Math for ML

Tanvir

Matrix multiplication

Element-wise

$A^{m\times n} \;\; B^{n\times p} = C^{m\times p}$

$$c_{i,j} = \begin{pmatrix} \cdots a_i \cdots \end{pmatrix} \cdot \begin{pmatrix} \cdot \\ \cdot \\ b_j \\ \cdot \\ \cdot \end{pmatrix}$$

```
np.einsum('ik,kj->ij', A, B)
```

Column-wise

$$\begin{pmatrix} \cdot \\ \cdot \\ c_k \\ \cdot \\ \cdot \end{pmatrix} = A \begin{pmatrix} \cdot \\ \cdot \\ b_k \\ \cdot \\ \cdot \end{pmatrix}$$

Row-wise

$$\begin{pmatrix} \cdots c_k \cdots \end{pmatrix} = \begin{pmatrix} \cdots a_k \cdots \end{pmatrix} \; B$$

Outer-product

$$C = \sum_{i=1}^{n} \begin{pmatrix} \cdot \\ \cdot \\ a_i \\ \cdot \\ \cdot \end{pmatrix} \begin{pmatrix} \cdots b_i \cdots \end{pmatrix}$$

Block-wise

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

$$C = AB = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

Column picture: $Ax = b$

$$A\,x = b$$

$$x_1 \begin{pmatrix} \cdot \\ \cdot \\ a_1 \\ \cdot \\ \cdot \end{pmatrix} + x_2 \begin{pmatrix} \cdot \\ \cdot \\ a_2 \\ \cdot \\ \cdot \end{pmatrix} = b$$

Notes:

- Linear combination of columns

Row picture: $xA = B$

$$x \begin{pmatrix} \cdots r_1 \cdots \\ \cdots r_2 \cdots \\ \cdots r_3 \cdots \end{pmatrix} = \left(x_1\, r_1 + x_2\, r_2 + x_3\, r_3 \right)$$

Notes:

- Linear combination of rows

Elimination: $A \to U$

$$E_n \cdots E_2 E_1 A = U$$

$$\begin{aligned} A\,x &= b \\ E_n \cdots E_2 E_1 A\,x &= E_n \cdots E_2 E_1 b \\ U\,x &= E_n \cdots E_2 E_1 b \end{aligned}$$

Practice: Solve the system of equations:

$$\begin{aligned} x - y - z + u &= 0 \\ 2x + 2z &= 8 \\ -y - 2z &= -8 \\ 3x - 3y - 2z + 4u &= 7 \end{aligned}$$

Matrix inverse

Square matrix

$$A^{-1}\,A = I = A\,A^{-1}$$

$$\begin{aligned} A\,x &= b \\ x &= A^{-1}\,b \end{aligned}$$

$$(AB)^{-1} = B^{-1}\,A^{-1}$$

$$\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$$

Notes:

- $\det(A) = 0$ implies that A is not invertible or A is singular.
- For $x \neq 0,\ A\,x = 0$ implies that A is singular.
- The Gauss-Jordan elimination finds the inverse: $E(A|I) = (I|A^{-1})$.

Practice: Conditions for inverse of $\begin{pmatrix} a & b & b \\ a & a & b \\ a & a & a \end{pmatrix}$ and the inverse when it exists.

Matrix Factorization

$A = L\,U$

$$\begin{aligned} E\,A &= U \\ A &= E^{-1}\,U \end{aligned}$$

$$L = E^{-1}$$

Notes:

- For $A^{n \times n}$, $\mathcal{O}(n^3)$.
- U has row echelon (staircase) form. In reduced row echelon form, $\mathrm{rref}(A)$, the pivots are 1's and above and below each pivot there are 0's.

Practice: When does LU factorization exist and what are the factors for $\begin{pmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{pmatrix}$.

Permutation matrices

Permutation matrices do row exchanges. There are $n!$ permutation matrices and they form a group.

$$P^{-1} = P^T$$

With row exchanges,

$$\begin{aligned} PA &= LU \\ A &= P^T LU \end{aligned}$$

Symmetric matrix

$$A^T = A$$

Notes:

- $A^T A$ is square and symmetric.

Vector space

Set of vectors closed under: (1) vector addition (2) scalar multiplication. Example: $\mathbb{R}^2$. The set $\{u \in \mathbb{R}^2 : u_i > 0\}$ is not a vector space.

Vector subspace

A vector space within a larger vector space.

Notes:

- A line through the origin is a subspace in $\mathbb{R}^2$. These vectors have two components, so the line-subspace is not the same as $\mathbb{R}^1$.
- The set with the zero vector is a subspace in $\mathbb{R}^2$: $\left\{(0,0)^T\right\}$.
- If S and T are subspaces, $S \cap T$ is a subspace. $S \cup T$ may not be a subspace.

Practice: $x_1 = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$, $x_2 = \begin{pmatrix} 2 \\ 4 \\ 0 \end{pmatrix}$

1. Find V_1 = Subspace generated by x_1; V_2 = Subspace generated by x_2. Describe $V_1 \cap V_2$.
2. Find V_3 = Subspace generated by $\{x_1,\ x_2\}$. Is $V_3 = V_1 \cup V_2$? Find a subspace S of V_3 such that $x_1 \notin S$, $x_2 \notin S$.

3. What is $V_3 \cap \{x - y \text{ plane}\}$?

Practice: Which are subspaces of $\mathbb{R}^3 = \left\{ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\}$?

