* <u>Eigenvalues and Eigenvectors.</u>

<u>Def:</u> Consider an matrix $A \in R^{n\times n}$, A scalar λ is called an <u>eigenvalue</u> of A if there is a nonzero vector $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$. Such $\vec{v}$ is called an <u>eigenvector</u> ~~core~~ corresponding to the eigenvalue λ.

What does it mean geometrically?
if $\vec{v}$ is an ~~eigenvalu~~ eigenvector of A, the image of $\vec{v}$ under A is <u>parallel</u> to $\vec{v}$.

. Note that an eigenvector cannot be $\vec{0}$, but an eigenvalue can be 0.
. Suppose 0 is an eigenvalue of A. What does that say about A? $\Rightarrow$ there must be some nontrivial vector $\vec{v}$ for which

$$A\vec{v} = 0\vec{v} = 0.$$

$\Rightarrow$ A is not invertible.

<u>Thm:</u> If $\vec{v}_1, \dots, \vec{v}_r$ are eigenvectors that correspond to <u>distinct</u> eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n\times n$ matrix A, then the set $\{\vec{v}_1, \dots, \vec{v}_r\}$ is linearly independent.

E.g. What are eigenvalues of a projector P?
Suppose λ is an eigenvalue of P and $\vec{v}$ is its corresponding eigenvector. Then

$$
\left.\begin{aligned}
& Pv = \lambda v \\
\Rightarrow \quad & PPv = \lambda Pv. \\
& P^2 v = \lambda^2 v. \\
\text{Since } P^2 = P, \quad & P^2 v = Pv = \lambda v
\end{aligned}\right\}.
$$

$$
\begin{aligned}
\Rightarrow \quad & \lambda v = \lambda^2 v. \\
& (\lambda^2 - \lambda) v = 0. \\
\text{but } v \neq \vec{0} \quad \Rightarrow \quad & \lambda^2 - \lambda = 0. \\
\Rightarrow \quad & \lambda = 1 \text{ or } \lambda = 0.
\end{aligned}
$$

Thm: All eigenvalues (all roots of the characteristic polynomials) of a symmetric matrix are real.

Thm: Eigenvectors of a symmetric matrix corresponding to different eigenvalues are orthogonal.

Pf. Let $A^T = A$. have eigenvectors v_1 and v_2 corresponding to eigenvalues $\lambda_1 \neq \lambda_2$.

Then

$$\langle A\vec{v}_1, \vec{v}_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle.$$

on the other hand,

$$
\begin{aligned}
\langle Av_1, v_2 \rangle = (Av_1)^T v_2 &= v_1^T A^T v_2 \\
&= v_1^T A v_2 \\
&= v_1^T \lambda_2 v_2 \\
&= \lambda_2 \langle v_1, v_2 \rangle.
\end{aligned}
$$

$$\Rightarrow \lambda_1 \langle v_1, v_2 \rangle = \langle A\vec{v}_1, v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle$$

Since $\lambda_1 \neq \lambda_2$, $\langle v_1, v_2 \rangle = 0$.

- <u>Symmetric positive definite matrices.</u>

Suppose λ is an eigenvalue of A and v its corresponding eigenvector.

$$\Rightarrow \quad Av = \lambda v.$$

$$v^T A v = v^T (\lambda v)$$

$$= \lambda \|v\|^2.$$

$$\Rightarrow \quad \lambda = \frac{v^T A v}{\|v\|^2}.$$

$\therefore$ A symmetric matrix is positive definite if and only if all of its eigenvalues are positive.

- Eigenvectors of symmetric p.d. matrices are orthogonal.

* Review of multivariable calculus.

<u>Def:</u> $f : \mathbb{R}^n \to \mathbb{R}$, then the derivative $\frac{df}{d\vec{x}}$ is a row vector

$$\frac{df}{d\vec{x}} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

if $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

<u>E.g.</u> 1) $f(\vec{x}) = \|x\|_2^2 = x_1^2 + x_2^2 + \dots + x_n^2$

Then

$$\frac{df}{dx} = (2x_1, 2x_2, \dots, 2x_n) = 2x^T.$$

~~2) If C is a kxn matrix, then~~

$$\cancel{\frac{d(C\vec{x})}{d\vec{x}} = C^T}$$

2) The derivative of a scalar product c^Tx where $\vec{c}$ is an n-dimensional vector, $\vec{c} \in \mathbb{R}^n$, is equal to $\vec{c}$:

$$\frac{d}{dx}(c^Tx) = c^T$$

3) Let $g(\vec{x}) = \|a - x\|^2$.

