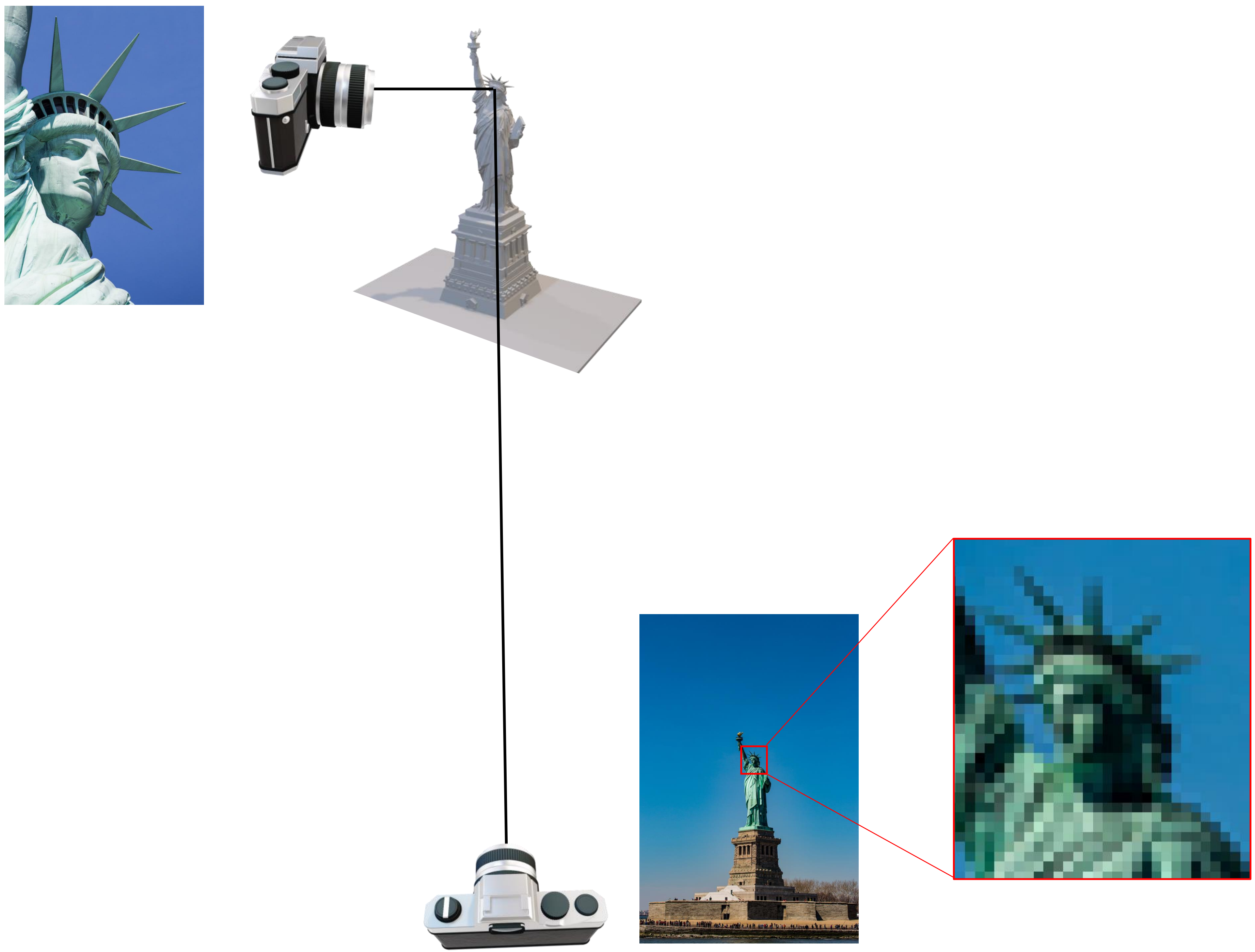
Normal Distribution Covariance Matrix

David Arnon

Triangulation



Covariance
Definition

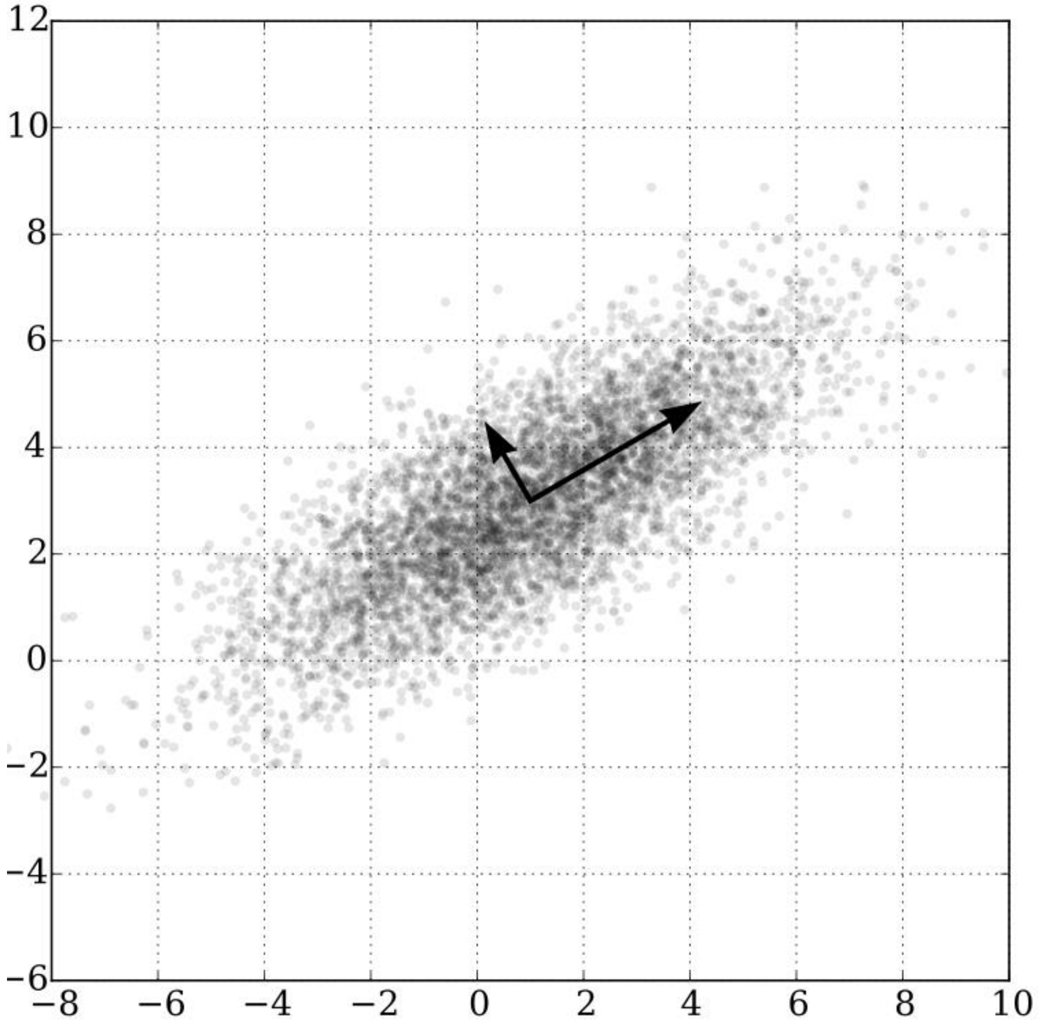
- Let $\boldsymbol{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T$ be a random vector

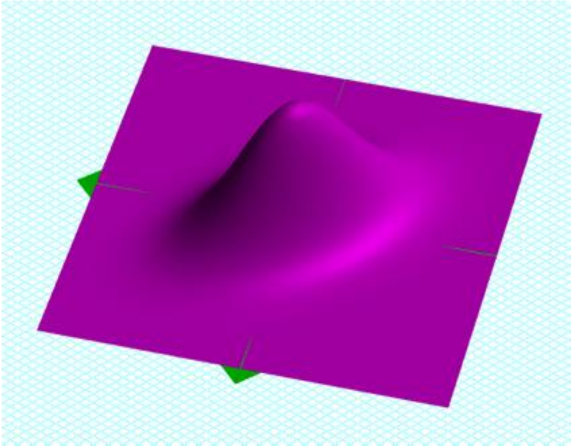
- We measure the coupling of the pair x_i, x_j by the Covariance
$$\mathrm{Cov}(x_i, x_j) = E_x\left[(x_i - \overline{x_i})\left(x_j - \overline{x_j}\right)\right] \underset{\substack{\uparrow \\ \boxed{\text{zero mean}}}}{=} E_x[x_i x_j]$$