1. $b_1 + b_2 - b_3 = 0$
2. $b_1 b_2 - b_3 = 0$
3. $\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$
4. $\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Matrix subspaces

Columnspace

$C(A)$: Linear combinations of columns of A.
$Ax = b$ has a solution when $b \in C(A)$.

$$C\left(\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \right) = C\left(\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} \right)$$

The column space of $\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$ is a two-dimensional subspace in $\mathbb{R}^4$.

Nullspace

$N(A)$: All solutions of $Ax = 0$.
We can use elimination to compute $N(A)$. $EA = U$. The number of pivots in U is $\mathrm{rank}(A)$.

With $A = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix}$ and $U = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, we now solve $Ux = 0$.
Here x_1, x_3 are pivot variables and x_2 and x_4 are free variables. We can have one special solution for each free variable by setting that free variable to 1 and the rest of the free variables to 0. $N(A)$ would then be linear combinations of the special solutions. For $A^{m \times n}$, there are $n - \mathrm{rank}(A)$ free variables.

$\mathrm{rref}(A) = \begin{pmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ has the form $\begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$ with I as the pivots and F as the free's. Special solutions are the columns of $\begin{pmatrix} -F \\ I \end{pmatrix}$.

$N\left(\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}\right)$ is a subspace of $\mathbb{R}^3$.

If $b \neq 0$, the set of solutions of $Ax = b$ is not a subspace.

$\mathrm{rank}(A) = \mathrm{rank}(A^T)$.

If A's columns are independent, $N(A)$ is empty.

Rowspace: $C(A^T)$

The space spanned by the row vectors. Same as $C\left(A^T\right)$. $\mathrm{rank}(C(A)) = \mathrm{rank}(C(A^T))$, because each row can have at most 1 pivot, each column can have at most 1 pivot. Also, a row is a constraint. Each new constraint reduces the number of free variables by 1 and thus increases pivot count by 1.

Left nullspace: $N(A^T)$

For $A^{m \times n}$, $\mathrm{rank}(N(A^T)) = m - \mathrm{rank}(C(A))$.

Practice: Suppose

$$B = \begin{pmatrix} 1 & & \\ 2 & 1 & \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Find a basis for and compute the dimension of each of the 4 fundamental subspaces.

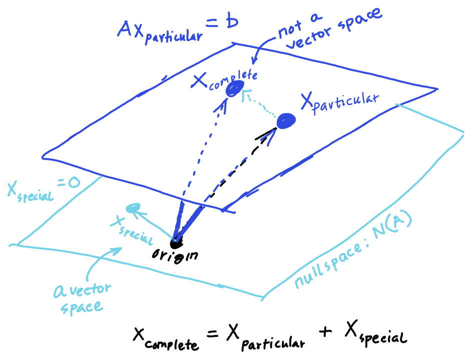
Solve $Ax = b$

Solvable only if $b \in \mathrm{col}(A)$. To solve, use Gaussian elimination.

$x_{\text{complete}} = x_{\text{particular}} + x_{\text{special}}$.

Say $Ax_1 = b$ and $Ax_2 = b$. Then, $A(x_2 - x_1) = 0$. So, difference between any two solutions, Δx, is in nullspace: $A\,\Delta x = 0$. Note, $x_2 = x_1 + \Delta x$.

We get $x_{\text{particular}}$ by setting all free variables to 0. We get one special nullspace solution per free variable. Linear combinations of x_{special}'s is the nullspace.



Practice: The set S of points $P(x, y, z)$ such that $x - 5y + 2z = 9$ is a ______ in $\mathbb{R}^3$. It is ______ to the ______ S_0 of $P(x, y, z)$ such that $x - 5y + 2z = 0$. All points of S have the form:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \\ 0 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} \\ 0 \\ 1 \end{pmatrix}$$

Rank

For $A^{m \times n}$, the rank r is the number of pivots during elimination. $r \leqslant m$, $r \leqslant n$.

Full column rank, $r = n < m$

0 or 1 solution.

No free variables. $N(A) = \{0\}$. Solution to $Ax = b$ exists only if $b \in C(A)$ and in that case it is unique.
$\text{rref}(A) = \begin{pmatrix} I \\ 0 \end{pmatrix}$

Full row rank, $r = m < n$

∞ solutions.

There are $(n - m)$ free variables. $Ax = b$ has $x_{\text{particular}}$ only if $b \in C(A)$. There are $(n - m)$ x_{special}'s whose linear combinations give $N(A)$.
$\text{rref}(A) = \begin{pmatrix} I & F \end{pmatrix}$

Full rank, $r = m = n$

1 solution

$r = m = n$. There is always a unique solution to $Ax = b$. $N(A) = \{0\}$.
$\text{rref}(A) = I$.

Not-full rank, $r < m, r < n$

0 or ∞ solutions.
$\mathrm{rref}(A) = \begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$

Practice: Find all solutions, depending on b_1, b_2, b_3:

$$\begin{aligned} x - 2y - 2z &= b_1 \\ 2x - 5y - 4z &= b_2 \\ 4x - 9y - 8z &= b_3 \end{aligned}$$

Linearly independent vectors

The columns are linearly independent vectors if $N(A) = \{0\}$.

Basis for a space is a set of linearly independent vectors that span the space. Given a space, every basis for the space has the same number of vectors and the number of vectors is the dimension of the space. $\mathrm{rank}(A)$ is the dimension of $C(A)$.

If $A^{n\times n}$ has linearly independent columns, A is invertible and the column vectors of A form a basis for $\mathbb{R}^n$. Dimension of $N(A) = n - \mathrm{rank}(A)$.
Practice: Find the dimension of the vector space spanned by the vectors:

$$\begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \\ -4 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \\ -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 0 \\ -2 \\ 0 \end{pmatrix}$$

And find a basis for that space.

