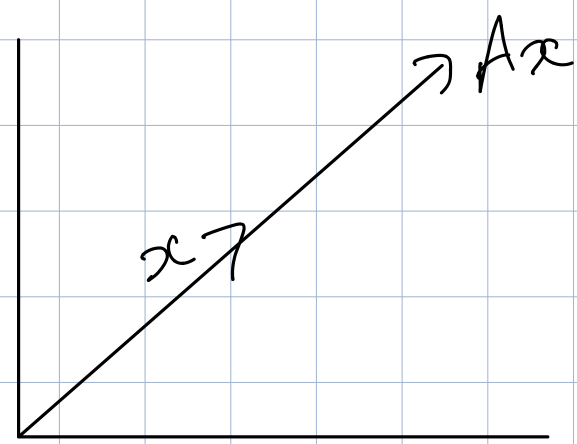
Eigenvalues and eigenvectors

$Ax = \lambda x$



Stretch x or shrink x but no change in direction

Case $\lambda = 0$: $Ax = 0$

$x \in N(A)$

Example : 1 Projection matrix P

Project onto a plane

$Px = x$ for x in plane.

$\lambda = 1$ is the eigenvalue & x is the eigenvector

$Px = 0 \quad \forall x$ perpendicular to the plane

$\lambda = 0$ and x is the eigenvector

(2) Permutation matrix

$$B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$Bx = x$ for $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$Bx = -x$ for $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

Finding the eigenvalues:

$$Ax = \lambda x$$

i.e., $(A - \lambda I)x = 0$

i.e., $(A - \lambda I)$ is singular

i.e., $\underbrace{\det(A-\lambda I)=0}$

↓

Characteristic polynomial of A which is of degree n

$(a_{11}-\lambda) \cdots\cdots (a_{nn}-\lambda)$

"n" roots = eigenvalues

For a given λ, $N(A-\lambda I)$ has to be found to obtain the eigenvectors.

Procedure to use for finding $N(A-\lambda I)$: Elimination.

Example:

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

$\det(A-\lambda I)=0$

$$\lambda^2 - 6\lambda + 8 = 0$$

$$\lambda_1 + \lambda_2 = 6, \quad \lambda_1 \lambda_2 = 8$$

Trace = sum of eigenvalues

Determinant = product —— " ——

Eigenvalues are $\lambda_1 = 4, \ \lambda_2 = 2$

$$A - 4I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$A - 2I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Letting $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$$B + 3I = A$$

$$Ax = (B+3I)x = \lambda x + 3x$$
$$= (\lambda+3)x$$

Eigen vecs of B = eigenvecs of A
(why?)

Notes:

1. If $Ax = \lambda_1 x$ and $Bx = \lambda_2 x$

then $\boxed{(A+B)x \overset{?}{=} (\lambda_1 + \lambda_2)x}$

does not always hold since x's need not be the same for A & B.

2. Symmetric matrices
↳ real eigenvalues

An example where eigenvalues aren't real.

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$ rotate by 90°

$$\det(A - \lambda I) = \lambda^2 + 1$$

$$\lambda_1 = i \quad , \quad \lambda_2 = -i$$

An example where we don't get two independent eigenvecs?

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

$$\lambda_1 = \lambda_2 = 3$$

$$(A - \lambda I) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is a eigenvec

& there is no other indep. eigenvec.

Similarity & diagonalization

Suppose A has n linearly independent eigenvectors $\{x_1 - - x_n\}$

$$S = \begin{bmatrix} | & & | \\ x_1 & - - - & x_n \\ | & & | \end{bmatrix}$$

Then, $$S^{-1} A S = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix},$$

and $\lambda_1 - - \lambda_n$ are the eigenvalues of A.

Proof:

$$AS = A\begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & & | \\ \lambda_1 x_1 & \cdots & \lambda_n x_n \\ | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\text{defined to be } \Lambda}$$

$$= S\Lambda$$

So, $AS = S\Lambda$

or, $S^{-1}AS = \Lambda$ [S invertible (why?)]

Claim:- If $\lambda_1, \cdots \lambda_n$ are distinct, then their eigenvectors $\{x_1, \cdots x_n\}$ are linearly independent.

