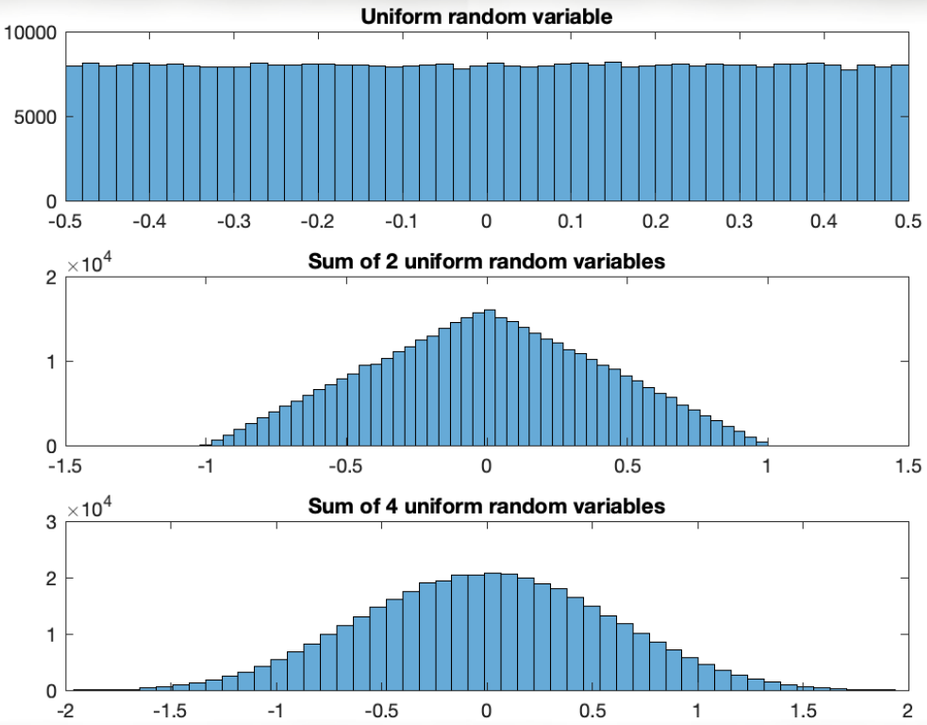
PCA: Principle Component Analysis

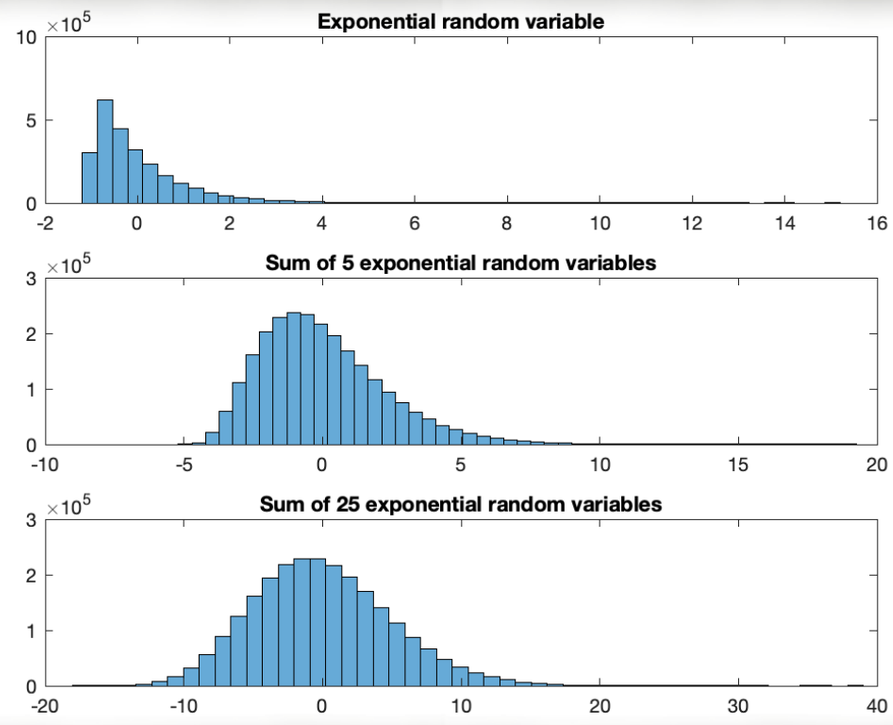
Dr. Keith Evan Schubert

Why Gaussian?

