Lecture 15: Matrix Completion, Correlated Random Variables, and Deconvolution

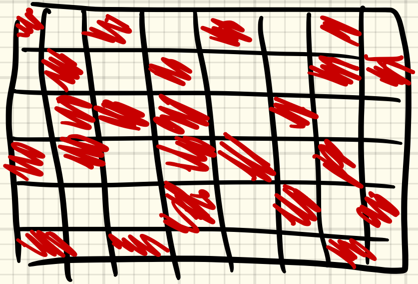
A. Matrix Completion

$$X \in \mathbb{R}^{n \times p}$$ (e.g. n movies, p customers)

assume X is low rank

only observe subset of entries of X

want to fill in remainder

 $= X$

$\Omega = \{(i,j) \text{ s.t. } X_{ij} \text{ is observed}\}$

$$\hat{X} = \arg\min_{M} \operatorname{rank}(M) \quad \text{s.t.} \quad M_{ij} = X_{ij} \ \forall (i,j) \in \Omega$$

intractable

alternative

$$\Rightarrow \quad \hat{X} = \arg\min_{M} \|M\|_* \quad \text{s.t.} \quad M_{ij} = X_{ij} \ \forall (i,j) \in \Omega$$

Computationally tractable

$\operatorname{rank}(X) = \#\{i : \sigma_i > 0\}$

$\|X\|_* = \sum_i \sigma_i$

= trace norm

= nuclear norm

Iterative Singular Value Thresholding

initialize: $\hat{X} = \text{zeros}(n,p)$

$\hat{X}_\Omega = X_\Omega$ ⟵ fill in obs. entries

set threshold or r

for $k = 1, 2, \ldots$

$[U, S, V] = \text{svd}(\hat{X});$

$\hat{S} = S \,.\!*\, (S > \text{threshold})$

or

$\hat{S} = \text{zeros}(n,p);$ for $\ell = 1, \ldots, r,$ $\hat{S}(r,r) = S(r,r);$

$\hat{X} = U \cdot \hat{S} \cdot V^T;$

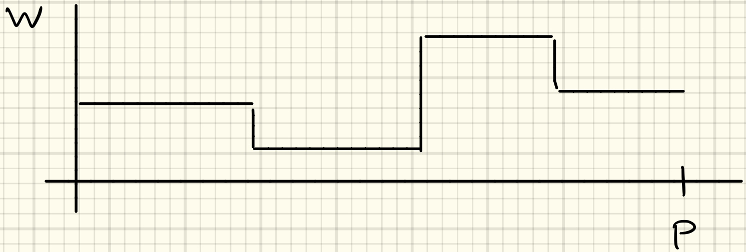
$\hat{X}_\Omega = X_\Omega;$

if converged, stop ⟶ $\|\hat{X} - \hat{X}_{old}\|_F < \varepsilon$

end

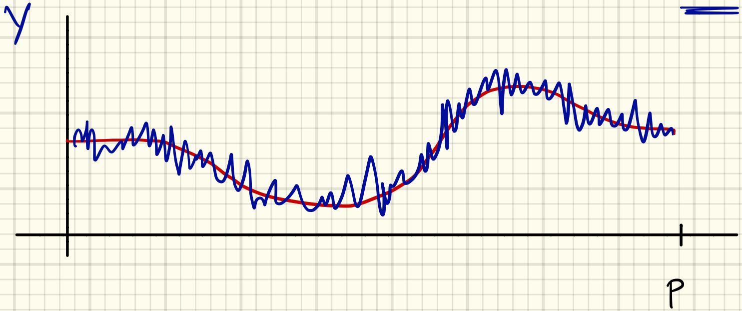
Deconvolution

have signal of interest $w \in \mathbb{R}^p$



we obs blurred version: $\underline{y_i} = \underline{\frac{1}{m}\sum_{j=0}^{m-1} w_{i+j}} + \varepsilon_i$

noise



$y \in \mathbb{R}^p$

define $X \in \mathbb{R}^{p \times p}$ s.t. $y = Xw + \varepsilon$

e.g. $m = 3$

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

X is "circulant"
each row is shifted version of previous row

To est w, $\hat{w} = (X^T X)^{-1} X^T y$

More general formulation:

$$\underbrace{y_i = \sum_{j=1}^{p} h_j \, w_{\underline{i+j} \bmod p}}_{\text{Convolution}} + \varepsilon_i$$

define $X \in \mathbb{R}^{p \times p}$

s.t. $y = Xw + \varepsilon$

$$X = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_p \\ h_p & h_1 & h_2 & \cdots & h_{p-1} \\ h_{p-1} & h_p & h_1 & \cdots & h_{p-2} \\ \vdots & & & & \\ h_2 & h_3 & h_4 & \cdots & h_1 \end{bmatrix}$$