Proof for $n=2$:

$$c_1 x_1 + c_2 x_2 = 0 \quad — \quad (1)$$

$$A(c_1 x_1 + c_2 x_2) = 0$$

$$c_1 \lambda_1 x_1 + c_2 \lambda_2 x_2 = 0 \quad — \quad (2)$$

$$c_1 (\lambda_1 - \lambda_2) x_1 = 0$$

$$\lambda_1 \neq \lambda_2, \quad x_1 \neq 0 \Rightarrow c_1 = 0$$

Similarly $c_2 = 0$

So, $\{x_1, x_2\}$ is a linearly indep. set

H.W:- Extend to the general case in "n" dimensions

Remarks:

(I) $A = S \Lambda S^{-1}$

S is not unique

x_1 is an eigenvec $\Rightarrow$ $c x_1$ is an eigenvec $\forall c \in \mathbb{R}$

So, each col. of S can be scaled to produce a new S.

(II) $A = S \Lambda S^{-1}$

Suppose col. 1. of S is y.

Col 1. of $S\Lambda = \lambda_1 y$

Col 1. of $AS = Ay$

Given $AS = S\Lambda$

$\Rightarrow Ay = \lambda_1 y \Rightarrow y$ is an eigenvec & λ_1 is an eigenval

III Powers of A.

Suppose λ is an eigenvalue & x an eigenvec of A.

$$A^2 x = A\lambda x = \lambda A x = \lambda^2 x$$

Suppose $S^{-1} A S = \Lambda$.

Is $S^{-1} A^2 S = \Lambda^2$?

$$(S^{-1} A S)(S^{-1} A S) = \Lambda^2$$

Works for a general k, i.e.,

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

IV Not all matrices are diagonalizable

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

Diagonalizability depends on "enough" eigenvecs

Invertibility — " — non-zero eigenvals

EXAMPLE

Fibonacci sequence

0, 1, 1, 2, 3, 5, - - -

$$F_{k+2} = F_{k+1} + F_k$$

What is F_{100}?

$$F_{k+2} = F_{k+1} + F_k$$

$$F_{k+1} = F_{k+1}$$

$$u_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}, \quad u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k$$

Start with u_0

$$u_0 \xrightarrow{A} u_1 \xrightarrow{A} u_2 \longrightarrow - - - -$$

$u_k = A^k u_0$ is a solution to

$$u_{k+1} = A u_k.$$

Suppose A has n independent eigenvecs

$$u_0 = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

$$A u_0 = c_1 \lambda_1 x_1 + c_2 \lambda_2 x_2 + \cdots + c_n \lambda_n x_n$$

$$u_k = A^k u_0 = c_1 \lambda_1^k x_1 + \cdots + c_n \lambda_n^k x_n$$

Back to Fibonacci!

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \quad p(\lambda) = \lambda^2 - \lambda - 1$$

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \quad \lambda_2 = \frac{1-\sqrt{5}}{2}$$

$$u_k = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2$$

$$u_0 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = c_1 x_1 + c_2 x_2$$

Need to find eigenvecs x_1, x_2 of A:

$$(A - \lambda I)(x) = \begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} x$$

$$x_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} \quad x_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$$

$$(A - \lambda_1 I)\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda_1^2 - \lambda_1 - 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$u_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \left(\frac{1}{\sqrt{5}}\right)\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} + \left(-\frac{1}{\sqrt{5}}\right)\begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$$

$$F_k = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k - \frac{1}{\sqrt{5}}\underbrace{\left(\frac{1-\sqrt{5}}{2}\right)^k}_{<1}$$

negligible contribution for large k

$$F_{100} \approx \frac{1}{\sqrt{5}}(1.618)^{100}$$

Spectral theorem

A is a real symmetric matrix

1. Eigenvalues of A are real
2. Eigenvecs corresponding to different eigenvalues are independent
3. A is orthogonally diagonalizable

i.e., $A = Q \Lambda Q^T$, where $Q^T Q = I$

$$A = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} - x_1^T - \\ \vdots \\ - x_n^T - \end{bmatrix}$$

Example:

$$A = \begin{bmatrix} 1 & -2 \\ -2 & -2 \end{bmatrix}$$

Eigenvalue $= -3, 2$

$$x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad x_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$q_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \qquad q_2 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

$$Q = \begin{bmatrix} | & | \\ q_1 & q_2 \\ | & | \end{bmatrix}$$ Check: $Q^T Q = I$

$$Q \Lambda Q^T = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

$$= A$$

Singular value decomposition

Every matrix cannot be diagonalized.