- $$\mathrm{Cov}(\boldsymbol{x}) = E_x[(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^T]$$
$$= E[\boldsymbol{x} \cdot \boldsymbol{x}^T] - \overline{\boldsymbol{x}} \cdot \overline{\boldsymbol{x}}^T$$

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & E(x_1 x_2) & \cdots & \\ E(x_2 x_1) & \sigma_{22}^2 & & \\ \vdots & & \ddots & \\ & & & \sigma_{nn}^2 \end{bmatrix}$$

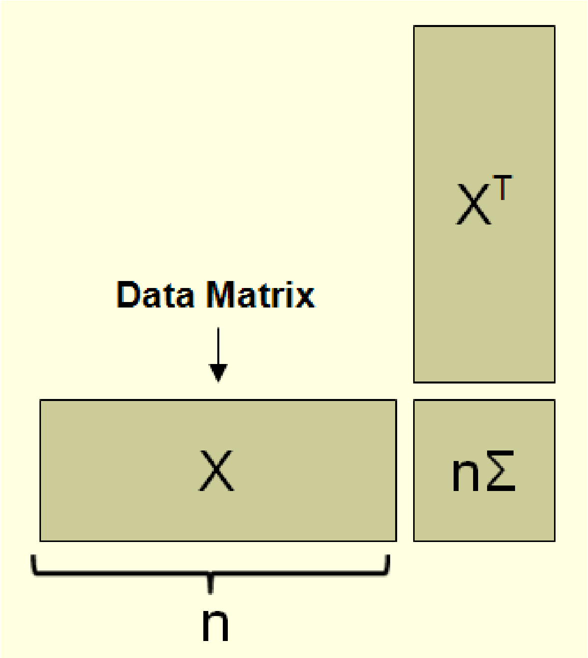
Covariance
Estimation



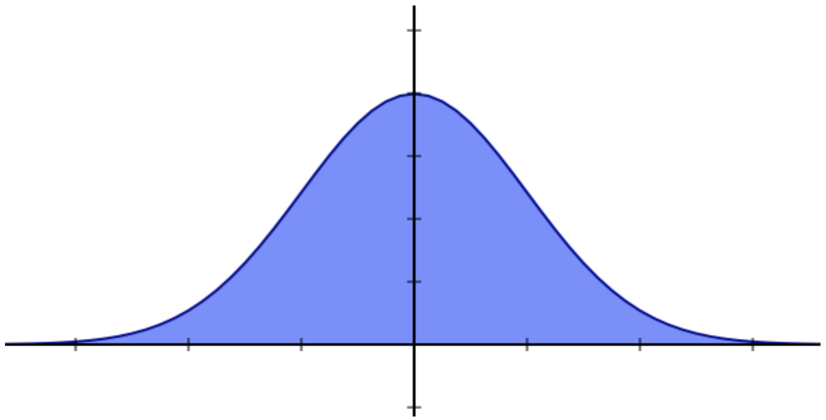


Covariance
Estimation

- $\Sigma = \frac{1}{n} X X^T$ is used as an approximation
  - $\Sigma = \frac{1}{n-1} X X^T$ may be better
- Σ is symmetric and positive semidefinite
  - $v^T (X X^T) v = (X^T v)^T X^T v = \|X^T v\|^2 \geq 0$
- Every symmetric positive semidefinite matrix Σ is a legal covariance matrix and can be expressed as $\Sigma = X X^T$



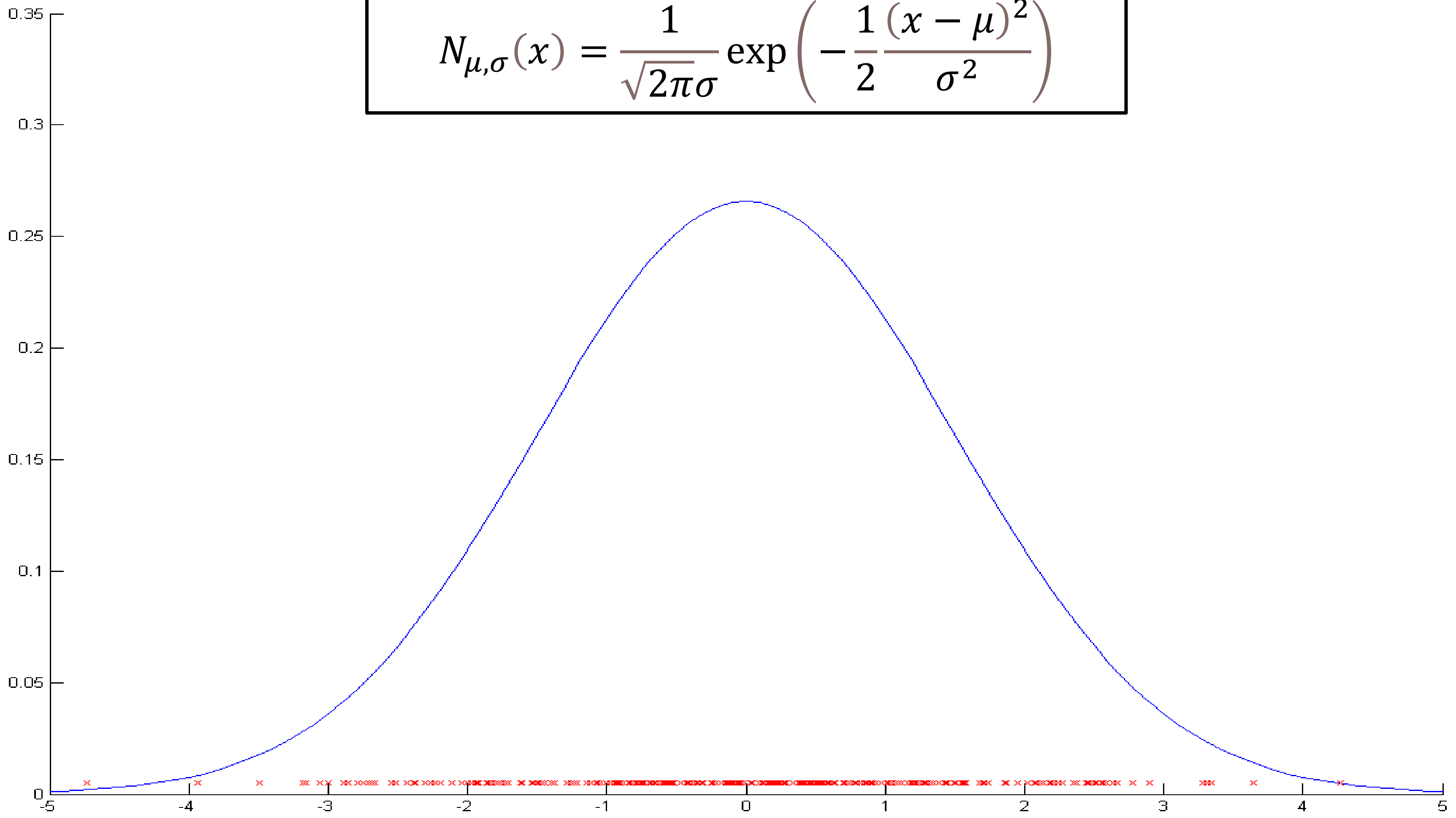
Normal Distribution
Overview



- The most prominent probability distribution
- Very tractable analytically
- Central limit theorem
  - The sum of many independent random variables has normal distribution
- In practice many observed random variables have bell shaped density function