X circulant

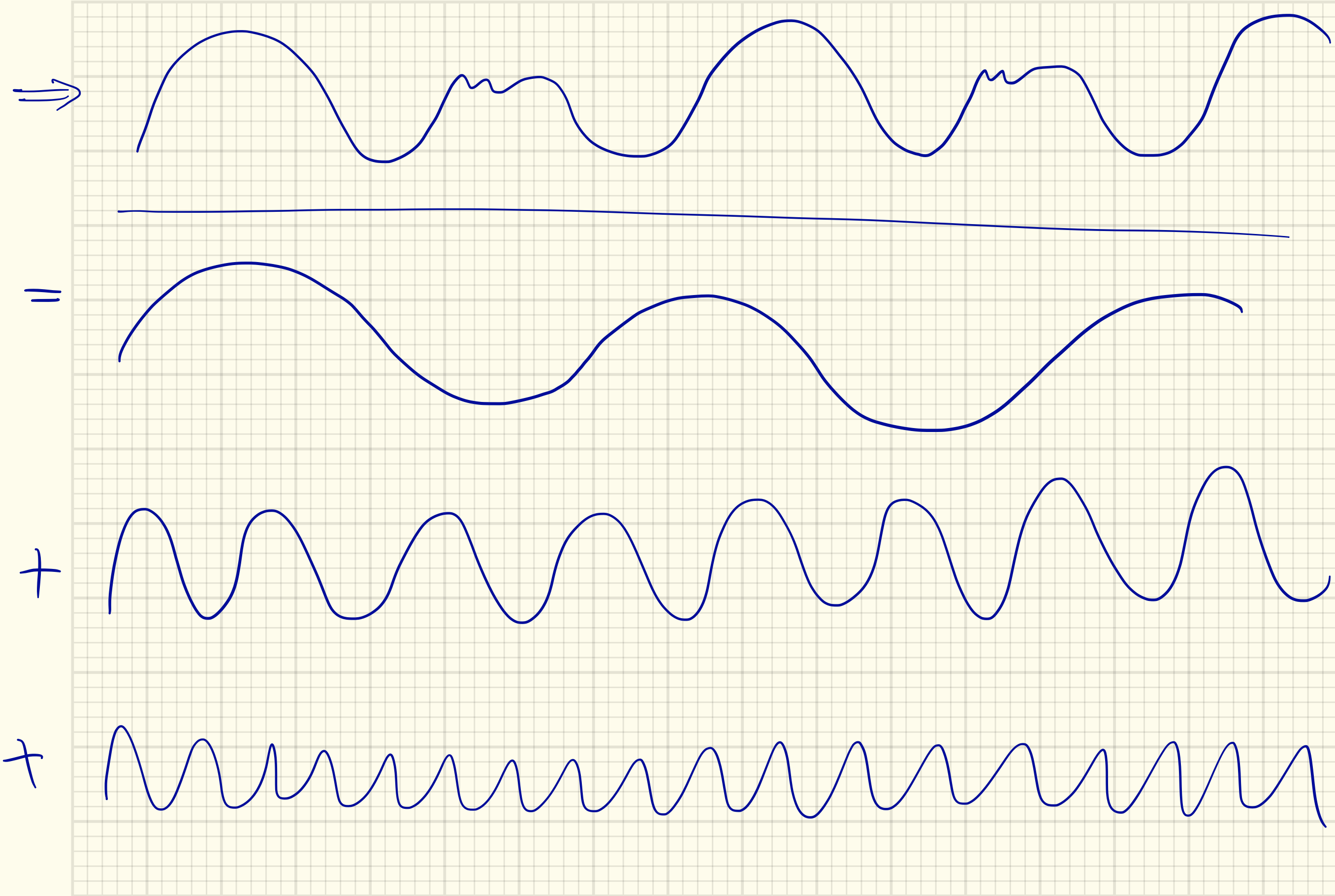
$X = V \Lambda V^H$ (V^H means "Hermitian" transpose $V^H = (V^T)^*$