But, any "real" $m \times n$ matrix A

can be written as

$$A = Q_1 \Sigma Q_2^T,$$

where $Q_1 \to m \times m$ matrix

$Q_2 \to n \times n$ matrix

Q_1, Q_2 orthogonal

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$, where

$$D = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix}, \quad \sigma_i > 0$$

Why does this decomposition hold for any matrix A?

A is $m \times n$

$A^T A$ is $n \times n$

$A^T A$ is symmetric and real

$\Rightarrow \exists$ a basis of orthonormal eigenvectors, say $\{x_1, \dots, x_n\}$

corresponding to eigenvalues $\{\lambda_1 \dots \lambda_n\}$

$$A^T A x_i = \lambda_i x_i \,, \quad i = 1 \dots n$$

$$\|x_i\|^2 = 1 \,, \quad x_i \cdot x_j = 0 \,, \; i \neq j$$

So,

$$(A^T A x_i) \cdot x_i = (\lambda_i x_i) \cdot x_i$$
$$= \lambda_i$$

$$(A^T A x_i) \cdot x_i = (A^T A x_i)^T x_i$$
$$= x_i^T A^T A x_i$$
$$= (A x_i)^T A x_i$$
$$= \|A x_i\|^2$$
$$\geq 0$$

So, $\lambda_i \geq 0$

Next Stop: $\sigma_i = \sqrt{\lambda_i}$

<u>Order the λ's:</u>

$\lambda_1, \ldots \lambda_r, \lambda_{r+1}, \ldots \lambda_n$

$\lambda_1 > 0 \ldots \lambda_r > 0, \lambda_{r+1} = \ldots = \lambda_n = 0$

Let $\sigma_i = \sqrt{\lambda_i}$ and

$$y_i = \frac{1}{\sigma_i} A x_i, \quad i = 1 \ldots r$$

$(y_i \in \mathbb{R}^m)$

$$\|y_i\| = \frac{1}{\sigma_i} \|A x_i\| = \frac{\sqrt{\lambda_i}}{\sigma_i} = 1$$

$$y_i \cdot y_j = \frac{1}{\sigma_i \sigma_j} A x_i \cdot A x_j$$

$$= \frac{1}{\sigma_i \sigma_j} x_i^T A^T A x_j$$

$$= \frac{1}{\sigma_i \sigma_j} x_i^T \lambda_j x_j$$

$$= \frac{\lambda_j}{\sigma_i \sigma_j} x_i^T x_j = 0$$

So, $\{y_1 --- y_r\}$ is a orthonormal set of vectors

Extend this set to a basis & make it orthonormal

Let $\{y_1 \cdots y_m\}$ be that set.

$$Q_1 = \begin{bmatrix} | & & | \\ y_1 & \cdots & y_m \\ | & & | \end{bmatrix}$$

$$Q_2 = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$$

$$\Sigma = Q_1^T A Q_2$$

$$= \begin{bmatrix} - & y_1^T & - \\ & \vdots & \\ - & y_m^T & - \end{bmatrix} \begin{bmatrix} | & & | \\ Ax_1 & \cdots & Ax_n \\ | & & | \end{bmatrix}$$

$$(\Sigma)_{ij} = y_i^T (A x_j)$$

If $j \le r$, $y_j = \frac{1}{\sigma_j} A x_j$

$$y_i^T (A x_j) = y_i^T (\sigma_j y_j)$$

$$= \sigma_j \, y_i^T y_j$$

$$= \begin{cases} \sigma_j & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

If $j > r$, then (see next page)

$$\|Ax_j\|^2 = \lambda_j = 0$$

So, $Ax_j = 0$

$\Rightarrow \quad y_i^T A x_j = 0$

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 & 0 \\ & \ddots & & \\ 0 & & \sigma_r & \\ & 0 & & 0 \end{bmatrix}$$

$$\Sigma = Q_1^T A Q_2$$

$\Leftrightarrow \quad Q_1 \Sigma Q_2^T = A$ $\blacksquare$

Remark:

$$AA^T = Q_1 \Sigma \Sigma^T Q_1^T$$

So, eigenvectors of AA^T go into Q_1

and,

$$A^T A = Q_2 \Sigma^T \Sigma Q_2^T$$

implying eigenvectors of $A^T A$ go into Q_2.