Normal Distribution
1D Gaussian

$$N_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$



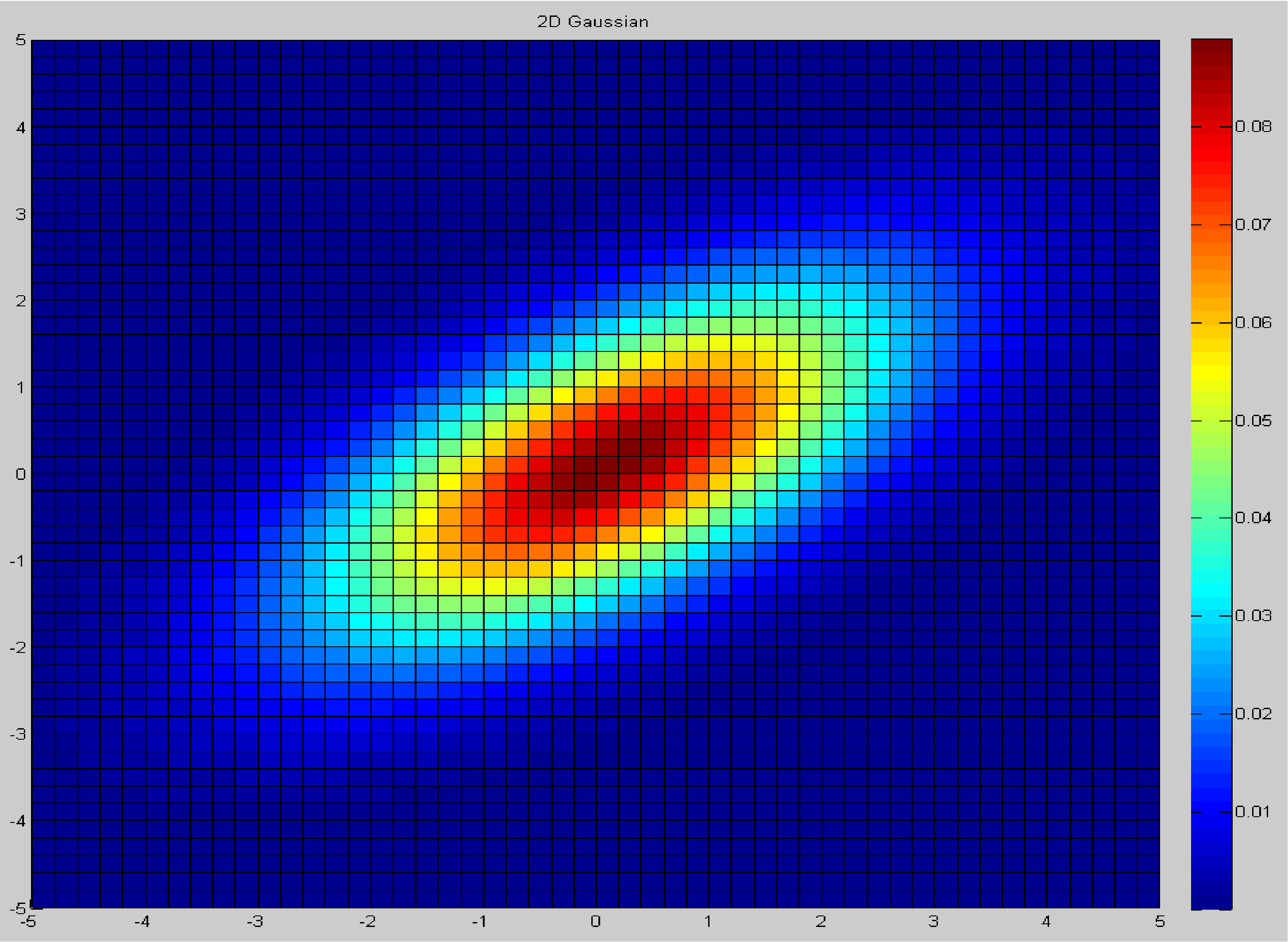
Normal Distribution
General Gaussian

- Given $\mu \in M_{n\times 1}$ and $\Sigma \in M_{n\times n}$ the PDF is given by:

$$\boxed{N_{\mu,\Sigma}(z) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\exp\left(-\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)\right)}$$

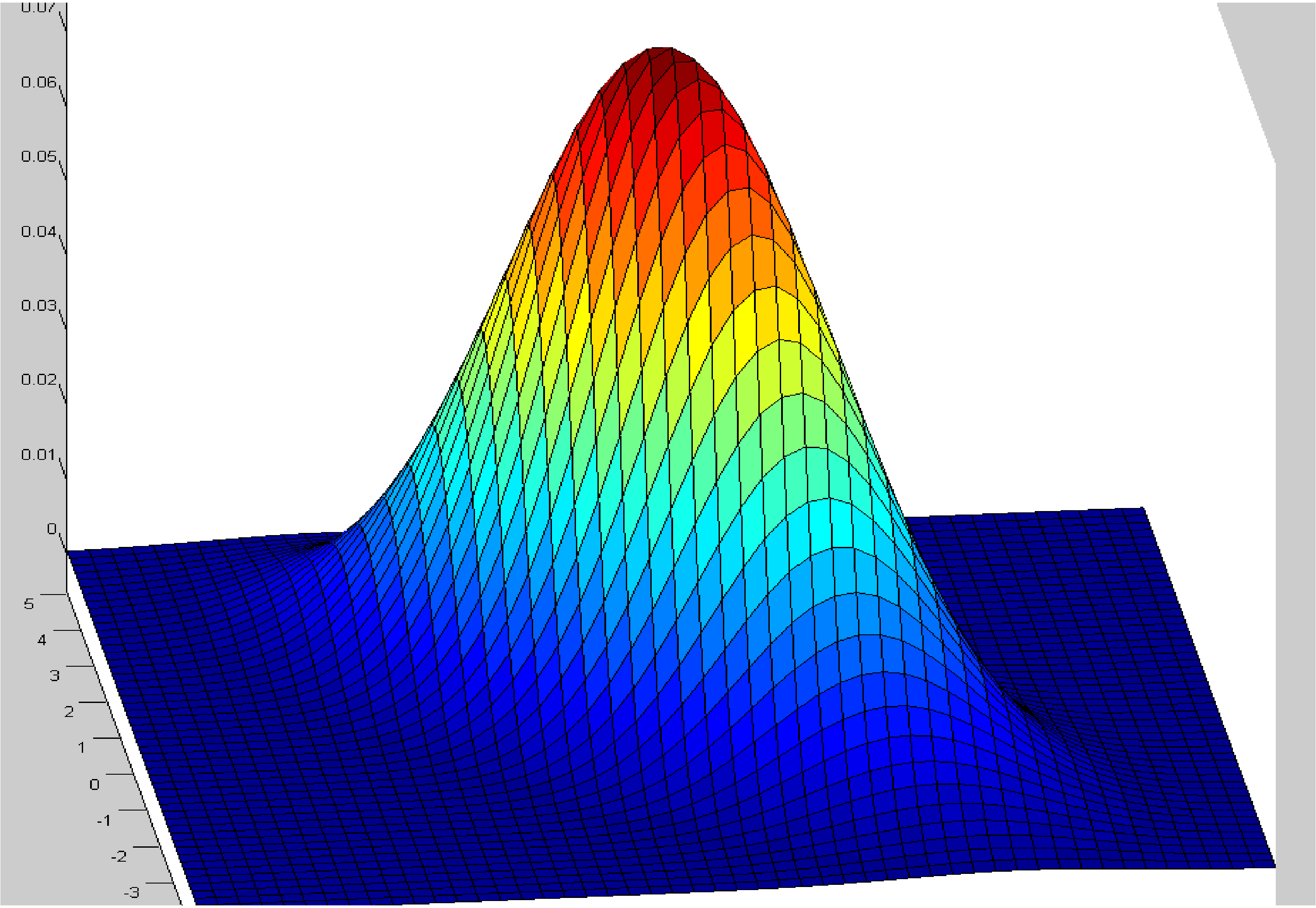
- Visualization in higher dimensions
(especially higher than 3) is more challenging

Gaussian 2D



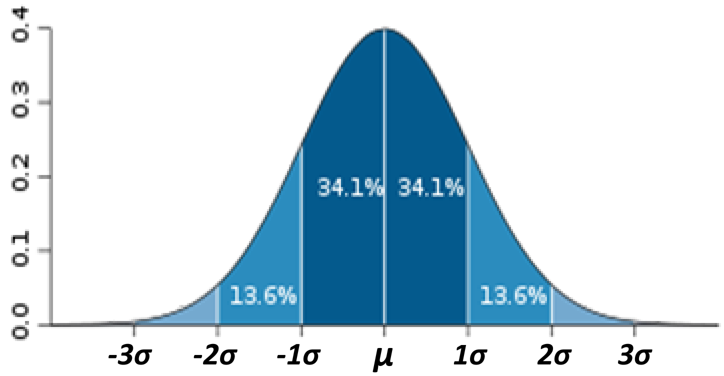
2D Gaussian
5
4
3
2
1
0
-1
-2
-3
-4
-5
-5
-4
-3
-2
-1
0
1
2
3
4
5
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01

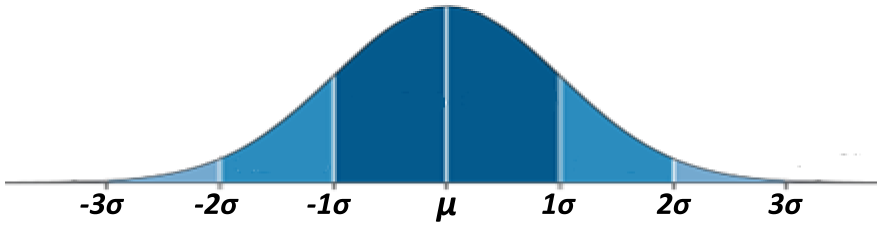
Gaussian 2D



Mahalanobis Distance

- Mahalanobis distance: $r = \frac{|x - \mu|}{\sigma}$
- 68-95-99.7 rule:





- For general dimensions: $r^2 = (z - \mu)^T \Sigma^{-1} (z - \mu)$ $\|z - \mu\|_{\Sigma}^2$

- Intuitively, measures the distance from the mean in standard deviation units.