Vector space of matrices

M = Set of all 3×3 matrices.

Practice: Show that the set of 2×3 matrices whose nullspace contains $\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ is a vector subspace and find a basis for it. What about the set of those whose column space contains $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$?

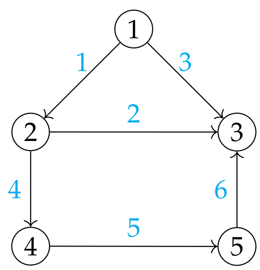
Graph

If A is the incidence matrix and x is potential at nodes. Then Ax is the potential differences across nodes. Say C is the conductance per edge and f is the external input and output.

$$\begin{aligned} e &= Ax \\ y &= Ce \\ A^T y &= f \end{aligned}$$

$$\boxed{A^T C A x = f}$$

Practice:



- Find incidence matrix A
- $N(A), N(A^T) =?$
- $\mathrm{Trace}(A^T A) =?$

Orthogonal vectors

The vectors x and y are orthogonal if $x^T y = 0$.

Orthogonal subspaces

The subspaces S and T are orthogonal if every vector in S is orthogonal to every vector in T.

For $A^{m \times n}$, $C(A) \perp N(A^T)$ and $C(A^T) \perp N(A)$. And, $\dim C(A) + \dim N(A^T) = m$, $\dim C(A^T) + \dim N(A) = n$.

Solve $Ax = b$ when no exact solution

For $A^{m \times n}$ where $m > n$ or where too many equations, we may not have a solution. Solve below instead:

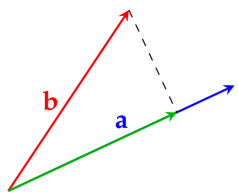
$$A^T A x = A^T b$$

$(A^T A)$ isn't always invertible. $N(A^T A) = N(A)$, $\mathrm{rank}(A^T A) = \mathrm{rank}(A)$. So, $A^T A$ is invertible if A has linearly independent columns.

Practice: S is spanned by $(1, 2, 2, 3)$ and $(1, 3, 3, 2)$.

1. Find a basis for $S^{\perp}$
2. Can every $v \in \mathbb{R}^4$ be written uniquely in terms of S and $S^{\perp}$?

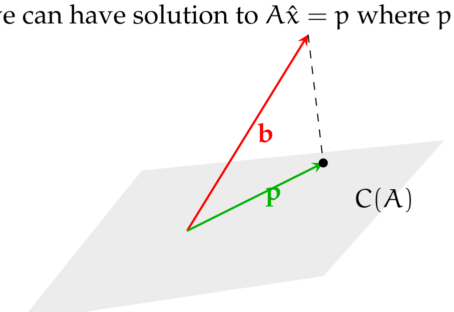
Projection



Projection of the vector b onto the vector a = $a\, \frac{a^T b}{a^T a} = Pb$.
Here the projection matrix $P = \frac{1}{a^T a}\, aa^T$.
P is symmetric. $C(P)$ is the line through a. $\mathrm{rank}(P) = 1$. $P^2 = P$.

Why project?

Ax is in $C(A)$ and when $b \notin C(A)$ we have no solution for $Ax = b$. However, we can have solution to $A\hat{x} = p$ where p is b's projection onto $C(A)$.



Here $(b - p) \perp C(A)$ or $A^T(b - p) = 0$.

$$A^T b = A^T p$$
$$A^T b = A^T A\hat{x}$$

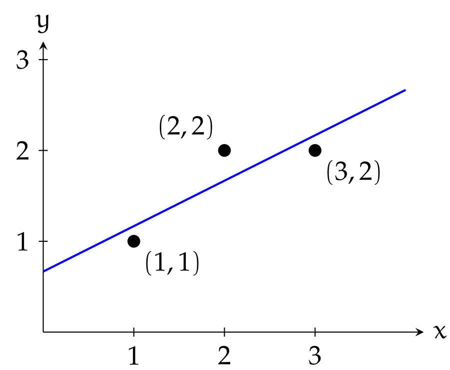
As long as A has linearly independent columns, $A^T A$ has inverse. In that case,

$$\hat{x} = \left(A^T A\right)^{-1} A^T b$$

$$p = A\hat{x} = A\, \left(A^T A\right)^{-1} A^T b$$

The projection matrix, $P = A\, (A^T A)^{-1} A^T$. P is symmetric. $P^2 = P$. On the other hand, $(I - P)$ projects b onto $N(A^T)$.

Least squares



$$\begin{aligned} c + d &= 1 \\ c + 2d &= 2 \\ c + 3d &= 2 \end{aligned}$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

$$\hat{x} = \begin{pmatrix} \hat{c} \\ \hat{d} \end{pmatrix}$$

Practice: Find the orthogonal projection matrix onto the plane $x + y - z = 0$.

Least squares minimizes $\|e\|^2 = \|Ax - b\|^2$

Investigate: If we are looking at the data plane (x-y), the error least squares is minimizing looks like vertical distance to the fitted line. If we are looking at C(A) and b, the error looks perpendicular to C(A). Geometrically, why?

Investigate: What if we minimized the perpendicular distance to the fitted line instead of vertical distance to the fitted line as is being done here in least squares?

Practice: Find the quadratic equation through the origin that is a best fit for the points: $(1, 1)$, $(2, 5)$, and $(-1, -2)$.

Orthogonal matrices

Orthogonal matrix Q has orthonormal columns. $Q^TQ = I$. If Q is square, $Q^T = Q^{-1}$. Projection matrix, $P_Q = QQ^T$.