EXAMPLE:

$$A = \begin{bmatrix} \sqrt{2} & 1 \\ 0 & \sqrt{2} \end{bmatrix}$$

Find SVD of A.

Is A diagonalizable?

No. A has $\sqrt{2}$ eigenvalue repeated twice.

$$A - \sqrt{2} I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

So, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector of A

& there aren't any more independent ones.

Finding the SVD

(see next page)

$$A^T A = \begin{bmatrix} 2 & \sqrt{2} \\ \sqrt{2} & 3 \end{bmatrix}$$

Eigenvalues: 4, 1

$$\sigma_1 = 2, \quad \sigma_2 = 1$$

$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Finding eigenvectors of $A^T A$:

$$A^T A - 4I = \begin{bmatrix} -2 & \sqrt{2} \\ \sqrt{2} & -1 \end{bmatrix}$$

So, $\begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}$ is an eigenvector of $A^T A$

$$x_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}$$

$$A^T A - I = \begin{bmatrix} \text{Fill} \\ \text{this} \end{bmatrix}$$

So, $\begin{bmatrix} \sqrt{2} \\ -1 \end{bmatrix}$ is another eigenvector of $A^T A$

$$x_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} \sqrt{2} \\ -1 \end{bmatrix}$$

So, $$Q_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} \sqrt{2} & 1 \\ 1 & -\sqrt{2} \end{bmatrix}$$

Using,

$$\sigma_1 y_1 = A x_1 \text{ \& } \sigma_2 y_2 = A x_2,$$

we obtain

$$y_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} \sqrt{2} \\ 1 \end{bmatrix}, \quad y_2 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ -\sqrt{2} \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} | & | \\ y_1 & y_2 \\ | & | \end{bmatrix}$$

Check, $A = Q_1 \Sigma Q_2^T$

Positive definiteness

Consider

$$F(x,y) = 7 + 2(x+y)^2 - y\sin y - x^3$$

$$f(x,y) = 2x^2 + 4xy + y^2$$

At a stationary point, the first derivatives vanish

$$\frac{\partial F}{\partial x} = 4(x+y) - 3x^2 = 0$$

$$\frac{\partial F}{\partial y} = 4(x+y) - y\cos y - \sin y = 0$$

$$\frac{\partial f}{\partial x} = 4x + 4y = 0$$

$$\frac{\partial f}{\partial y} = 4x + 2y = 0$$

So $(x,y) = (0,0)$ is a "stationary point" for both F and f.

Question: Whether $(0,0)$ is a minima / maxima / saddle?

Answer: Find the second derivatives at $(0,0)$

$$\frac{\partial^2 F}{\partial x^2} = 4 = \frac{\partial^2 f}{\partial x^2}$$

$$\frac{\partial^2 F}{\partial x \partial y} = 4 = \frac{\partial^2 f}{\partial x \partial y}$$