Cholesky Decomposition

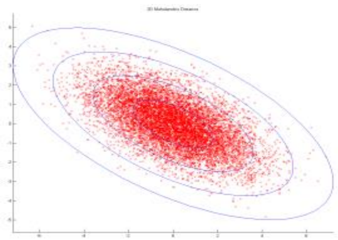
- $\Sigma = C^T C$ is the Cholesky decomposition of Σ if C is upper triangular
  - Every symmetric positive semidefinite matrix has a Cholesky decomposition.

- The locus of points with Mahalanobis distance r is $\boxed{\left\{ rC^T u \middle| \|u\| = 1 \right\}}$

$$(rC^T u)^T \Sigma^{-1} (rC^T u) = r^2 u^T C (C^{-1} C^{-T}) C^T u = r^2 u^T u = r^2$$

$$\boxed{\Sigma^{-1} = C^{-1} C^{-T}}$$

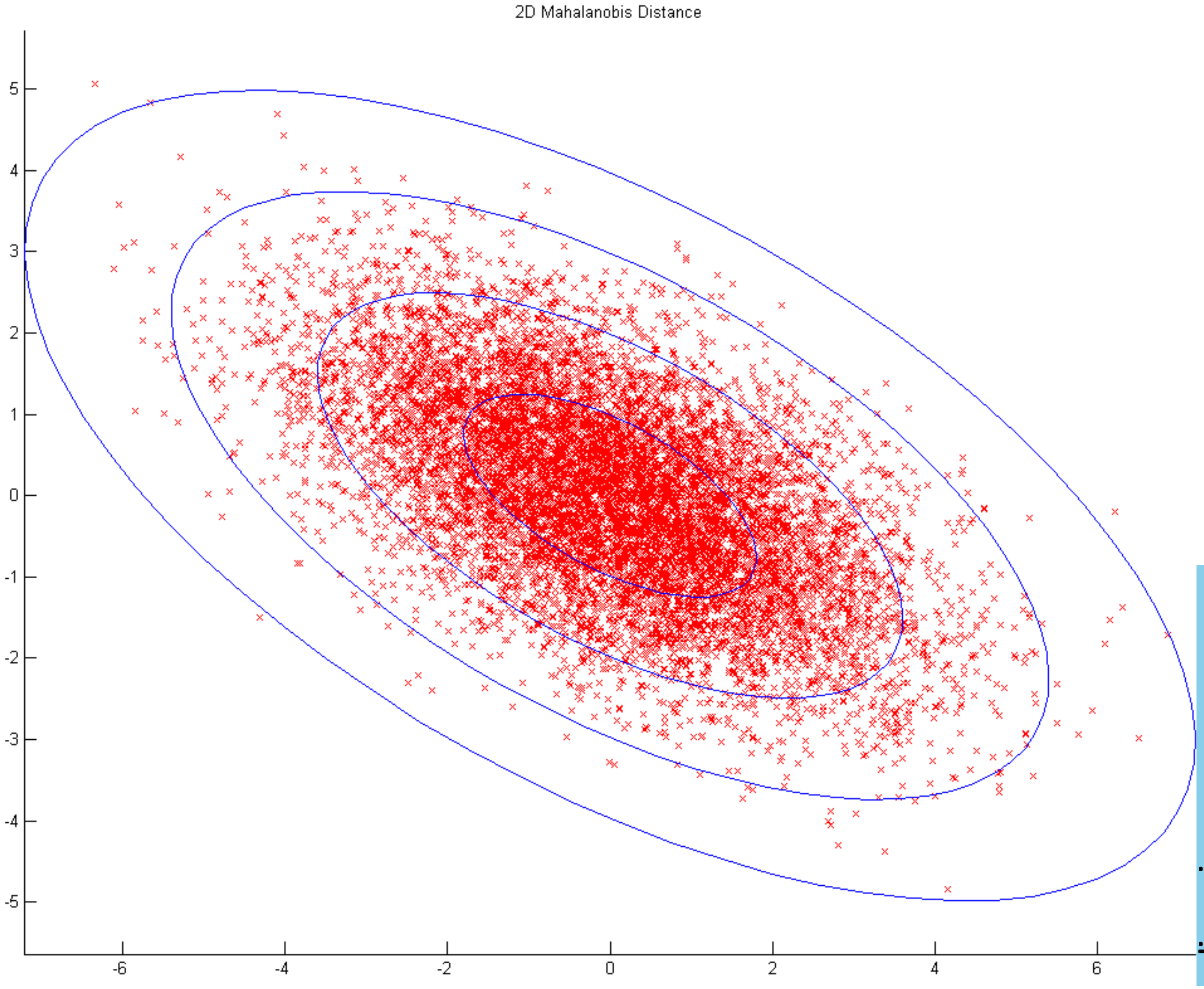
- Used as an intuitive visualization of the covariance matrix.

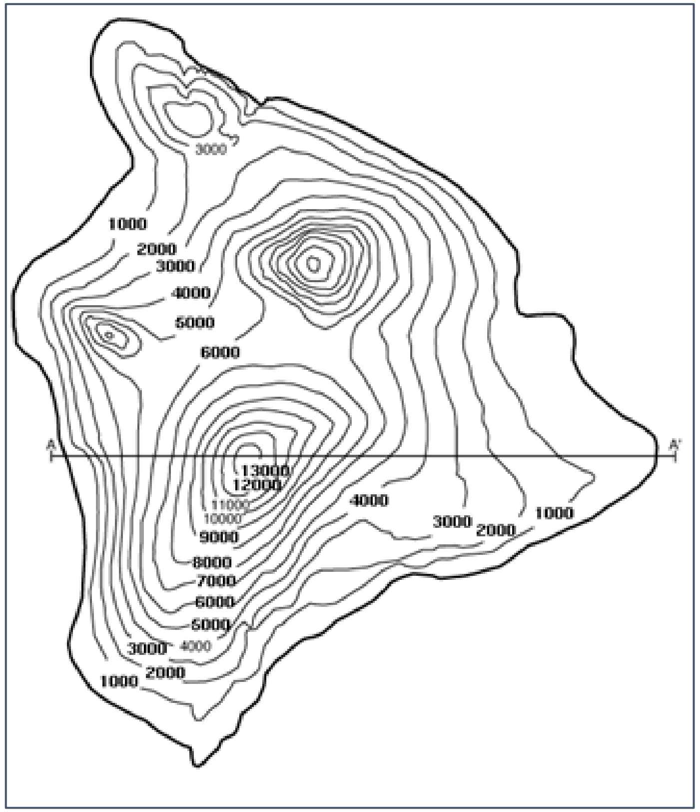


Cholesky Decomposition

- André-Louis Cholesky (1875-1918)
- French with Polish roots
  - Mathematician
  - Cartograph, Geodetic surveyor
    - Crete, North Africa
  - Military Officer

