 axioms above.

1. The set of $m \times n$ matrices $\mathscr{M}_{mn}$ with matrix addition and scalar multiplication.
2. The set of functions $\mathscr{F}(\Omega) : \Omega \to \mathbb{R}$ with addition and scalar multiplication defined pointwise.

Subspaces

Consider the subset W of a vector space V, so that $W \subset V$. For W to be a subspace of V, it must also satisfy the 10 axioms shown above. Now as W is a subset of V, 6 axioms are automatically inherited from the enclosing space V.

Therefore only the following axioms need to be satisfied in W:

- Axiom 1: Closure under vector addition
- Axiom 4: Identity element of vector addition
- Axiom 5: Inverse elements of vector addition
- Axiom 6: Close under scalar multiplication

However, if Axioms 1 and 6 are established, then Axioms 4 and 5 will inherit from the vector space structure of V. So it suffices to check only Axioms 1 and 6.

Therefore any subset W of a vector that is closed under vector addition and scalar multiplication is a subspace of that vector space.

5.3.1 Examples of Subspaces

Subspaces of $\mathbb{R}^n$:

1. Lines, planes and higher-dimensional analogues in $\mathbb{R}^n$ *passing through the origin.*

Subspaces of $\mathscr{M}_{mn}$:

1. The set of all *symmetric* $n \times n$ matrices $\mathbf{A}$ such that $\mathbf{A} = \mathbf{A}^\top$, denoted $\mathscr{S}_n \subset \mathscr{M}_{nn}$.
2. The set of all *skew symmetric* $n \times n$ matrices $\mathbf{A}$ such that $\mathbf{A} = -\mathbf{A}^\top$, denoted $\mathscr{K}_n \subset \mathscr{M}_{nn}$.

Subspaces of $\mathscr{F}(\Omega)$:

1. The set of all *polynomials* of degree n or less, denoted $\mathscr{P}_n \subset \mathrm{F}(\Omega)$.
2. The set of all *continuous functions*, denoted $C(\Omega) \subset \mathscr{F}(\Omega)$.
3. The set of all functions with *continuous derivatives* on Ω, for example, $C^1(\Omega) \subset C(\Omega)$ is the set of all functions with continuous first derivatives.
4. The set of all functions f defined on $[0, 1]$ satisfying $f(0) = f(1)$.

General Vector Space Terminology

The notions of linear combination, linear independence and span are all unchanged for general vector spaces.

Let $c_1, \ldots, c_k \in \mathbb{R}$ be scalars:

- The linear combination of a set of vector $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is a vector of the form $\boldsymbol{v} = c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k$.
- A set of vectors $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is linearly independent if the only solution to $c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$ is the trivial solution $c_1 = \cdots = c_k = 0$.
- The span of a set of vectors $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is the set of all linear combinations of that set, denoted $\mathrm{span}(S)$.

A set of vectors $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is a *basis* for a vector space V if

- S is linearly independent.
- $\mathrm{span}(S) = V$.

The vector space is of dimension k there are k vectors in its basis. Note that not all vector spaces have a basis. For example, function spaces, such as C are infinite-dimensional.

The standard bases for some vector spaces are shown below:

- $\mathbb{R}^3$: $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$
- $\mathscr{M}_{22}$: $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
- $\mathscr{S}_{22}$: $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$
- $\mathscr{K}_{22}$: $S = \left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$
- $\mathscr{P}_3$: $S = \{1, x, x^2, x^3\}$

Linear Transformations

A linear transformation T is a mapping from a vector space V to a vector space W

$$T : V \to W$$

satisfying the following properties for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and all scalars $k \in \mathbb{R}$:

1. $T(k\boldsymbol{u}) = kT(\boldsymbol{u})$
2. $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$

These defining properties allow us to characterise a linear transformation completely by considering how the basis vectors from V map to W. Any vector $\boldsymbol{v} \in V$ can be written in terms of a basis $B = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$, such that

$$\boldsymbol{v} = x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n.$$

Therefore

$$\begin{aligned} T(\boldsymbol{v}) &= T(x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n) \\ &= T(x_1\boldsymbol{v}_1) + \cdots + T(x_n\boldsymbol{v}_n) \\ &= x_1 T(\boldsymbol{v}_1) + \cdots + x_n T(\boldsymbol{v}_n) \end{aligned}$$

Now consider the coordinate vector of $\boldsymbol{w} \in W$ relative to the basis $B' = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_m\}$ is

$$\begin{aligned} \boldsymbol{w} &= b_1\boldsymbol{w}_1 + \cdots + b_m\boldsymbol{w}_m \\ &= B'\boldsymbol{b} \\ &= B'(\boldsymbol{w})_{B'} \end{aligned}$$

so that the linear transformation can be expressed as follows

$$\begin{aligned} T(\boldsymbol{v}) &= \boldsymbol{w} \\ x_1 T(\boldsymbol{v}_1) + \cdots + x_n T(\boldsymbol{v}_n) &= b_1\boldsymbol{w}_1 + \cdots + b_m\boldsymbol{w}_m \\ \begin{bmatrix} | & & | \\ T(\boldsymbol{v}_1) & \cdots & T(\boldsymbol{v}_n) \\ | & & | \end{bmatrix} \boldsymbol{x} &= B'\boldsymbol{b} \\ \begin{bmatrix} | & & | \\ (T(\boldsymbol{v}_1))_{B'} & \cdots & (T(\boldsymbol{v}_n))_{B'} \\ | & & | \end{bmatrix} \boldsymbol{x} &= \boldsymbol{b} \\ \mathbf{A}\boldsymbol{x} &= \boldsymbol{b} \\ (T)_{B',B}(\boldsymbol{v})_B &= (\boldsymbol{w})_{B'}. \end{aligned}$$