$V^H V = I$)

when X circulant,
then

- diagonal elts of Λ = Fourier transform coeff's of h
- V is Fourier transform matrix

Fourier transform: represent signals as weighted sums of sinusoids



Fourier transform of h

$$\lambda_1 = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} h_i \, \underbrace{e^{-j2\pi(i-1)0/p}}_{\text{zero freq sine wave} = \text{constant}}$$

$$\lambda_2 = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} h_i \, e^{-j2\pi(i-1)1/p}$$

$$\vdots$$

$$\underline{\lambda_k} = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} h_i \, e^{-j2\pi(i-1)\underline{(k-1)}/p}$$

$(k-1) \rightarrow$ freq $\frac{k-1}{p}$

$$\Longrightarrow h_i = \frac{1}{\sqrt{p}} \sum_{k=1}^{p} \lambda_k \, e^{+j2\pi \underbrace{(i-1)}_{\text{sample}} \underbrace{(k-1)}_{\text{frequency}}/p}$$

(write h as weighted sum of sinusoids)

These sinusoids correspond to the columns of V!

$$v_k = k^{th} \text{ col of } V = \frac{1}{\sqrt{p}} \left[e^{j2\pi(k-1)\cdot 0/p} \quad e^{j2\pi(k-1)1/p} \quad \cdots \quad e^{j2\pi(k-1)(p-1)/p} \right]^T$$

$$\Rightarrow \quad h = V\lambda \quad , \quad \lambda = [\lambda_1, \lambda_2, \ldots, \lambda_p]^T$$

note

$$v_k^H = k^{th} \text{ row of } V^H = \frac{1}{\sqrt{p}} \left[\underline{e^{-j2\pi(k-1)\cdot 0/p}} \quad \underline{e^{-j2\pi(k-1)\cdot 1/p}} \quad \cdots \quad \underline{e^{-j2\pi\frac{(k-1)(p-1)}{p}}} \right]$$

$$\lambda_k = v_k^H h$$

$$\underline{\lambda} = V^H h$$

$y = Xw = \underline{V}\underline{\Lambda}\underline{V^H w}$

eigendecomposition of $X = V\Lambda V^H$

diag elts of Λ = λ_k's.

Consider LS est of w:

$$\hat{w} = (X^H X)^{-1} X^H y$$

$$= (V\Lambda V^H V\Lambda V^T)^{-1} V\Lambda V^H y$$

$$= V\Lambda^{-1} V^H y$$

interp:

1. $V^H y$ = DFT y
2. scale DFT coeffs by dividing by DFT coeff of h
3. inverse DFT

Signal processing perspective:

$$y = w * h \cancel{+ \varepsilon}$$

$$Y = W \cdot H$$

$$\Rightarrow \hat{W} = Y ./ H$$

1. $Y = \mathrm{dft}(y)$
2. $H = \mathrm{dft}(h)$, $\hat{W} = Y ./ H$
3. $\hat{w} = \mathrm{idft}(\hat{W})$

Consider Ridge Regression

$$\hat{w} = \left(V \Lambda^2 V^H + \lambda I\right)^{-1} V \Lambda V^H y$$

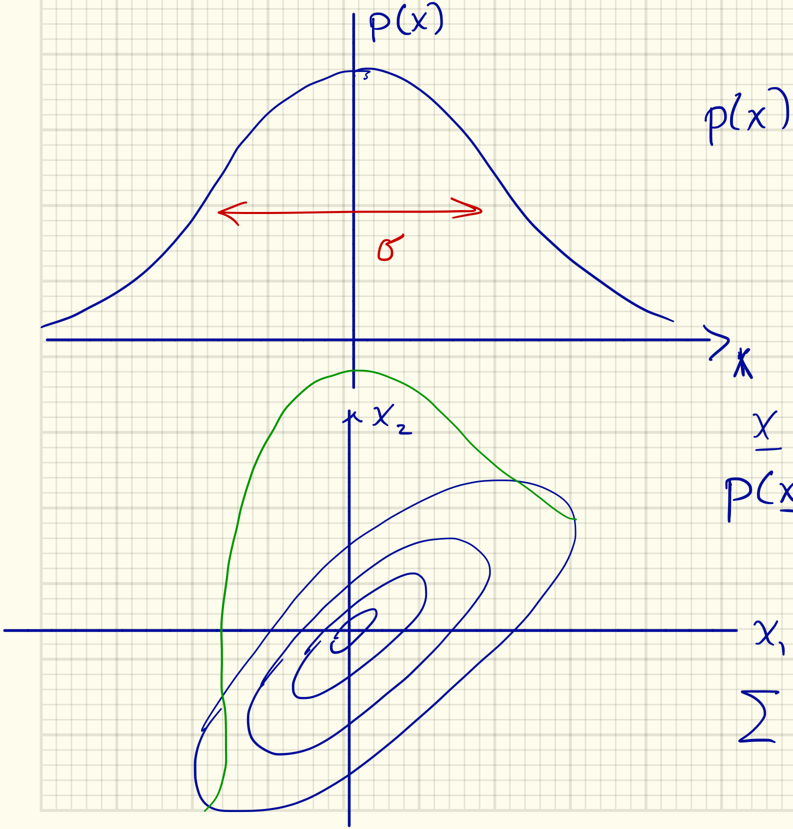
$$= \left(V \left(\Lambda^2 + \lambda I\right) V^H\right)^{-1} V \Lambda V^H y$$