Mahalanobis Distance
Gaussian 2D Visualization





Courtesy of GIS3015 Map Blog Andrea Davis

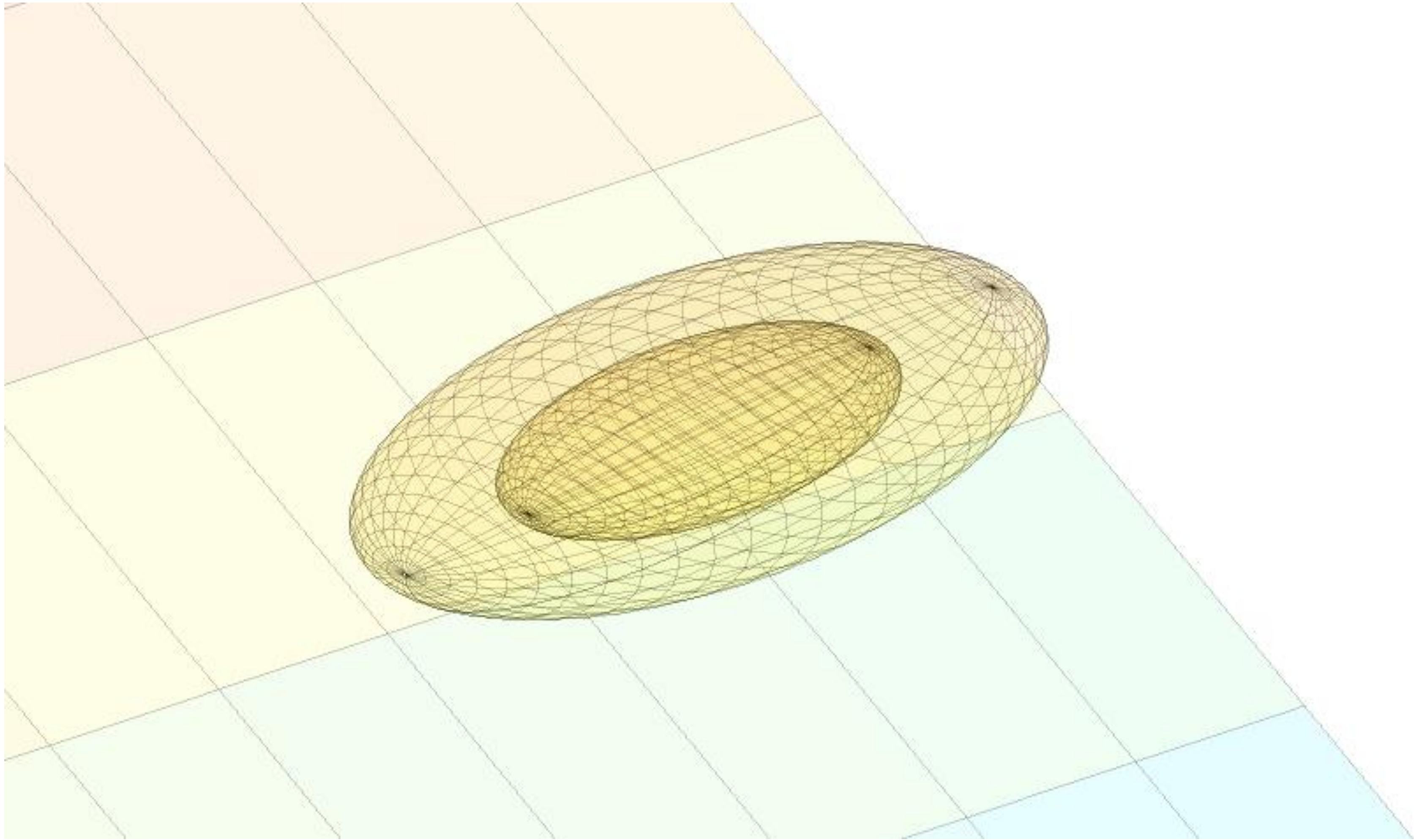
Gaussian Cumulative Distribution Function
Probability bound by Mahalanobis distance, per dimension

Mahalanobis dist.

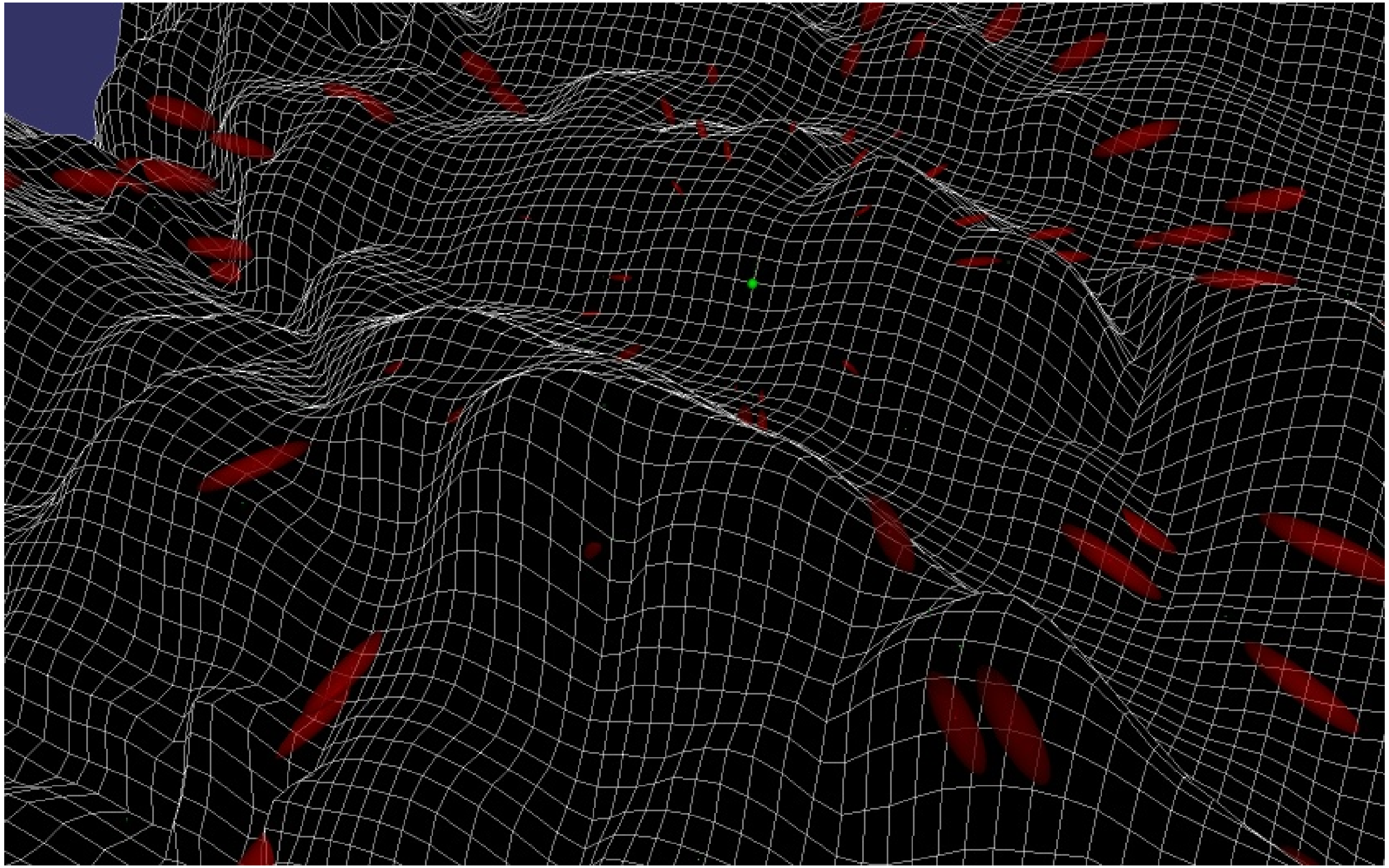
dimension	< 1	< 2	< 3	< 4	> 5
1d	68.3 %	95.4 %	99.7 %	99.99 %	1 : 1744k
2d	39.3 %	86.5 %	98.9 %	99.97 %	1 : 268k
3d	19.9 %	73.9 %	97.1 %	99.89 %	1 : 65k
4d	9.0 %	59.4 %	93.9 %	99.70 %	1 : 20k
5d	3.7 %	45.1 %	89.1 %	99.32 %	1 : 7k
6d	1.4 %	32.3 %	82.6 %	98.62 %	1 : 3k