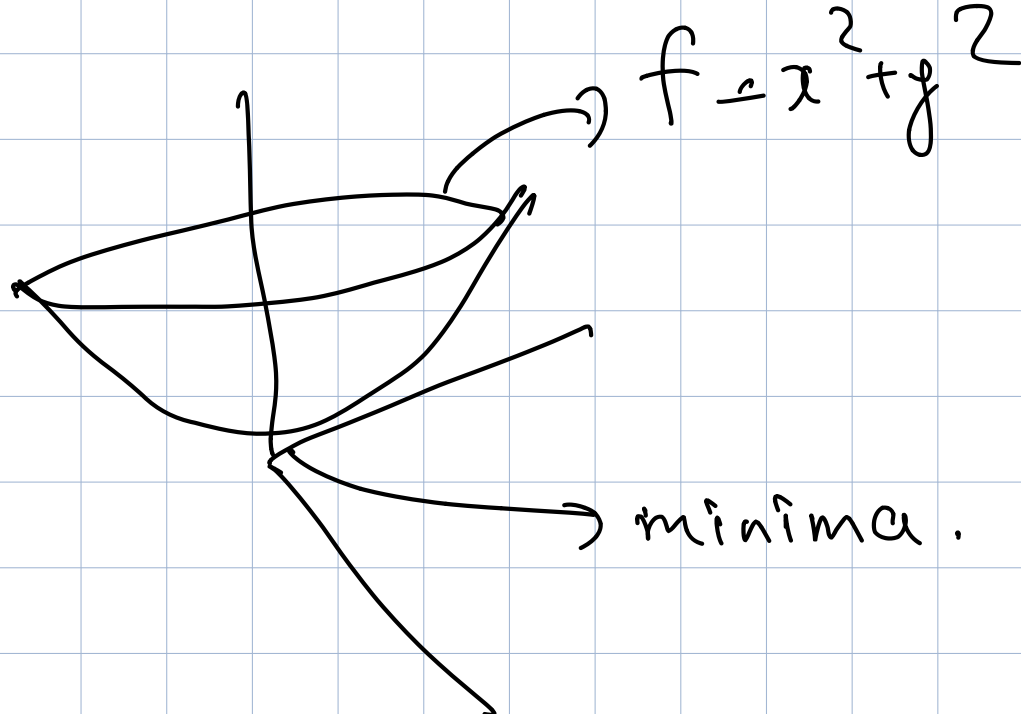
$$\frac{\partial^2 F}{\partial y^2} = 2 = \frac{\partial^2 f}{\partial y^2}$$

So, F and f behave identically near origin, i.e.,

F has a minimum iff f has a minimum

<u>Remark:</u>

Every quadratic $ax^2+2bxy+cy^2$ has a stationary point at $(0,0)$.



A function f that vanishes at $(0,0)$ & is strictly positive at other points is "positive definite".

Question: What conditions on a, b, c ensure $f = ax^2 + 2bxy + cy^2$ is p.d.?

Necessary conditions:

(I) If $f > 0$, then $a > 0$ [Look at $(1,0)$]

(II) If $f > 0$, then $c > 0$ [Look at f value at $(0,1)$]

But $a > 0$, $c > 0$ is not enough to ensure $f > 0$

e.g. $f = x^2 - 10xy + y^2$

Trick (Complete the square):

$$f = ax^2 + 2bxy + cy^2$$

$$= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2$$

(III) If $f > 0$, then

$$ac > b^2$$

So, $f = ax^2 + 2bxy + cy^2$ is p.d.

if and only if

$a > 0$ and $ac > b^2$

See next page

Any function $F(x,y)$ has a minimum at a point (x,y) where

$$\frac{\partial F}{\partial x} = 0 = \frac{\partial F}{\partial y}$$

$$\frac{\partial^2 F}{\partial x^2} > 0 \quad \text{and}$$

$$\left(\frac{\partial^2 F}{\partial x^2}\right)\left(\frac{\partial^2 F}{\partial y^2}\right) > \left(\frac{\partial^2 F}{\partial x \partial y}\right)^2$$

Note: Quadratic part of F:

$$\frac{x^2}{2}\frac{\partial^2 F}{\partial x^2}(\alpha,\beta) + xy\frac{\partial^2 F}{\partial x \partial y}(\alpha,\beta) + \frac{y^2}{2}\frac{\partial^2 F}{\partial y^2}(\alpha,\beta)$$

Remarks:-

① If $ac = b^2$, then

f is positive semidefinite if $a > 0$ &

negative semi-definite if $a < 0$

② Saddle point if $ac < b^2$

Connection to linear algebra:

$$ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$v = [x \ \ y]^T$$

$$ax^2 + 2bxy + cy^2 \quad \text{is} \quad v^T A v$$

where $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$

$v^T A v$ in $\mathbb{R}^n$

$$[x_1 \ - - \ x_n] \begin{bmatrix} a_{11} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

$v = (x_1 \cdots x_n)^T$

$f(v) = a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + a_{nn}x_n^2$

At $v = (0, \cdots 0)$, $f = 0$

So, $(0, \cdots 0)$ is a stationary point

Next question is to check if f has a minima/maxima/saddle at origin.