Gram-Schmidt

A has linearly independent columns. Gram-Schimdt algorithm transforms A into an orthogonal matrix Q. The first column is fine. For the second column, to make it perpendicular to the first column, subtract its projection onto the first column – the remainder is our new second column. For the third column, subtract its projections onto the first and onto the new second columns. So on. In the end, normalize these new columns to build Q.

For $\begin{pmatrix} \vdots & \vdots & \vdots \\ a & b & c \\ \vdots & \vdots & \vdots \end{pmatrix}$ with a, b, c linearly independent.

$$\alpha = a$$

$$\beta = b - \frac{\alpha^T b}{\alpha^T \alpha}\alpha$$

$$\gamma = a - \frac{\alpha^T c}{\alpha^T \alpha}\alpha - \frac{\beta^T c}{\beta^T \beta}\beta$$

$\alpha \perp \beta \perp \gamma$.

$$Q = \begin{pmatrix} \vdots & \vdots & \vdots \\ q_1 & q_2 & q_3 \\ \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & \vdots \\ \frac{\alpha}{||\alpha||} & \frac{\beta}{||\beta||} & \frac{\gamma}{||\gamma||} \\ \vdots & \vdots & \vdots \end{pmatrix}$$

$A = QR$ where R is upper triangular.

Practice: Find q_1, q_2, q_3 (orthonormal) from a, b, c (columns of A). Then write A as QR.

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{pmatrix}$$

```python
A = np.array(
    [
        [1, 2, 4],
        [0, 0, 5],
        [0, 3, 6]
    ]
)
a, b, c = A[:, 0], A[:, 1], A[:, 2]

project = lambda b, a: (a.T @ b)/(a.T @ a) * a
alpha = a
beta = b - project(b, alpha)
gamma = c - project(c, alpha) - project(c, beta)
Q = np.vstack(
    [
        alpha/la.norm(alpha),
        beta/la.norm(beta),
        gamma/la.norm(gamma)
    ]
)
R = Q.T @ A
```

Determinants

$\det(A)$ is a number associated with a square matrix and useful in eigen values.

Fundamental properties

- $\det(I) = 1$
- Row-exchange flips sign of $\det(A)$
- Linear in each row:
  - $\det\left(\begin{pmatrix} t \cdot a & t \cdot b \\ c & d \end{pmatrix}\right) = t \cdot \det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)$
  - $\det\left(\begin{pmatrix} a + a' & b + b' \\ c & d \end{pmatrix}\right) = \det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) + \det\left(\begin{pmatrix} a' & b' \\ c & d \end{pmatrix}\right)$

Derived properties

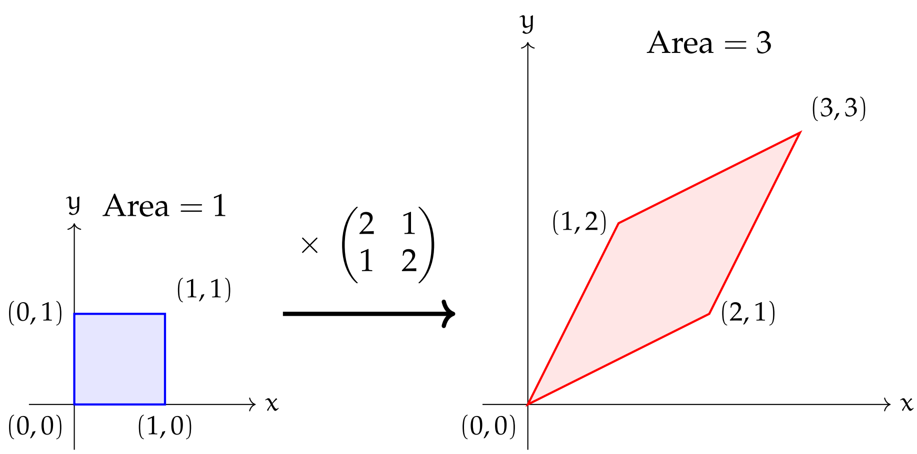
- Two identical rows $\implies \det(A) = 0$
- Elimination step: $\text{row}_j = \text{row}_j - c \cdot \text{row}_i$, does not change $\det(A)$
- Row of 0's $\implies \det(A) = 0$

- For upper triangular $U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix}$, $\det(U) = \prod_{i=1}^{n} u_{ii}$
- $\det(A) = 0$ when A is singular.
- $\det(AB) = \det(A) \cdot \det(B)$.
  - $\det(AA^{-1}) = \det(A) \cdot \det(A^{-1}) \implies \det(A^{-1}) = \frac{1}{\det(A)}$.
  - $\det(A^2) = (\det(A))^2$.
  - $\det(2A) = 2^n \det(A)$.
- $\det(A^T) = \det(A)$.
  - All row properties are now applicable for columns.

From the properties:

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \det\left(\begin{pmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{pmatrix}\right) = a \cdot \left(d - \frac{c}{a}b\right) = ad - bc$$

$\boxed{\text{det(A) is the volume of unit cube after multiplying by A}}$



$\det(A)$ gives a measure of distortion multiplying by A causes to the space.