Mahalanobis Distance

Gaussian 3D Visualization

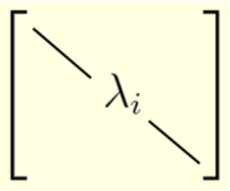


Gaussian 3D Visualization

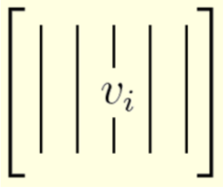


SVD
Singular Value Decomposition

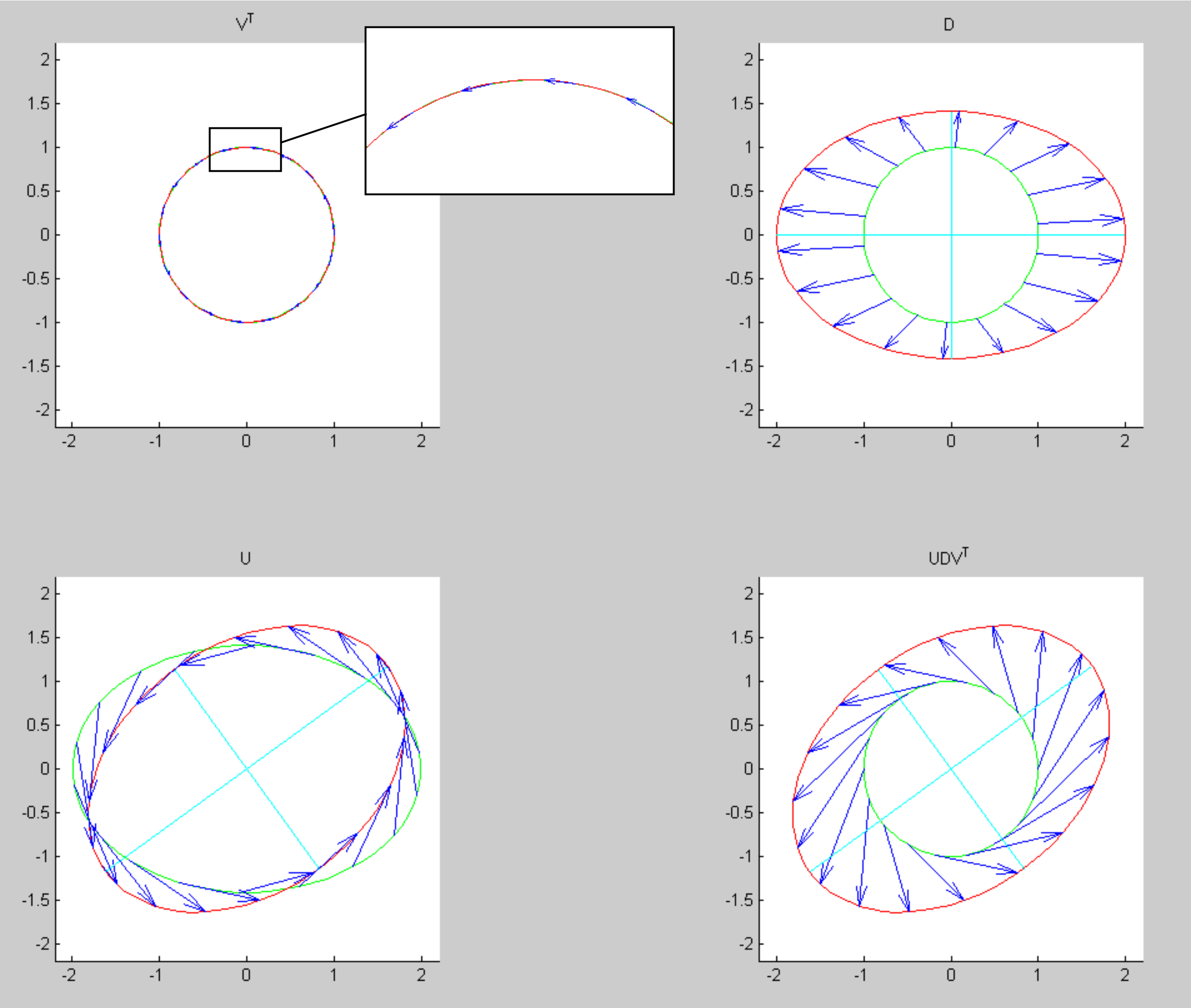
- $A = UDV^T$ is the SVD of A if:
  - $U \in M_{m\times m}$ Orthonormal $\quad (U^TU = I_{m\times m})$
  - $V \in M_{n\times n}$ Orthonormal $\quad (V^TV = I_{n\times n})$
  - $D \in M_{m\times n}$ Diagonal with non-negative entries ordered in descending order.
- D diagonal entries are:
  - called **singular values** of A
  - square root of the **eigenvalues** of A^TA



