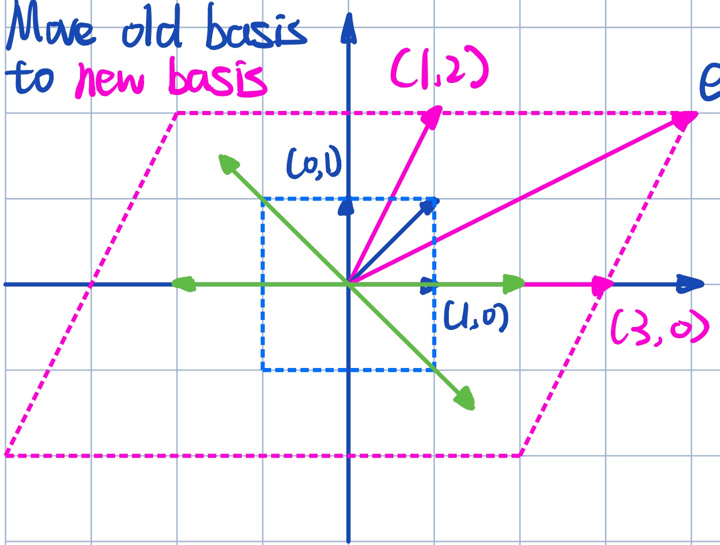
1. Eigenvalue decomposition of square matrices M

$\begin{cases} T(\vec{a}+\vec{b}) = T(\vec{a}) + T(\vec{b}) \\ T(c\vec{a}) = cT(\vec{a}) \end{cases}$

$M \in \mathbb{R}^{N\times N}$ represent a linear transformation to vector $v \in \mathbb{R}^N$

Mx: Apply transformation to v. $\mathbb{R}^N \to \mathbb{R}^N$

Move old basis to new basis



e.g. $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \times \overset{x}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$

If these new basis are linear dependent

→ After transformation, the dimension of space is reduced.

→ $\det(M)=0$, $\mathrm{rank}(M) < N$

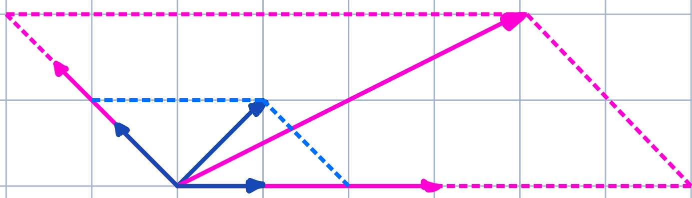
Within such transformations, there could be certain directions (up to rank(M)) that are only scaled:

direction $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is only scaled by 3

direction $\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$ is only scaled by 2

$Mv = \lambda v$.

If the coordinate is changed so that the eigenvector is exactly the basis to represent transformation M:



$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \times \overset{x}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

basis is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \times \overset{x}{\begin{bmatrix} 2 \\ \sqrt{2} \end{bmatrix}} = \begin{bmatrix} 6 \\ 2\sqrt{2} \end{bmatrix}$$

basis is $\begin{array}{cc} V_1 & V_2 \end{array}$ $\begin{bmatrix} 1 & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} \end{bmatrix}$

This can be expressed as

$$\overset{V}{\begin{bmatrix} 1 & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} \end{bmatrix}} \times \overset{\Lambda}{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}} \times \overset{V^{-1}}{\begin{bmatrix} 1 & -\frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} \end{bmatrix}^{-1}} \times \overset{x}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

- Change back to original basis ← Scaling along new basis ← Change to new basis (which are the eigenvectors)

  By calculating the projection on these basis

- Inverse transform of A. ← Scaling along original basis ← Transformation A.

$\hookrightarrow M = V \Lambda V^{-1}$

If M is symmetric:

- M represent scaling along perpendicular directions
- Eigenvectors of M are orthogonal to each other
- Transformation A is pure rotation by letting V contains orthonormal basis vectors. This yields $V^{-1} = V^T$
- M can be expressed as $V \Lambda V^T$