Practice: Find the determinant of:

$$A = \begin{pmatrix} 101 & 201 & 301 \\ 102 & 202 & 302 \\ 103 & 203 & 303 \end{pmatrix}, B = \begin{pmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{pmatrix}, C = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & -4 & 5 \end{pmatrix}, D = \begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & 4 \\ -3 & -4 & 0 \end{pmatrix}$$

Iterative formula

$$\det(A) \;=\; \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Where,

- S_n is the symmetric group of all permutations of n elements of size $n!$.
- $$\operatorname{sgn}(\sigma) = \begin{cases} +1, & \text{if } \sigma \text{ is an even permutation,} \\ -1, & \text{if } \sigma \text{ is an odd permutation.} \end{cases}$$

For example:

$$\begin{aligned}
\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) &= \det\left(\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 & b \\ c & d \end{pmatrix}\right) \\
&= \det\left(\begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix}\right) + \det\left(\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}\right) \\
&= \det\left(\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\right) + \det\left(\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}\right) \quad \text{// single element in each row and column} \\
&= \det\left(\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\right) - \det\left(\begin{pmatrix} c & 0 \\ 0 & b \end{pmatrix}\right) \\
&= ad - bc
\end{aligned}$$

Recursive cofactor formula

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{j=1}^{n} a_{ij} (-1)^{i+j} \det(A_{-i,-j}).$$

$A_{-i,-j}$ is the $(n-1) \times (n-1)$ matrix without i-th row and j-th column of A.

Practice: Find the determinants of

$$A = \begin{pmatrix} x & y & 0 & 0 & 0 \\ 0 & x & y & 0 & 0 \\ 0 & 0 & x & y & 0 \\ 0 & 0 & 0 & x & y \\ y & 0 & 0 & 0 & x \end{pmatrix}, \; B = \begin{pmatrix} x & y & y & y & y \\ y & x & y & y & y \\ y & y & x & y & y \\ y & y & y & x & y \\ y & y & y & y & x \end{pmatrix}$$

Eigenvalues and Eigenvectors

Associated with square matrices. A typical vector x changes direction when A acts on it, therefore Ax points to a different direction than x. Eigenvectors are special. For eigenvectors, Ax is on the same straight line as x: $Ax = \lambda x$.

If A is singular, $Ax = 0$ for some $x \neq 0$. In that case, $\lambda = 0$ is an eigenvalue. For a projection matrix P there are two types of eigenvectors,

$$Px = \begin{cases} x, & x \text{ is on the plane}, \lambda = 1 \\ 0 & x \perp \text{plane}, \lambda = 0 \end{cases}$$

For permutation matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

$$A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \lambda = 1$$
$$A \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \lambda = -1$$

For general $A^{n \times n}$,

$$\sum_{i=0}^{n} \lambda_i = trace(A) = \sum_{i=1}^{n} a_{ii}$$

$$\begin{aligned} Ax &= \lambda x \\ (A - \lambda I)x &= 0, \quad // \ (A - \lambda I) \text{ is singular} \\ \det(A - \lambda I) &= 0, \quad // \text{ characteristic equation} \end{aligned}$$

$$(Ax = \lambda x) \implies (A + cI)x = (\lambda + c)x$$

The 90° rotation matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has complex eigenvalues: i and $-i$.
A symmetric matrix has real eigenvalues.
For a triangular matrix, the eigenvalues are the diagonal elements. For example, $\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$ has eigenvalues: 3 and 3. Since eigenvalue is repeated, there aren't two independent eigenvectors.

Practice: Given the invertible

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -2 \\ 0 & 1 & 4 \end{pmatrix}$$

Find the eigenvalues and eigenvectors of A^2 and $\left(A^{-1} - I\right)$.

Diagonalization

$A^{n \times n}$ has n independent eigenvectors and say they are the columns of S.

$$\begin{aligned} AS &= S\Lambda \\ S^{-1}AS &= \Lambda \qquad \texttt{//diagonalization} \\ A &= S\Lambda S^{-1} \qquad \texttt{//factorization} \end{aligned}$$

Here Λ is a diagonal matrix with eigenvalues in the diagonal.

$$A^k = S\Lambda^k S^{-1}$$

We have n independent eigenvectors:

- If all eigenvalues are distinct. In contrast, $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ has a single eigenvector.

Fibonacci sequence can be written in matrix form:

$$u_k = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} u_{k-1}$$

Here, $u_k = \begin{pmatrix} f_k \\ f_{k-1} \end{pmatrix}, u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

The eigenvalues of $F = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, are $\frac{1 \pm \sqrt{5}}{2}$. Since, F has distinct eigenvalues, it has two independent eigenvectors. F is diagonalizable. $F^k = S\Lambda^k S^{-1}$. So, the Fibonacci sequence is growing like $\left(\frac{1+\sqrt{5}}{2}\right)^k \approx 1.62^k$. The eigenvectors of F are: $\begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix}, \begin{pmatrix} \lambda_2 \\ 1 \end{pmatrix}$.

Practice: Find a formula for C^k where:

$$C = \begin{pmatrix} 2b - a & a - b \\ 2b - 2a & 2a - b \end{pmatrix}$$

Calculate C^{100} when $a = b = -1$.

Symmetric matrix

For $A = A^T$:

- Eigenvalues are real
- Eigenvectors are $\perp$
- $A = Q\Lambda Q^T = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T$. Here, $q_i q_i^T$ is a projection matrix. So, multiplication by symmetric A is like projecting onto the n mutually perpendicular eigenvectors, scaled by eigenvalues.
- Number of positive pivots = Number of positive eigenvalues.

Positive definite matrix

A positive definite matrix is symmetric with following properties:

- All eigenvalues are positive.
- All pivots are positive.
- All sub-determinants are positive.