$$= V \underbrace{\frac{\Lambda}{\Lambda^2 + \lambda I}}_{\begin{bmatrix} \frac{\lambda_1}{\lambda_1^2+\lambda} & & & \\ & \frac{\lambda_2}{\lambda_2^2+\lambda} & & \\ & & \ddots & \\ & & & \frac{\lambda_p}{\lambda_p^2+\lambda} \end{bmatrix}} V^H y$$

1. dft(y)
2. rescaling dft coeffs (avoid dividing by zeros)
3. inverse dft

Some λ_k's close to zero

Multivariate normal / Gaussian distribution



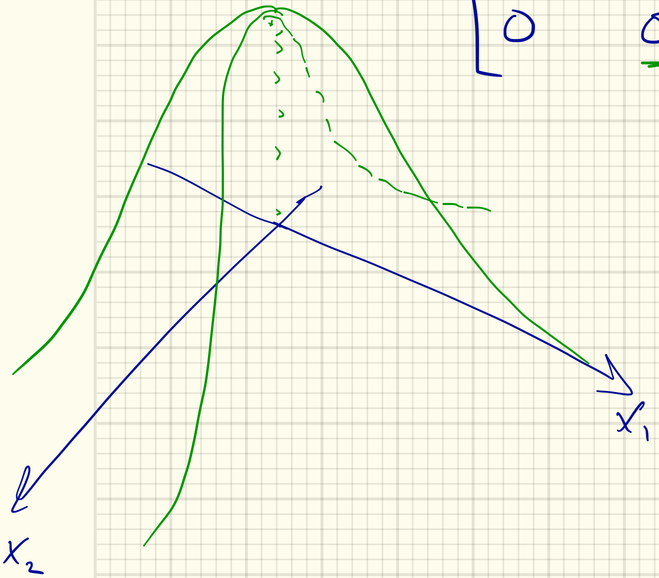
$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2} x^2}$$

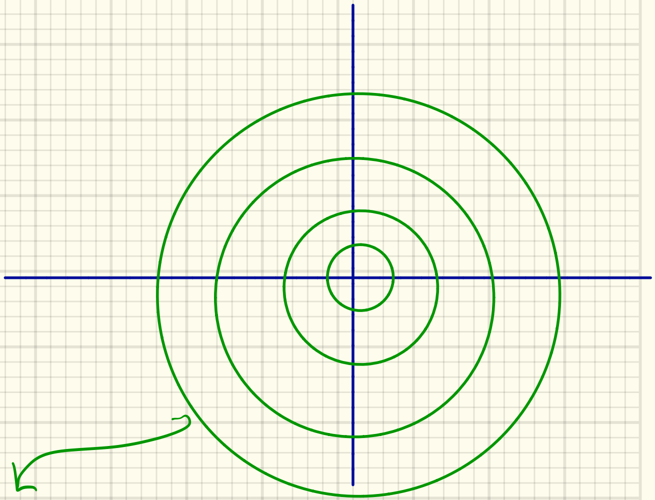
$\underline{x} \in \mathbb{R}^p$

$$p(\underline{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\, e^{-\frac{1}{2}x^T\Sigma^{-1}x}$$