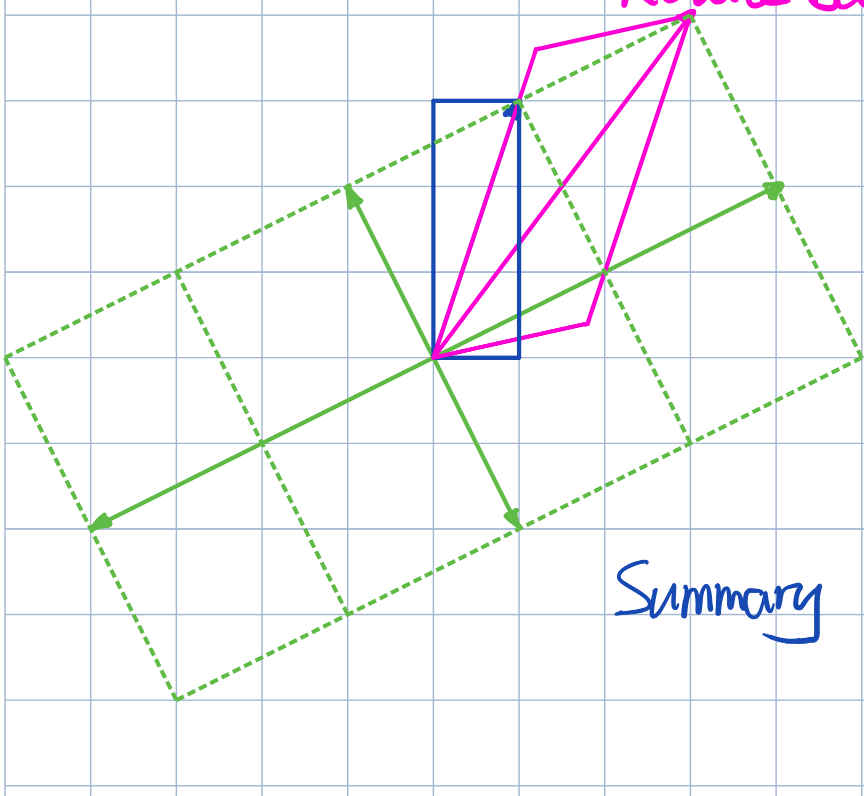
E.g. $M = \begin{bmatrix} \frac{9}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{6}{5} \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & \\ & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}^{-1}$

— Change back | Scale along new basis | Chage basis

$= \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 2 & \\ & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}^{-1}$

— Rotate back | Scale along old basis | Rotate

$= \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$



Summary
- rely on eigenvalues and eigenvectors : $Mx = \lambda x$
- M must be $N \times N$

2. Singular value decomposition of non-square matrices

$$\underset{M}{\begin{bmatrix} 4 & 3 & 1 \\ 5 & 0 & 2 \end{bmatrix}} \times \underset{x}{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} = \begin{bmatrix} 8 \\ 7 \end{bmatrix}$$

Linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$:

move $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$

move $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ to $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$

move $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ to $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

- For $M \in \mathbb{R}^{P \times N}$ with rank r
  - $M v = \sigma u$

$v \in \mathbb{R}^N$ — σ: Singular value — $u \in \mathbb{R}^P$

  - $M = U S V^T$

V: orthonormal basis in $\mathbb{R}^N$ $\quad V^{-1} = V^T$

U: orthonormal basis in $\mathbb{R}^P$

$$\underset{P \times N}{M} \; \underset{N \times 1}{x} = \underset{P \times P}{U} \quad \underset{P \times N}{S} \quad \underset{N \times N}{V^T} \quad \underset{N \times 1}{x}$$

$$S = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$$

Change basis in $\mathbb{R}^P$ ← Scale along new basis add/discard dimension ← Change basis in $\mathbb{R}^N$

Rotation in $\mathbb{R}^P$ ← Scale along old basis add/discard dimension ← Rotation in $\mathbb{R}^N$

3. Affine transformation of multivariative normal distribution — Principle components analysis

$x \in \mathbb{R}^N \quad x \sim N(\mu, \Sigma)$

$\rightarrow p(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$

Here suppose that x is centered: $\mu = 0$

Σ: $N \times N$ Covariance matrix

$$\begin{array}{c|ccc} & 1 & 2 \cdots & N \\ \hline 1 & Var(x_1) & Cov(x_1, x_2) & Cov(x_1, x_N) \\ 2 & Cov(x_2, x_1) & Var(x_2) & Cov(x_2, x_N) \\ \vdots & & & \\ N & Cov(x_N, x_1) & Cov(x_N, x_2) & Var(x_N) \end{array} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$\rightarrow Ax + b \sim N(A\mu + b, A\Sigma A^T)$

$\rightarrow$ In our case, $Ax \sim N(0, A\Sigma A^T)$

$\rightarrow$ By selecting A, It is possible to make $A\Sigma A^T$ diagnoal.

Such operation decorrelate x:

$Cov(x_i, x_j) = 0$ for $i \neq j$

$\rightarrow$ Achieved by eigenvalue decomposition to Σ

$\Sigma = V \Lambda V^T$, set $A = V^T$

$\rightarrow A\Sigma A^T = V^T V \Lambda V^T V = \Lambda$

Λ is the diagnoal matrix of eigenvalues of Σ.

Each eigenvalue represent variance of each feature in Ax

- Σ is symmetric and positive definite
  - A represent pure rotations / A contains orthonormal basis
  - All eigenvalues of $\Sigma > 0 \rightarrow$ matches variance
- Additional properties
  - Ax is the projection of x on directions of orthonormal basis in V ($Ax = V^T x = V^{-1}x$).