Practice: Explain why each of the following is true:

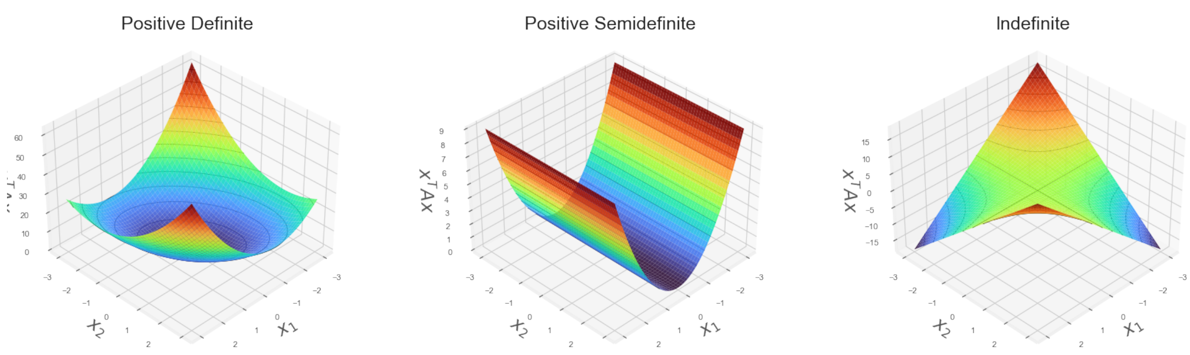
a) Every positive definite matrix is invertible.

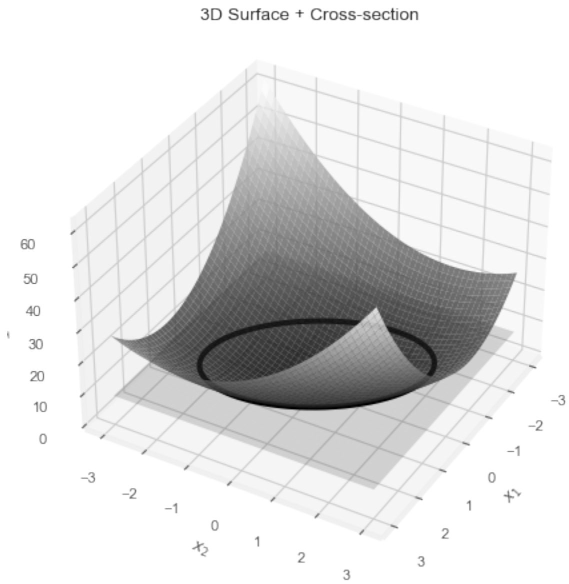
b) The only positive definite projection matrix is P=I.

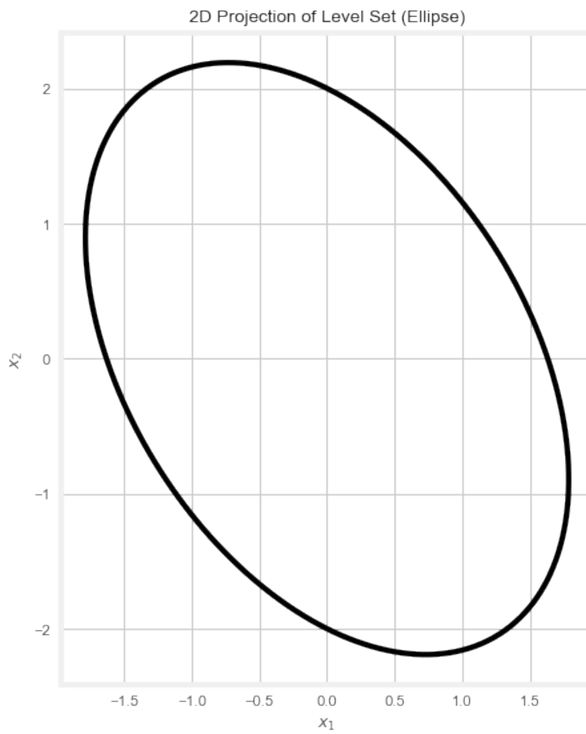
c) D is diagonal with positive entries is positive definite.

d) S is symmetric with $\det(S) > 0$ might not be positive definite.

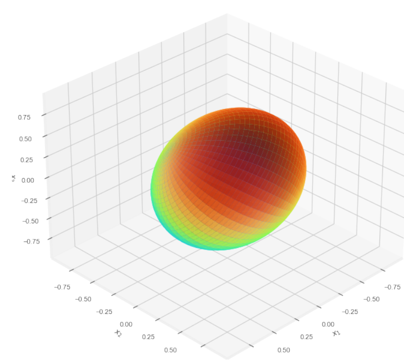
$\mathbf{x^T A x} > \mathbf{0} \implies$ positive definite.







For the positive definite matrix $\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, each level set $x^\top Ax = c(> 0)$ is an ellipsoid with 3 axes along the three perpendicular eigenvectors and the lengths of the axes are inversely proportional eigenvalues. For bigger eigenvalues, it grows faster in that direction, so for the points in the same level, those axes are shorter.



Ellipsoid defined by $x^\top Ax = 1$.

Practice: For which values of c is

$$B = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2+c \end{pmatrix}$$

- positive definite?
- positive semidefinite?

If A is (symmetric) positive definite, A^{-1} is also positive definite.
If A and B are positive definite, $A + B$ is positive definite.
For $A^{m\times n}$, $A^T A$ is symmetric. $x^T A^T Ax = (Ax)^T (Ax) >= 0$. It is zero, when $Ax = 0$. So, $A^T A$ is positive definite when A has independent columns.

Similar matrices

$A^{n\times n}$ and $B^{n\times n}$ are similar if for some (invertible) M, $B = M^{-1}AM$. For example, when A has n independent eigenvectors, $A = S\Lambda S^{-1}$. So, A and Λ are similar.

$$\boxed{\text{Similar matrices have same eigenvalues}}$$

Eigenvectors of B are $(M^{-1}x)$'s where x's are A's eigenvectors.