Then

$$\frac{dg}{dx} = -2(a-x)^T.$$

<u>Pf:</u> $g(\vec{x}) = \sum_{i=1}^{n} (a_i - x_i)^2$

$$\frac{\partial g(\vec{x})}{\partial x_i} = -2(a_i - x_i)$$

$$\Rightarrow \cancel{\frac{\partial g}{\partial x_i}}$$

$$\Rightarrow \quad \frac{dg}{dx} = \left(\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n}\right)$$

$$= \left(-2(a_1 - x_1), \ldots, -2(a_n - x_n)\right).$$

$$= -2\left(a_1 - x_1, \ldots, a_n - x_n\right)$$

$$= -2(a - x)^T.$$

· **Def:** $f : \mathbb{R}^n \to \mathbb{R}^m$, then

$$\frac{\partial f}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

E.g. $f(x) = x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_2 \\ 2x_1 + 4x_2 \end{bmatrix}$.

Find $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$, and $\frac{\partial f}{\partial \vec{x}}$.

$$\frac{\partial f}{\partial \vec{x}} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \Rightarrow \frac{\partial f}{\partial x_1} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{and} \quad \frac{\partial f}{\partial x_2} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

* **Principal Component Analysis: (PCA).**

Data set $X = \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^m$.
(centered at 0, i.e, $E[X] = 0$).
Let ~~$\{q_1\}$~~ $\{b_1, \dots, b_m\}$ be an O.N.B. of $\mathbb{R}^m$

Facts: 1) $x_i = \sum_{j=1}^{m} \alpha_{ji} b_j$.

2) $\alpha_{ji} = \langle x_i, b_j \rangle = x_i^T b_j$ (why?).

3) Consider a subspace V spanned by $\{b_1, \dots, b_M\}$.
Let $Q = [b_1 \dots b_M]$.

$\Rightarrow$ orth. $\text{proj}_V x = \hat{x} = QQ^T x$.

~~• PCA: Consider a subspace, says, $V = \text{span}\{b_1, \dots, b_m\}$~~

• PCA: Find a subspace $V = \text{span}\{b_1, \dots, b_n\}$ which minimizes the cost:

$$J = \frac{1}{n} \sum_{i=1}^{n} \| x_i - \underbrace{\sum_{j=1}^{M} \alpha_{ji} b_j}_{\tilde{x}_i} \|_2^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \| x_i - \tilde{x}_i \|_2^2$$

$$\frac{\partial J}{\partial(\beta_{ji}, b_j)} = \frac{\partial J}{\partial \tilde{x}_i} \frac{\partial \tilde{x}_i}{\partial(\beta_{ji}, b_i)}$$

Fact: $\frac{\partial J}{\partial \tilde{x}_i} = -\frac{2}{n}(x_i - \tilde{x}_i)^T$.

chain rule: $\frac{\partial J}{\partial(\beta_{ji})} = \frac{\partial J}{\partial \tilde{x}_i} \frac{\partial \tilde{x}_i}{\partial \beta_{ji}}$

Note that: $\frac{\partial \tilde{x}_i}{\partial \beta_{ji}} = b_j \quad , \quad j = 1, \ldots, M.$

$$\Rightarrow \frac{\partial J}{\partial \beta_{ji}} = -\frac{2}{n}(x_i - \tilde{x}_i)^T b_j.$$

$$= -\frac{2}{n}\left(x_i - \sum_{l=1}^{M} \beta_{li} b_l\right)^T b_j$$

$$\overset{\text{O.N.B.}}{=} -\frac{2}{n}\left(x_i^T b_j - \beta_{ji} \underbrace{b_j^T b_j}_{=1}\right).$$

$$= -\frac{2}{n}\left(x_i^T b_j - \beta_{ji}\right)$$

$$= 0.$$

$$\Leftrightarrow \quad \beta_{ji} = x_i^T b_j = \langle b_j, x_i \rangle.$$

That is, $\tilde{x}_i$ is the orthogonal projection of x_i onto $\text{span}\{b_1, \ldots, b_M\}$.

$$\tilde{x}_i = \Big(\sum_{j=1}^{M} b_j b_j^T\Big) x_i .$$

$$\Rightarrow \quad x_i - \tilde{x}_i = \Big(\sum_{j=M+1}^{m} b_j b_j^T\Big) x_i = \sum_{j=M+1}^{m} \langle b_j, x_i \rangle b_j$$

Rewrite the cost function

$$J = \frac{1}{n} \sum_{i=1}^{n} \|x_i - \tilde{x}_i\|_2^2 .$$

$$= \frac{1}{n} \sum_{i=1}^{n} \Big\| \sum_{j=M+1}^{m} \langle b_j, x_i \rangle b_j \Big\|_2^2 .$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=M+1}^{m} \langle b_j, x_i \rangle^2 .$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=M+1}^{m} b_j^T x_i x_i^T b_j$$

$$= \cancel{\frac{1}{n}} \sum_{j=M+1}^{m} b_j^T \underbrace{\Big(\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T\Big)} b_j .$$

$$= \sum_{j=M+1}^{m} b_j^T \; S \; b_j .$$

$$= \mathrm{trace}\Big(\underbrace{\Big(\sum_{j=M+1}^{m} b_j b_j^T\Big)}_{\text{projection matrix}} S\Big).$$

$$\Rightarrow \boxed{J = \sum_{j=M+1}^{m} b_j^T S b_j .}$$

data covariance matrix

First consider $m=2 \Rightarrow b_1$ and b_2.