Examples:-

① $f = 2x^2 + 4xy + y^2$

$$A = \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix}$$

saddle at origin

(2) $f = 2xy$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$ Saddle

(3) A is 3×3

$$f = 2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2$$

$$f = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Check that f has a minimum at origin

Def: Matrix A is p.d.

if $v^T A v > 0 \quad \forall v \in \mathbb{R}^n$.

Test for positive-definiteness

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$ is p.d. if

$a > 0$, $ac - b^2 > 0$

$\Rightarrow$ both eigenvalues are > 0.

Fact:-

Each of the following tests are

both necessary & sufficient for p.d.

① $v^T A v > 0 \quad \forall v$

② All eigenvalues of A are > 0

③ All the pivots (without row exchanges) are > 0.

$$f(x,y) = ax^2 + 2bxy + cy^2$$

$$= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2$$

f is p.d if $f > 0$
$\forall\ (x,y) \neq (0,0)$ & 0 at $(0,0)$

f is p.d

$\Updownarrow$

$a > 0$ and $ac > b^2$

Recall the example
$$f(x,y) = 2x^2 + 4xy + y^2$$

This function has a saddle at origin because $ac < b^2$

brief remark on saddle:-

$f_1 = 2xy$ and $f_2 = x^2 - y^2$

$ac - b^2 = -1$ for both

Connection to linear algebra:-

$v^T A v$ in $\mathbb{R}^2$ $\quad ax^2 + 2bxy + cy^2$

$$= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$v^T A v$ in $\mathbb{R}^n$ $\quad \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$

"real-symmetric"
A matrix A is positive definite

if (i) $v^T A v > 0 \quad \forall v \in \mathbb{R}^n, v \neq 0$.

or

(ii) All eigenvalues of A are > 0

or

(iii) All upper left submatrices have positive determinant

or

(iv) All the pivots are > 0

Suppose (i) holds

$$Ax = \lambda x$$

$$x^T A x = x^T \lambda x = \lambda \|x\|^2$$

$$x \neq 0, \quad x^T A x > 0 \Rightarrow$$

$\lambda > 0$ i.e., (ii) holds.

Suppose (ii) holds.

Using spectral theorem, we obtain a orthonormal basis of eigenvectors.

Let $\{x_1 \dots x_n\}$ be that basis

Any x, $\quad x = c_1 x_1 + \cdots + c_n x_n$

$$Ax = c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n$$

$$x^T A x = (c_1 x_1^T + \cdots + c_n x_n^T)$$
$$\times (c_1 \lambda_1 x_1 + \cdots + c_n \lambda_n x_n)$$
$$= c_1^2 \lambda_1 + \cdots + c_n^2 \lambda_n$$
$$> 0 \text{ since } \lambda_i > 0 \ \forall i$$

So, (i) holds.

Brief notes on (iii) & (iv)

In 2×2 case, $\quad ac > b^2$

$$\Leftrightarrow \det(A) > 0$$

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

consider $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

$ac > b^2$ but

$A = -I$ not p.d.

infact, A is negative definite

Condition (iii) requires

$A_1 = [a_{11}]$ $A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$A_3 = \begin{bmatrix} a_{11} & & \\ & \ddots & \\ & & a_{33} \end{bmatrix}$ --- $A_n = A$

$\det(A_i) > 0 \quad i = 1 \dots n$

On condition (iv)

Consider $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

LDL^T

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

If $x = \begin{bmatrix} u \\ v \\ w \end{bmatrix}$, then

$$L^T x = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$= \begin{bmatrix} u - v/2 \\ v - 2w/3 \\ w \end{bmatrix}$$

$$x^T A x = x^T L D L^T x$$

$$= (L^T x)^T D \, L^T x$$

$$= 2\left(u - \frac{v}{2}\right)^2 + \frac{3}{2}\left(v - \frac{2w}{3}\right)^2$$

$$+ \frac{4}{3} w^2 > 0$$

So, in this example, (iv) $\Rightarrow$ (i)

Check:

$$ax^2 + 2bxy + cy^2$$

$$= a\left(x + \frac{b}{a}y\right)^2 + \frac{ac-b^2}{a}\,y^2$$

a, $\dfrac{ac-b^2}{a}$ are the pivots

for the 2×2 matrix A.