Therefore the linear transformation between the vector spaces V and W can be represented as the transformation of coordinate vectors relative to the bases B and B', denoted $(T)_{B',B}$, that is, the matrix $\mathbf{A}$.

Definition 5.1 (Isomorphism)**.** A linear transformation $T : V \to W$ is an isomorphism between V and W if there exists a bijection between the two vector spaces.

All n dimensional vector spaces V are isomorphic to $\mathbb{R}^n$. This is a result of the coordinate vectors of V with respect to the basis B that allow us to represent each vector $\boldsymbol{v} \in V$ as a linear combination of the standard basis vectors in $\mathbb{R}^n$.

Fundamental Subspaces of T

The four fundamental subspaces also generalise to arbitrary linear transformations.

Given the linear transformation $T : V \to W$:

- The set of all vectors in V that map to W is the **image** of T, denoted $\operatorname{im}(T)$.
- The set of all vectors in W that is mapped to by a vector in V is the **range** of T, denoted $\operatorname{range}(T)$.
- The set of all vectors in V that T maps to $\mathbf{0}_W$ is the **kernel** of T, denoted $\ker(T)$.

If the range of T is finite-dimensional, its dimension is the **rank** of T, and if the kernel of T is finite-dimensional, its dimension is the **nullity** of T, so that the rank-nullity theorem continues to hold.

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V).$$

Inner Product Spaces

To describe geometric properties of vector spaces, we introduce an operation called the inner product that acts on the vectors in a vector space. The inner product associates a pair of vectors to a real number, and is delimited by angle brackets

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}.$$

This operation must satisfy the following axioms. For $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $k \in \mathbb{R}$:

Axiom	Meaning
Symmetry	$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$
Linearity in the first argument	$\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
Linearity in the first argument	$\langle k\boldsymbol{u}, \boldsymbol{v} \rangle = k\langle \boldsymbol{u}, \boldsymbol{v} \rangle$
Positive semi-definitiveness	$\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$, $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0 \iff \boldsymbol{u} = \mathbf{0}$

Table 3: Inner product axioms.

A vector space that defines such an operation called an inner product space.

5.7.1 Examples of Inner Products

In $\mathbb{R}^n$:

- $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^\top \boldsymbol{v}$. This is the standard inner product called the "dot product".
- $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^\top \mathbf{A}\boldsymbol{v}$ where $\mathbf{A}$ is positive definite. This is a weighted inner product, which can be used in weighted least squares.

For matrices $\mathbf{A}, \mathbf{B} \in \mathscr{M}_{mn}$, the standard inner product is defined

$$\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{Tr}(\mathbf{A}^\top \mathbf{B}).$$

For continuous function spaces, consider $f, g \in C([a, b])$ where the inner product operation is defined by the integral

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, \mathrm{d}x.$$

As the vector spaces $\mathscr{M}_{mn}$ and $\mathscr{P}_n$ are isomorphic to $\mathbb{R}^{mn}$ and $\mathbb{R}^{n+1}$ respectively, we can use the inner product definitions from $\mathbb{R}^n$ for these spaces also.

Likewise we can also consider the following integral definition with a continuous weight function $w(x)$ that is positive for all $x \in [a, b]$:

$$\langle f, g \rangle = \int_a^b f(x)\, g(x)\, w(x)\, \mathrm{d}x.$$

5.7.2 Norms

Having defined an inner product, we can also define the norm as

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

From this definition, we maintain the expected properties of the Euclidean norm:

- $\|\boldsymbol{v}\| \geq 0$ and $\|\boldsymbol{v}\| = 0$ iff $\boldsymbol{v} = \mathbf{0}$.
- $\|k\boldsymbol{v}\| = |k|\|\boldsymbol{v}\|$ for $k \in \mathbb{R}$.
- $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$, which is the triangle inequality.

For matrices, the inner product inherited from $\mathbb{R}^{mn}$ leads to the following definitions of norms:

$$\|\mathbf{A}\| = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{\operatorname{Tr}(\mathbf{A}^\top \mathbf{A})} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$$

which is known as the Frobenius norm.

For continuous functions $f \in C([a, b])$:

$$\|f(x)\| = \sqrt{\langle f(x), f(x) \rangle} = \sqrt{\int_a^b f(x)^2\, \mathrm{d}x}.$$

5.7.3 Orthogonality

Similarly, we can say that two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if

$$\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0.$$

Using this definition we can show that all