Jordan's theorem

Every $A^{n\times n}$ is similar to a Jordan matrix, J.

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_k \end{pmatrix}.$$

Where a $(m \times m)$ Jordan block:

$$J_m(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

Practice: Which of the following statements are true? Explain.

(a) If A and B are similar matrices, then $2A^3 + A - 3I$ and $2B^3 + B - 3I$ are similar.

(b) If A and B are 3 matrices with eigenvalues $1, 0, -1$, then A and B are similar.

(c) The matrices $J_1 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$ and $J_2 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ are similar.

Singular value decomposition, SVD

For any $A^{m \times n}$, we have $A = U\Sigma V^T$ where U and V are orthogonal and Σ is diagonal. This is a generalization of say $A = Q\Lambda Q^T$ for symmetric, positive definite A.
Here:

A is $m \times n$

V is $n \times n$

Σ is $m \times n$

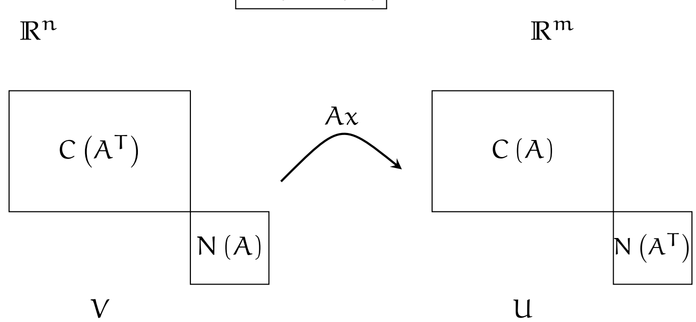
U is $m \times m$

$AV = U\Sigma$. We are diagonalizing A using two different bases: U (in column space) and V (in rowspace).

$A^TA = \left(V\Sigma^TU^T\right)U\Sigma V^T = V\Sigma^2V^T$, which is the eigen decomposition of A^TA. Now we solve $AV = U\Sigma$ to find U and thus SVD of A.

$v_1, v_2, \ldots, v_r$: orthonormal basis for rowspace of A.
$u_1, u_2, \ldots, u_r$: orthonormal basis for columspace of A.
$v_{r+1}, \ldots, v_n$: orthonormal basis for nullspace of A.
$u_{r+1}, \ldots, u_m$: orthonormal basis for nullspace of A^T

$$\boxed{Av_i = \sigma_i u_i}$$



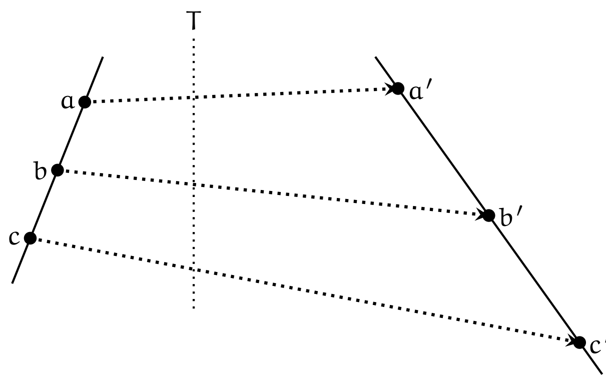
Practice: Find the singular value decomposition of the matrix

$$C = \begin{pmatrix} 5 & 5 \\ -1 & 7 \end{pmatrix}$$

Linear transformations

Properties:

- $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$
- $T(c\mathbf{v}) = cT(\mathbf{v})$



equal spacing preserved (not distance)

If for an n-dimensional space, we have a basis: $v_1, v_2, \ldots, v_n$ and we know $T(v_1), T(v_2), \ldots, T(v_n)$, we know $T(v)$ for any vector v.

Coordinates $(c_1, c_2, \ldots, c_n)$ come from a basis:

$$v = \sum_{i=1}^{n} c_i v_i$$

$A^{m\times n}\, x^{n\times 1} = y^{m\times 1}$ is like $T : \mathbb{R}^n \to \mathbb{R}^m$. We need an n-dimensional input basis and an m-dimensional output basis.

Rule to find A corresponding to T, given input basis $\{v_1, v_2, \ldots, v_n\}$ and output basis $\{w_1, w_2, \ldots, w_m\}$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Here:

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i$$

$$\boxed{A \times \text{input-coordinates} = \text{output-coordinates}}$$

For example, say $T = \frac{d}{dx}$.

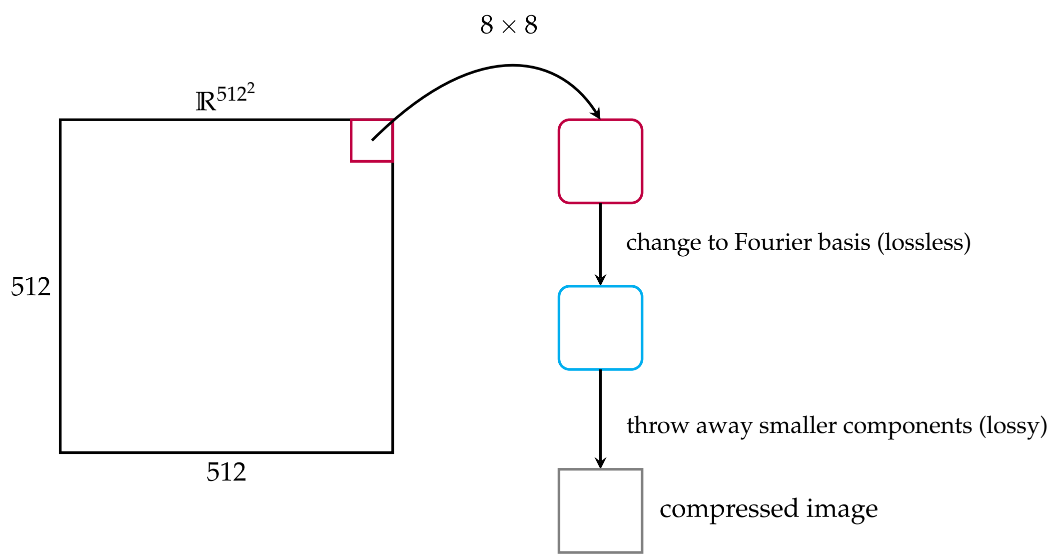
$$\begin{array}{lll} \text{Input: } c_1 + c_2 x + c_3 x^2 & & \text{basis: } 1, x, x^2 \\ \text{Output: } \quad c_2 + 2c_3 x & & \text{basis: } 1, x \end{array}$$