Principal Component Analysis

Feature Selection:

Start with as many features as you can collect, & then find a good subset of features.

Principal Component Analysis (PCA):

Idea: Project onto a lower dimensional space such that (i) reconstruction error is minimized

$\Updownarrow$

(ii) maximize the variance of projected data.

Given: $\{x_1, \ldots, x_n\}$, $x_i \in \mathbb{R}^d$

Goal: Project to a m-dimensional subspace
(m: input parameter)

Let $\mathcal{B} = \{u_1, \ldots, u_m\}$ be a orthonormal basis for a m-dimensional subspace

Extend $\mathcal{B}$ to a basis for $\mathbb{R}^d$: $u_1, \ldots u_m, u_{m+1}, \ldots, u_d$

Any vector $x \in \mathbb{R}^d$ can be written as

$$x = \sum_{j=1}^{d} \alpha_j u_j \,, \quad \text{where} \quad \alpha_j = x^T u_j$$

In particular, for $i = 1 \dots n$,

$$x_i = \sum_{j=1}^{d} (x_i^T u_j) u_j$$

Approximate x_i by $\tilde{x}_i$ as follows:

$$\tilde{x}_i = \sum_{j=1}^{m} z_{ij} u_j + \sum_{j=m+1}^{d} \beta_j u_j$$

Find z_{ij}, β_j to minimize

$$J = \frac{1}{n} \sum_{i=1}^{n} \| x_i - \tilde{x}_i \|^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\| \sum_{j=1}^{m} (x_i^T u_j - z_{ij}) u_j + \sum_{j=m+1}^{d} (x_i^T u_j - \beta_j) u_j \right\|^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{m} (x_i^T u_j - z_{ij})^2 + \sum_{j=m+1}^{d} (x_i^T u_j - \beta_j)^2 \right]$$

$$\frac{\partial J}{\partial z_{ij}} = 0 \quad \Rightarrow \quad 2(x_i^T u_j - z_{ij}) = 0 \Rightarrow z_{ij} = x_i^T u_j$$

$$\frac{\partial J}{\partial \beta_j} = 0 \quad \Rightarrow \quad \frac{1}{n} \sum_{i=1}^{n} (x_i^T u_j - \beta_j) = 0 \Rightarrow \beta_j = \left(\frac{1}{n} \sum x_i \right)^T u_j$$

$$= \bar{x}^T u_j$$

So,
$$\tilde{x}_i = \sum_{j=1}^{m} (x_i^T u_j)\, u_j + \sum_{j=m+1}^{d} (\bar{x}^T u_j)\, u_j$$

$$x_i - \tilde{x}_i = \sum_{j=m+1}^{d} (x_i^T u_j - \bar{x}^T u_j)\, u_j$$

$$\| x_i - \tilde{x}_i \|^2 = \sum_{j=m+1}^{d} \left((x_i - \bar{x})^T u_j \right)^2$$

$$J^* = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=m+1}^{d} \left((x_i - \bar{x})^T u_j \right)^2$$

$$= \frac{1}{n} \sum_{j=m+1}^{d} \sum_{i=1}^{n} \left((x_i - \bar{x})^T u_j \right)^T \left((x_i - \bar{x})^T u_j \right)$$

$$= \frac{1}{n} \sum_{j=m+1}^{d} \sum_{i=1}^{n} u_j^T (x_i - \bar{x}) (x_i - \bar{x})^T u_j$$

$$= \sum_{j=m+1}^{d} u_j^T \Bigg[\underbrace{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T}_{= C} \Bigg] u_j$$

$$J^* = \sum_{j=m+1}^{d} u_j^T C u_j$$

Consider a simpler case

$$\min_{u} u^T C u \quad \text{s.t.} \quad u^T u = 1$$

Lagrangian: $L(u,\lambda) = u^T C u + \lambda(1 - u^T u)$

$\nabla_u L(u,\lambda) = 0 \Rightarrow Cu = \lambda u$

So, $u^T C u = \lambda$

C is real-symmetric $\Rightarrow$
all eigenvalues are real &
$\exists$ a orthonormal basis of eigenvectors

To minimize $J^* = \sum_{j=m+1}^{d} u_j^T C u_j$,

Choose $u_{m+1}, \ldots, u_d$ to be $(d-m)$ eigenvectors corresponding to $(d-m)$ "least" eigenvalues.