- V columns are the **eigenvectors** of A^TA



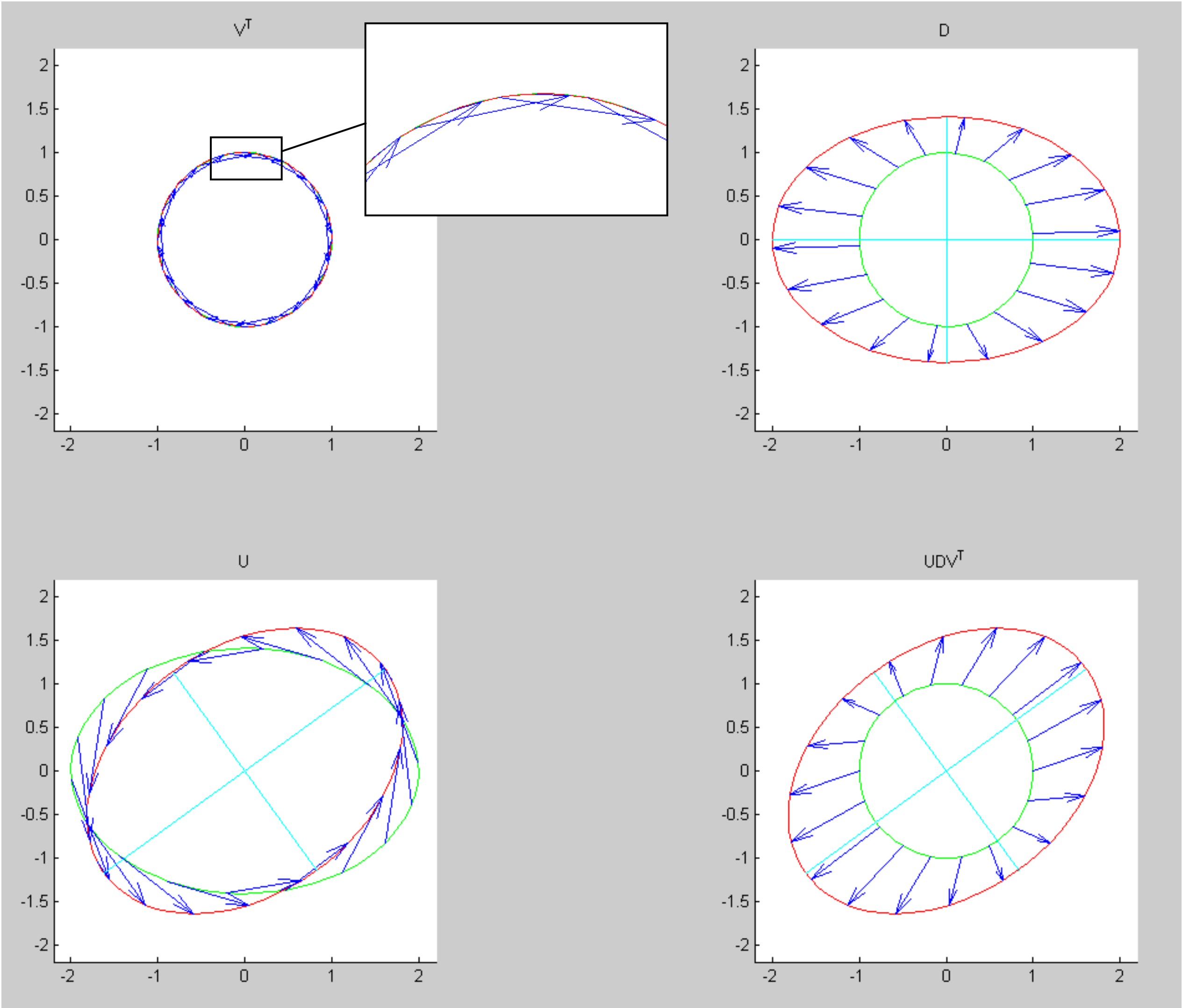
SVD: $A = UDV^T$



Covariance
SVD

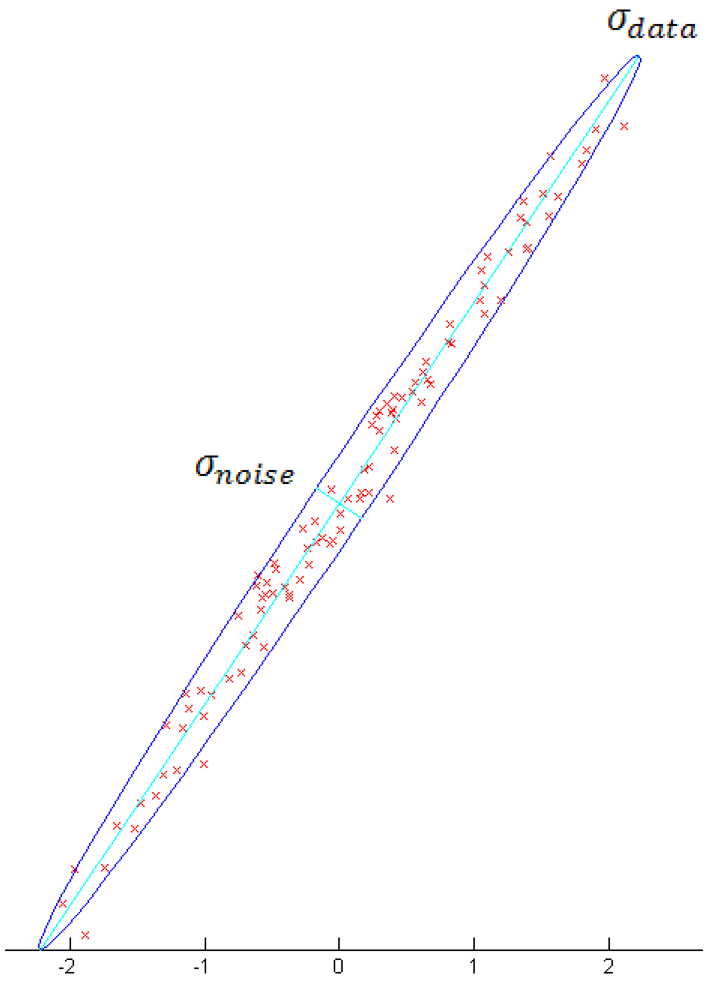
- Covariance matrix SVD properties
  - $\Sigma = UDU^T$ (i.e. $V = U$)
  - Since Σ is symmetric (hermitian) the spectral theorem holds
  - $\Sigma = XX^T = (USV^T)(USV^T)^T = USV^TVSU^T = US^2U^T$
- D diagonal entries are the eigenvalues of Σ
- U columns are the eigenvectors of Σ

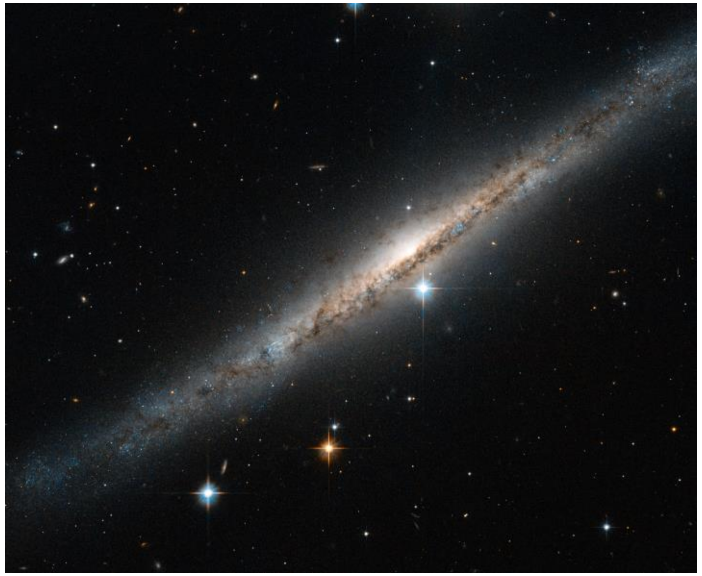
SVD
Covariance Matrix



Covariance
Semantics

- How can we recognize the directions with small variance?
- How can we remove the noise from the data?