$\Sigma = E[xx^T]$ = covariance matrix

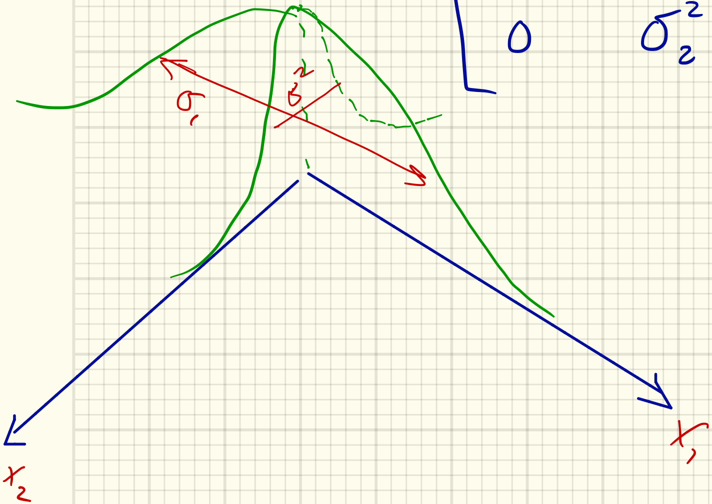
$$p = 2, \quad \Sigma = \begin{bmatrix} \underline{\sigma^2} & 0 \\ 0 & \underline{\sigma^2} \end{bmatrix}$$

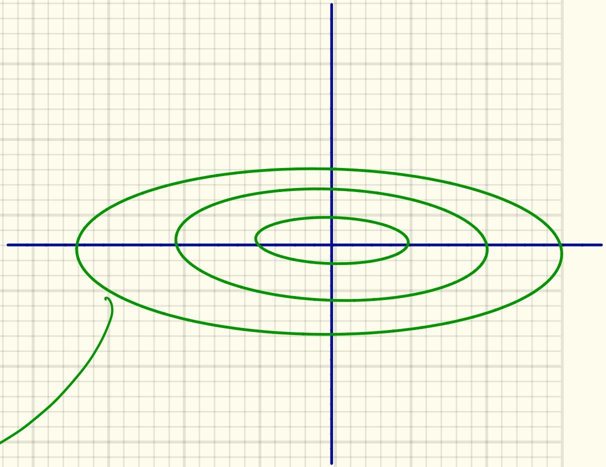




$$\{ x \in \mathbb{R}^2 \ \text{s.t.} \ \|x\| = \gamma \}$$

$$p = 2, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$





$$\left\{ x \in \mathbb{R}^2 : \underbrace{x^T \Sigma^{-1} x} = \gamma \right\}$$

$$\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} = \gamma$$

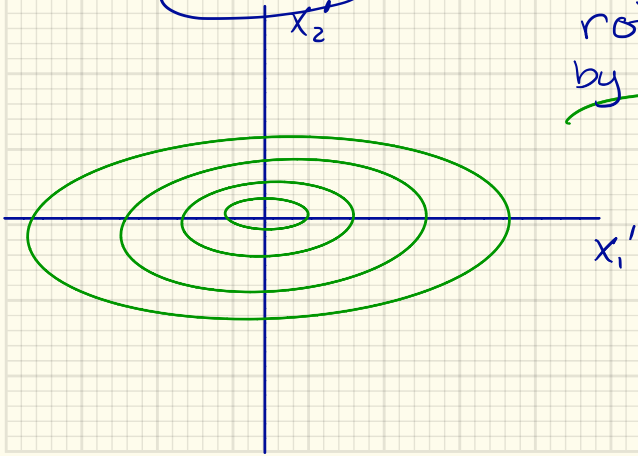
$p = 2$, Σ arbitrary positive definite matrix.