So, $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

Practice: Let $T(A) = A^T$, A is 2×2.

1) Why is T linear? What is T^{-1}?
2) Write down the matrix of T in
   - $V_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, V_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, V_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, V_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$
   - $W_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, W_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, W_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, W_4 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
3) Eigenvalues and eigenvectors of T?

Change of basis



Another good basis for image compression would be wavelets. The relevant algorithms: FFT and FWT. Eigenbasis would have been the best with a diagonal matrix, but expensive to find the eigenvectors.

If the linear transformation T has matrix A with respect to basis $(v_1, v_2, \ldots, v_n)$ and matrix B with respect to basis $u_1, u_2, \ldots, u_n$ then A and B are similar matrices. In other words, $B = M^{-1}AM$.

Practice: The vector space of all polynomials in x of degree $\leqslant 2$ has a basis $1, x, x^2$. Let w_1, w_2, w_3 be a different basis, of polynomials whose values at $x = -1, 0, 1$ are given by:

x	w_1	w_2	w_3
-1	1	0	0
0	0	1	0
1	0	0	1

a) Express $y(x) = -x + 5$ in this basis!

b) Find the change of basis matrices $(1, x, x^2) \rightleftarrows (w_1, w_2, w_3)$.

c) Find the matrix of "taking derivatives" in both bases!

Left-, right-, pseudo-inverse

When two-sided inverse does not exist.

Left inverse

For $A^{m \times n}$, full column rank, $\text{rank}(A) = n < m$, $N(A) = \{\mathbf{0}\}$, $Ax = b$ has 0 or 1 solution. The left inverse, A_l^{-1}, is $\left(A^T A\right)^{-1} A^T$.

$\boxed{AA_l^{-1} \text{ is the projection matrix for columnspace.}}$

Right inverse

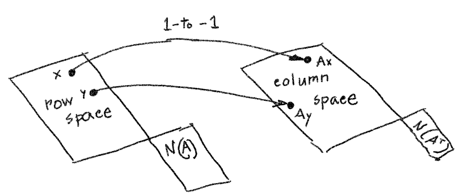
Full row rank, $\text{rank}(A) = m < n$, $N\left(A^T\right) = \{\mathbf{0}\}$, $n - m$ free variables, $Ax = b$ has ∞ number of solutions. Right inverse, A_r^{-1}, is $A^T \left(AA^T\right)^{-1}$.

$\boxed{A_r^{-1}A \text{ is the projection matrix for rowspace.}}$

Pseudo inverse

Always exists.

For $x, y \in C\left(A^T\right)$, $x \neq y \implies Ax \neq Ay$. For such x's, Ax's can be inversed. This is pseudo inverse: $A^+(Ax) = x$.



$$A = U\Sigma V^T$$
$$A^+ = V\Sigma^+ U^T$$

Here $\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & 0 \\ & \frac{1}{\sigma_2} & & & \\ & & \ddots & & \vdots \\ & & & \frac{1}{\sigma_r} & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix}$

Practice: Given $A = \begin{pmatrix} 1 & 2 \end{pmatrix}$

i) What is A^+?

ii) AA^+ and A^+A

iii) If $x \in N(A)$ what is A^+Ax?

iv) If $x \in C(A^T)$ what is A^+Ax?

Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x}_n = \bar{x}_{n-1} - \frac{1}{n}\left(x_n - \bar{x}_{n-1}\right)$$

Notes:

- May lie outside of the sample.
- Outliers affect the mean.
- Adding c to each data point adds c to the mean.
- Multiplying each data point by c, multiplies the mean by c.

Variance

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$s_n^2 = \frac{n-1}{n} s_{n-1}^2 + \frac{1}{n}\left(x_n - \bar{x}_{n-1}\right)\left(x_n - \bar{x}_n\right)$$

Notes:

- Average spread around the mean.
- Hard to interpret because in squared unit. So, the standard deviation, $\sqrt{s^2}$, is used.
- Adding c to each data point does not change the variance.
- Multiplying each data point by c, multiplies the variance by c^2.

Covariance Matrix

References

- Prof. Strang's lectures on Linear Algebra
  - Introduction to linear algebra
  - Applied numerical linear algebra
  - Linear algebra done right
  - The Matrix Cookbook
- Math for ML Specialization – Coursera
  - Mathematics for machine learning
  - The Matrix Calculus You Need For Deep Learning
  - Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach
  - An illustrative guide to multivariable and vector calculus
- Introduction to probability
  - Introduction to probability, statistics, and random processes
  - Introduction to mathematical statistics