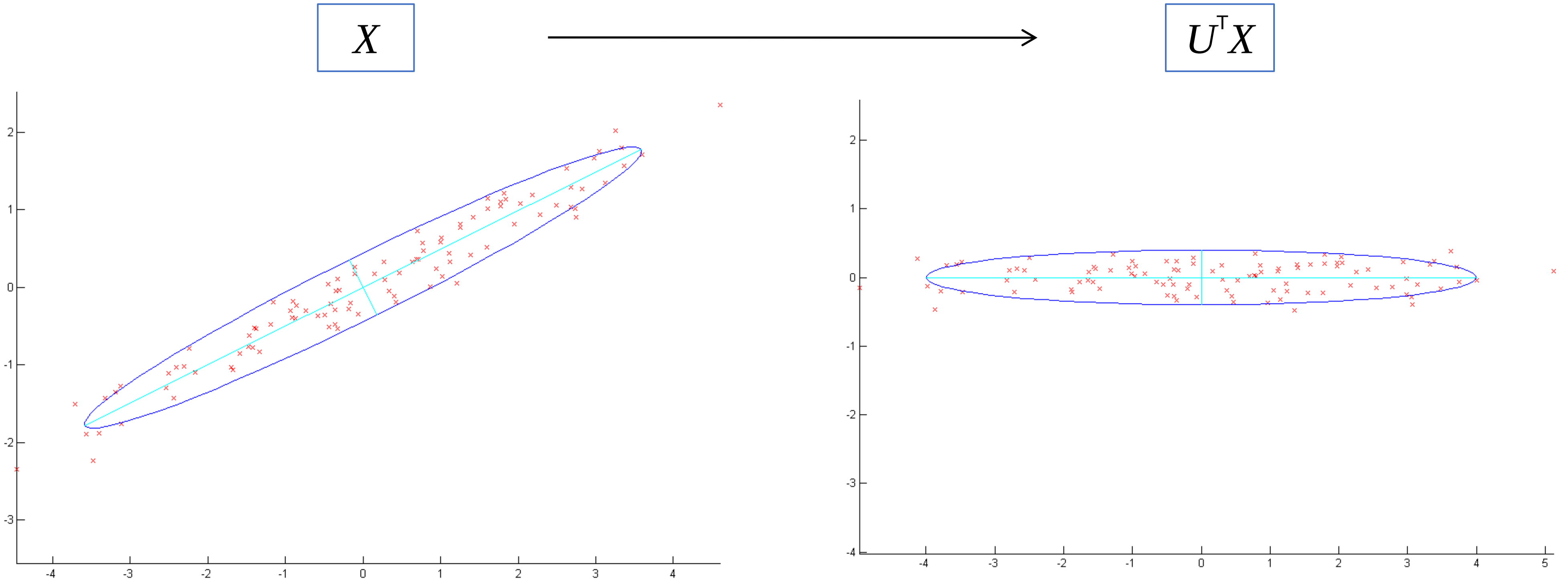




Courtesy of ESA/Hubble & NASA

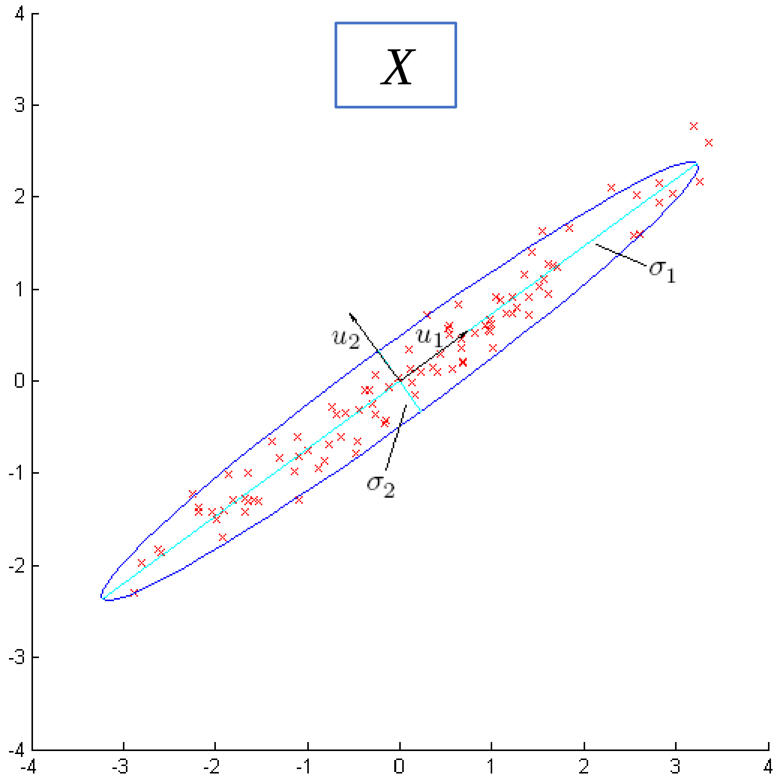
Covariance
Principal Components

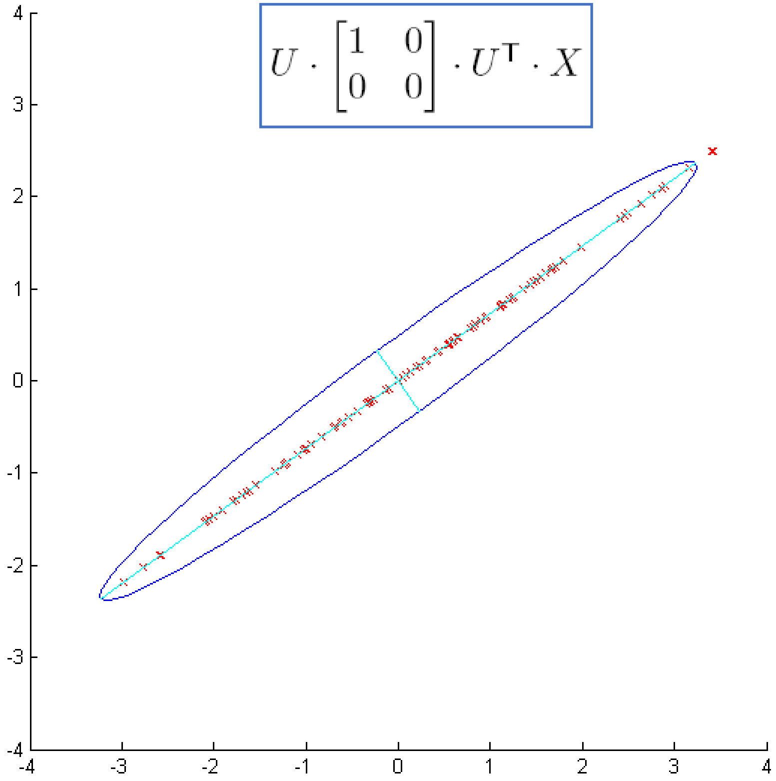
- $\frac{1}{n}XX^T = \Sigma = UDU^T \quad \rightarrow \quad nD = (U^TX)(U^TX)^T$
  - U^TX is decorrelated.
  - D diagonal holds the variance of U^TX on each axis.
  - U columns are called the *principal components* of X



Covariance
Principal Components

- $\frac{1}{n}XX^T = \Sigma = UDU^T = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \cdot \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \cdot \begin{bmatrix} - u_1 - \\ - u_2 - \end{bmatrix}$

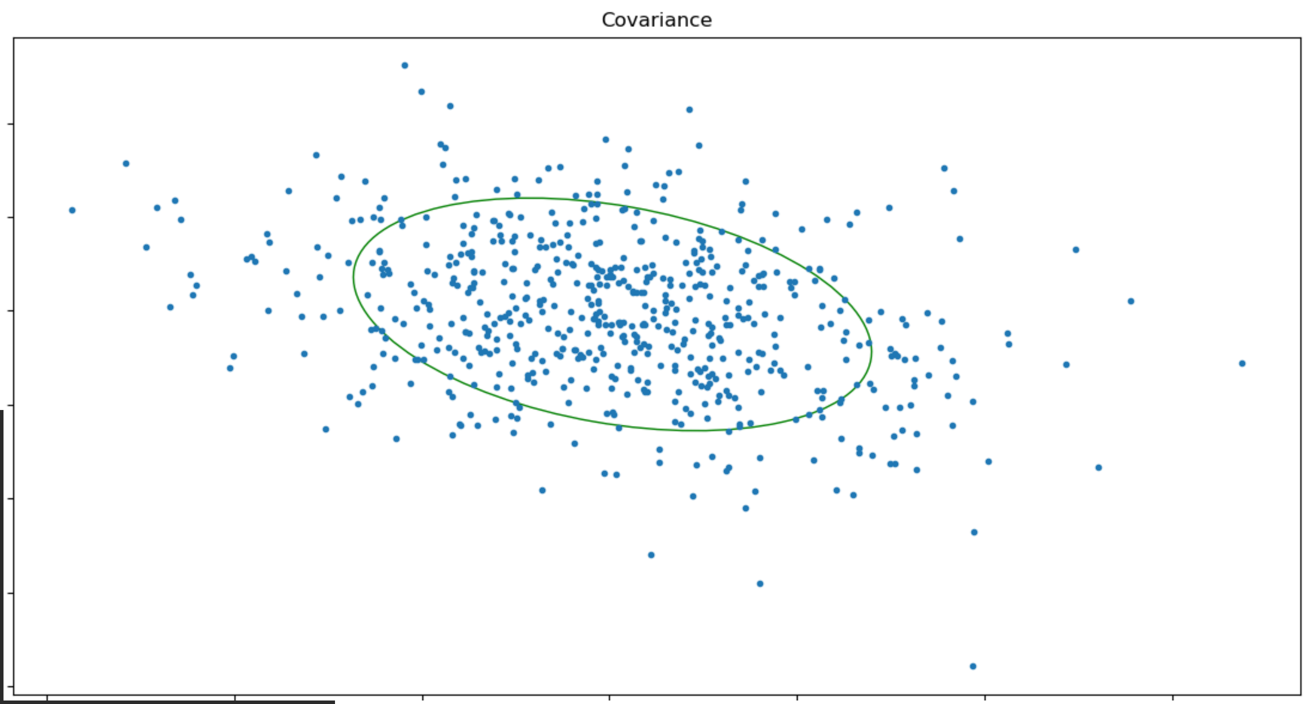




Covariance
Drawing

```python
def draw_cov(points):
    mean = np.mean(points, axis=1).reshape([2, 1])
    p_centered = points - mean
    n = points.shape[1]
    cov = (1.0/n) * p_centered @ p_centered.T
    u, d, _ = np.linalg.svd(cov)
    sig1 = math.sqrt(d[0])
    sig2 = math.sqrt(d[1])
    angle = math.atan2(u[0, 1], u[0, 0]) * 180 / np.pi

    plt.figure()
    plt.plot(points[0, :], points[1, :], '.')
    plt.axis('equal')
    ellipse = Ellipse(mean, sig1, sig2, angle=angle, fill=False)
    plt.gca().add_patch(ellipse)
```



class `matplotlib.patches.Ellipse`(*xy*, *width*, *height*, *angle=0*, ***kwargs*)

Bases: `matplotlib.patches.Patch`

A scale-free ellipse.

Parameters:

xy : (float, float)

xy coordinates of ellipse centre.

width : float

Total length (diameter) of horizontal axis.

height : float

Total length (diameter) of vertical axis.

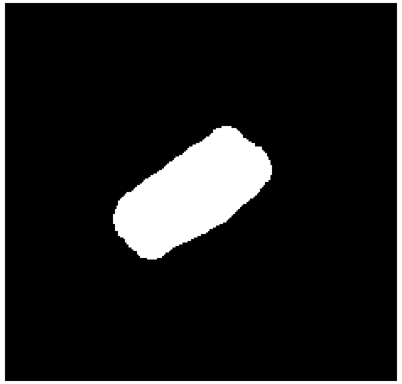
angle : scalar, optional

Rotation in degrees anti-clockwise.

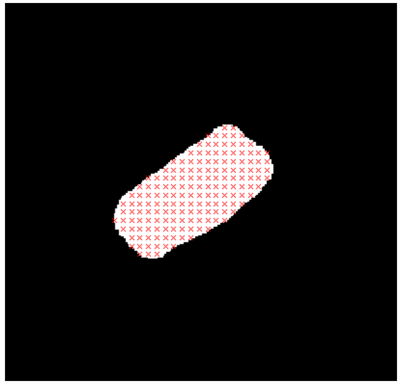
Covariance
Principal Components

Image

Detection



X



SVD