$$\{ x \in \mathbb{R}^2 : x^T \Sigma^{-1} x = \gamma \}$$

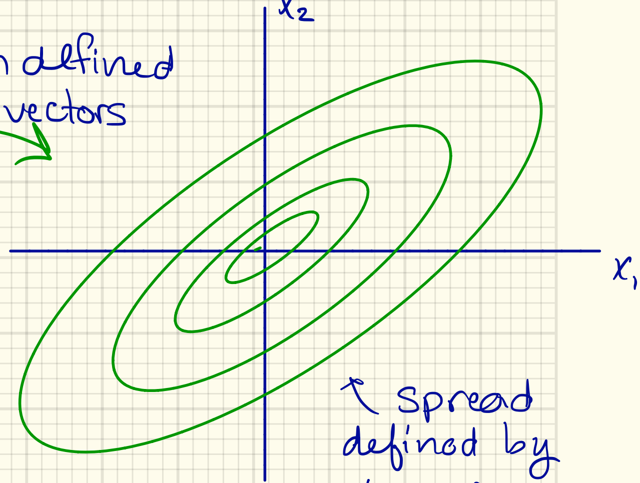
$$\boxed{\Sigma = V \Lambda V^T}$$

$$x^T \Sigma^{-1} x = x^T V \Lambda^{-1} V^T x = x' \underline{\Lambda}^{-1} x' = \gamma$$

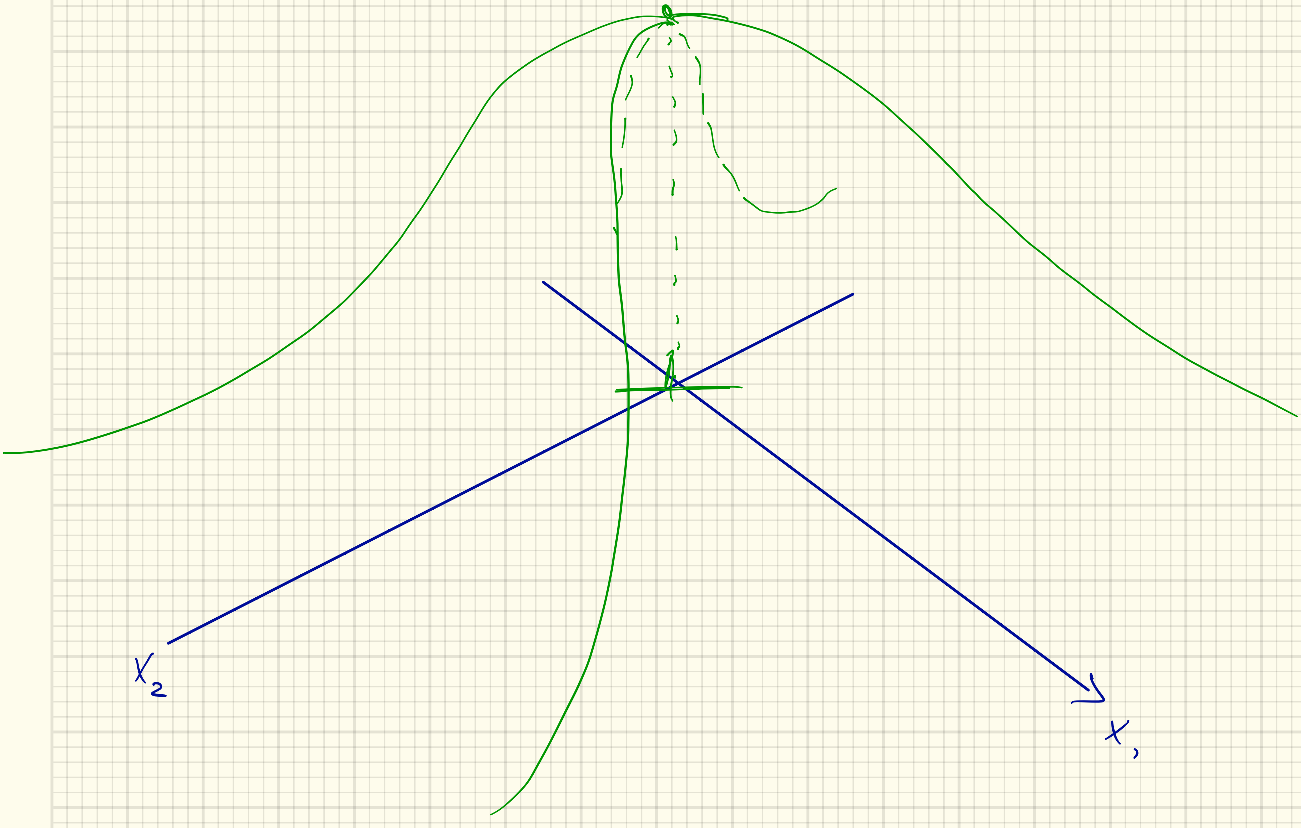
let $x' = V^T x$



rotation defined by eigenvectors



spread defined by eigenvalues.


x_2
x_1