$u_1, \ldots u_m$ correspond to the top-m eigenvalues of C.

PCA as maximizing variance:

Consider projection onto a line, given by a unit vector u_1.

For an x_i, the projection onto line along u_1 is
$(x_i^T u_1) u_1$

Mean is $(\bar{x}^T u_1) u_1$

Variance is $(x_i^T u_1 - \bar{x}^T u_1)^2$

Sum over all points to obtain

$$\frac{1}{n} \sum_{i=1}^{n} (x_i^T u_1 - \bar{x}^T u_1)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} u_1^T (x_i - \bar{x})(x_i - \bar{x})^T u_1$$

$$= u_1^T C u_1 \text{ , where } C = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$$

So, $\max_{u_1} u_1^T C u_1$ s.t. $u_1^T u_1 = 1$

is achieved by the eigenvector corresponding to the highest eigenvalue.

The logic can be extended to the case when $m > 1$.

For instance, we need u_1, u_2 s.t $\|u_1\| = 1, \|u_2\| = 1$ & $u_1^T u_2 = 0$

& projected variance is maximized

It can be shown that picking the eigenvectors corresponding to top-2 eigenvalues maximizes projected variance & so on.

In general, to perform PCA,

pick the top-m eigenvalues, find the corresponding eigenvectors $\{u_1 \dots u_m\}$

$u_i \to$ principal directions

& projected values $\to$ principal components.

PCA in higher dimensions

$$\{x_1 \dots x_n\} \qquad x_i \in \mathbb{R}^d$$

$$d \gg n$$

PCA requires calculating the eigenvectors of

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

Goal: Formulate the problem alternatively as finding the eigenvectors of a $n \times n$ matrix.

Notice that $\text{rank}(C) \leq n$,

which implies $(d-n)$ eigenvalues are zero.

Thus, it is not necessary to find $(d-n)$ eigenvectors

Let $A = \begin{bmatrix} (x_1-\bar{x})^T \\ \vdots \\ (x_n-\bar{x})^T \end{bmatrix}$ $\qquad C = \frac{1}{n} A^T A$

Let u_i be an eigenvector of C corresponding to eigenvalue $\lambda_i > 0$

n×n matrix

Claim: λ_i is an eigenvalue of $\frac{1}{n} A A^T$

<u>Proof:</u>

$$\lambda_i (A u_i) = A (\lambda_i u_i)$$
$$= A \left(\frac{1}{n} A^T A u_i\right)$$
$$\lambda_i (A u_i) = \frac{1}{n} A A^T (A u_i)$$

So, λ_i is an eigenvalue of $\frac{1}{n} A A^T$

$\blacksquare$

" It is enough to find eigenvectors of $\frac{1}{n} A A^T$ "

because ______________

Suppose v_i is an eigenvector of $\frac{1}{n} A A^T$. Then,

$$\frac{1}{n} A A^T v_i = \lambda_i v_i$$
$$\frac{1}{n} A^T A (A^T v_i) = \lambda_i (A^T v_i)$$

implying $(A^T v_i)$ is an eigenvector of $\frac{1}{n} A^T A = C$.

<u>PCA as whitening transform :-</u>

Features $\{x_1, \ldots x_n\}$ normalized & centered

i.e., zero mean $\bar{x}$ & unit variance.

Goal: Make the features uncorrelated.

Let $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_d$ be the eigenvalues of C

with corresponding eigenvectors $u_1, \ldots u_d$

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$$

$$\tilde{U} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{bmatrix}$$

Finding eigenvectors requires solving

$$C \tilde{U} = \tilde{U} \Lambda$$

Feature transformation:

$$z_i = \Lambda^{-1/2} \tilde{U}^T x_i$$

Covariance of $\{z_1, \ldots z_n\}$:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^T$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Lambda^{-1/2} \tilde{U}^T x_i x_i^T \tilde{U} \Lambda^{-1/2}$$

$$= \Lambda^{-1/2} \tilde{U}^T C \tilde{U} \Lambda^{-1/2}$$

$$= \Lambda^{-1/2} \tilde{U}^T \tilde{U} \Lambda \Lambda^{-1/2}$$

$$= I$$

■