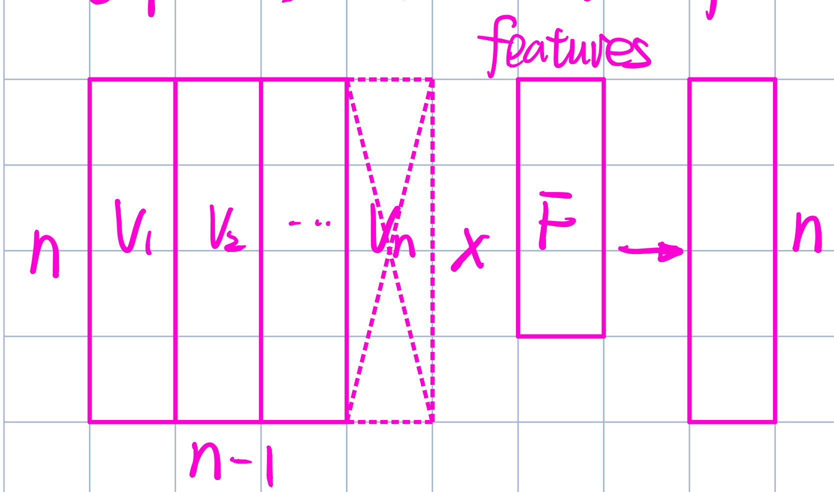
    By selecting V to be the eigenvectors of Σ
    - the largest diagonal element in $V \Lambda V^T$ is maximized
    - the second-to-largest diagonal element in $V \Lambda V^T$ is maximized
    - the third-to-largest diagonal element in $V \Lambda V^T$ is maximized

      $\vdots$
    - the smallest diagonal element in $V \Lambda V^T$ is minimized

- For reducing dimensions
  - we can discard some elements in Ax with low variance, which contains few information

$$
n\underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}}_{n}^{-1} x = Ax \xrightarrow[\text{elements}]{\text{Discard}} \overset{\text{features}}{F}\ n-1
$$

- (optional) we can transfer F back to original basis



4. Application: How to do PCA with samples.

Method 1: From covariance matrix

① Collet data samples $X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}_{N\times M}$ (centered) $x \in \mathbb{R}^M$

② Estimate covariance matrix

$$(N-1)\Sigma = \begin{bmatrix} \sum_{i=1}^N x_{i1}^2 & \sum_{i=1}^N x_{i1}x_{i2} & \cdots & \sum_{i=1}^N x_{i1}x_{iM} \\ \sum_{i=1}^N x_{i2}x_{i1} & \sum_{i=1}^N x_{i2}^2 & \cdots & \sum_{i=1}^N x_{i2}x_{iM} \\ \sum_{i=1}^N x_{i3}x_{i1} & & \ddots & \\ \sum_{i=1}^N x_{iM}x_{i1} & \sum_{i=1}^N x_{iM}x_{i2} & & \sum_{i=1}^N x_{iM}^2 \end{bmatrix} = X^T X$$

M

M

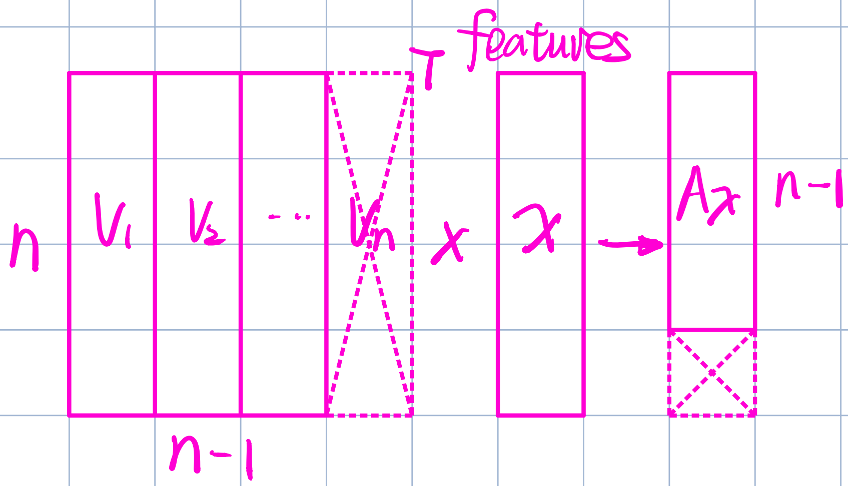
③ Do eigenvalue decomposition

$\Sigma = V \Lambda V^T$

④ Select $A = V^T$

$Ax \sim N(0, A\Sigma A^T) = N(0, \Lambda)$

⑤ Discard elements in Ax with low variance (corresponding to small eigenvalues in Λ) / Remove corresponding basis in A



This method is costy: need to calculate $\mathbf{X}^T\mathbf{X}$

Method 2: To obtain V and Δ, just directly do SVD to $\mathbf{X}$

$$\mathbf{X} = USV^T$$

$$\mathbf{X}^T\mathbf{X} = VSU^TUSV^T = V S^2 V^T$$
$$= V \Delta V^T$$