Suppose $J = b_2^T S b_2$.

$\Rightarrow$ $\min J = b_2^T S b_2$ subject to $b_2^T b_2 = 1$.

$$L = b_2^T S b_2 + \lambda(1 - b_2^T b_2).$$

$$\frac{\partial L}{\partial \lambda} = 1 - b_2^T b_2 = 0 \Leftrightarrow b_2^T b_2 = 1.$$

$$\frac{\partial L}{\partial b_2} = 2 b_2^T S - 2\lambda b_2^T = 0 \Leftrightarrow S b_2 = \lambda b_2.$$

$$\Rightarrow J = b_2^T S b_2 = b_2^T (\lambda b_2) = \lambda.$$

$\Rightarrow$ J is minimized if λ is the smallest eigenvalue corresponding to to the eigenvector b_2.

$\Rightarrow$ we want to project our data points onto the space spanned by the eigenvector corresponding to the largest eigenvalue.

$\Rightarrow$ For general case, b_j for $j = M+1, \dots, m$.

J is minimized if b_j are eigenvectors of S corresponding to the $m-M-1$ smallest eigenvalues of S

$[\, S b_j = \lambda_j b_j,\ j = M+1, \dots, m \Rightarrow J = \sum_{j=M+1}^{m} \lambda_j \,]$

$\Rightarrow$ ~~the~~ We need to project our data points onto the principal subspace spanned by ~~the M M larg~~ the M eigenvectors corresponding to the M largest eigenvalues.

* Key steps of PCA algorithm:

standardize the data:

1) Compute the mean μ of the data set of X
2) Mean subtraction: $\tilde{x}_i = x_i - \mu$.
3) Divide the data by its standard deviation.

$$\tilde{x}_i(j) = \frac{x_{ij} - \mu}{\sigma(X^{(j)})}$$

$$\Rightarrow \quad \bar{X}$$

4) Compute the eigenvectors and eigenvalues of the data covariance matrix

$$S = \frac{1}{n} \bar{X} \bar{X}^T$$

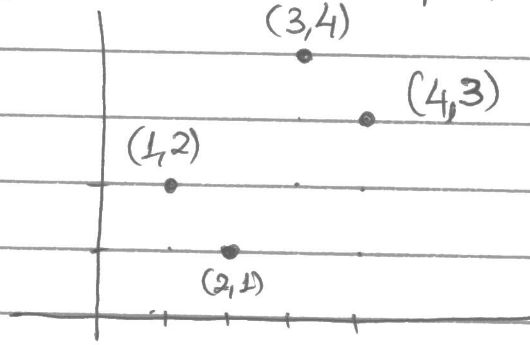
5) Chooses the eigenvectors associated with the M largest eigenvalues to be the basis of the principal subspace.

6) Collect these eigenvectors in a matrix

$$B = [b_1 \dots b_M]$$

7) Orthogonal projection of the data onto the principal axis using the projection matrix BB^T.

Example: (See textbook p. 413).



$$X = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix}.$$

- Compute XX^T:

$$XX^T = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 30 & 28 \\ 28 & 30 \end{bmatrix}.$$

- Find eigenvalues:

$$\det \begin{bmatrix} 30-\lambda & 28 \\ 28 & 30-\lambda \end{bmatrix} = 0.$$

$$(30-\lambda)(30-\lambda) - 28^2 = 0.$$

$$(58-\lambda)(2-\lambda) = 0.$$

$\Rightarrow$ $\lambda = 58$ and $\lambda = 2$.
are eigenvalues

- Find eigenvectors:
  - For $\lambda = 58$:

$$\begin{bmatrix} 30-58 & 28 \\ 28 & 30-58 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0.$$

$$\Rightarrow \begin{bmatrix} -28 & 28 \\ 28 & -28 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \text{ is an eigenvector}$$

For $\lambda = 2$:

$$\begin{bmatrix} 30-2 & 28 \\ 28 & 30-2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0.$$

$$\begin{bmatrix} 28 & 28 \\ 28 & 28 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

$\Rightarrow \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ is an eigenvector.

$\rightarrow$ Matrix of eigenvectors:

$$B = \begin{array}{cc} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \\ \begin{array}{cc} \uparrow & \uparrow \\ b_1 & b_2 \end{array} \end{array}$$

These vectors are orthonormal.

Consider

$$X^T B = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 3/\sqrt{2} & 1/\sqrt{2} \\ 3/\sqrt{2} & -1/\sqrt{2} \\ 7/\sqrt{2} & 1/\sqrt{2} \\ 7/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}.$$