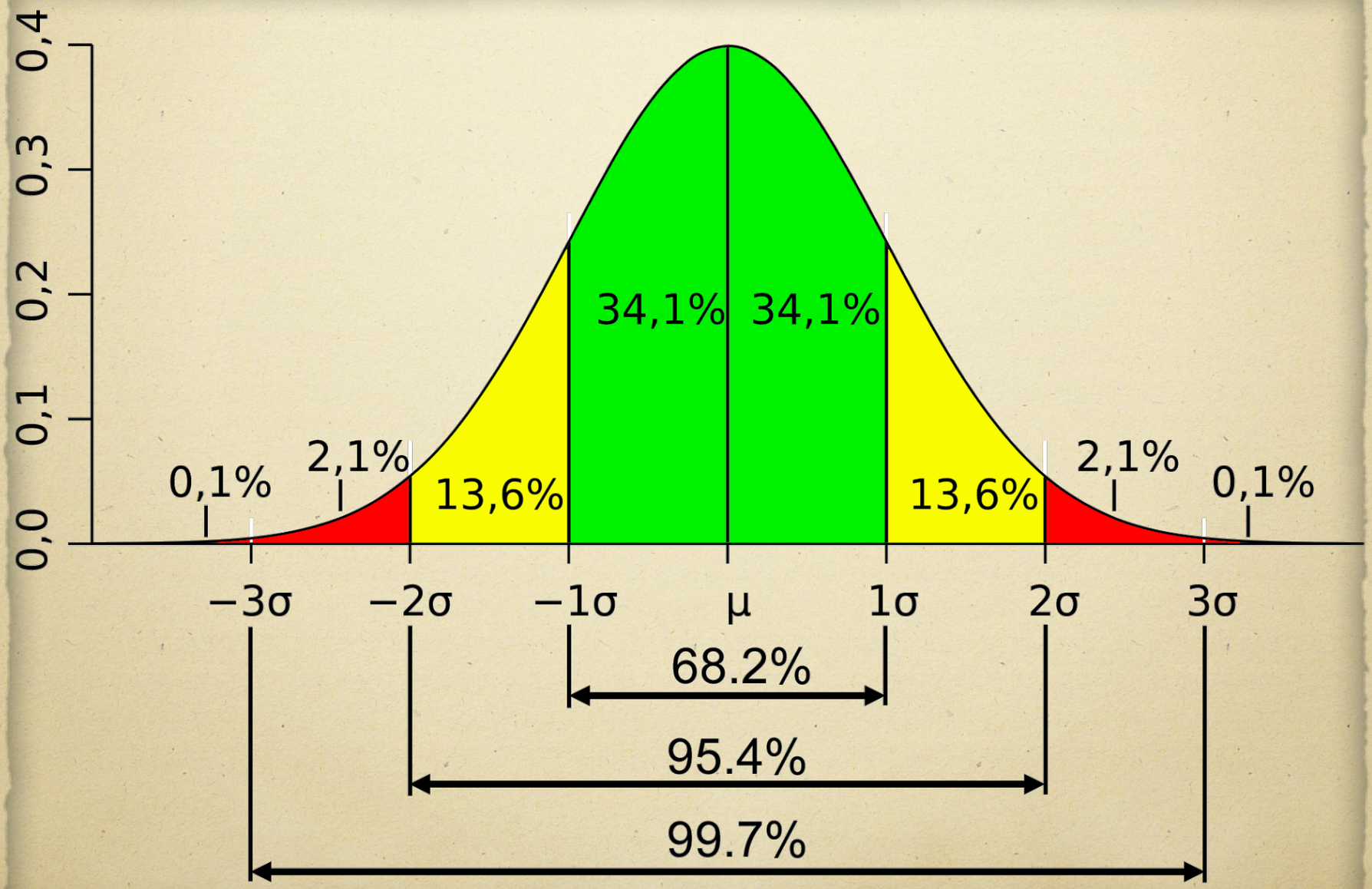
- Gaussian through a linear system produces Gaussian output

Why Gaussian?



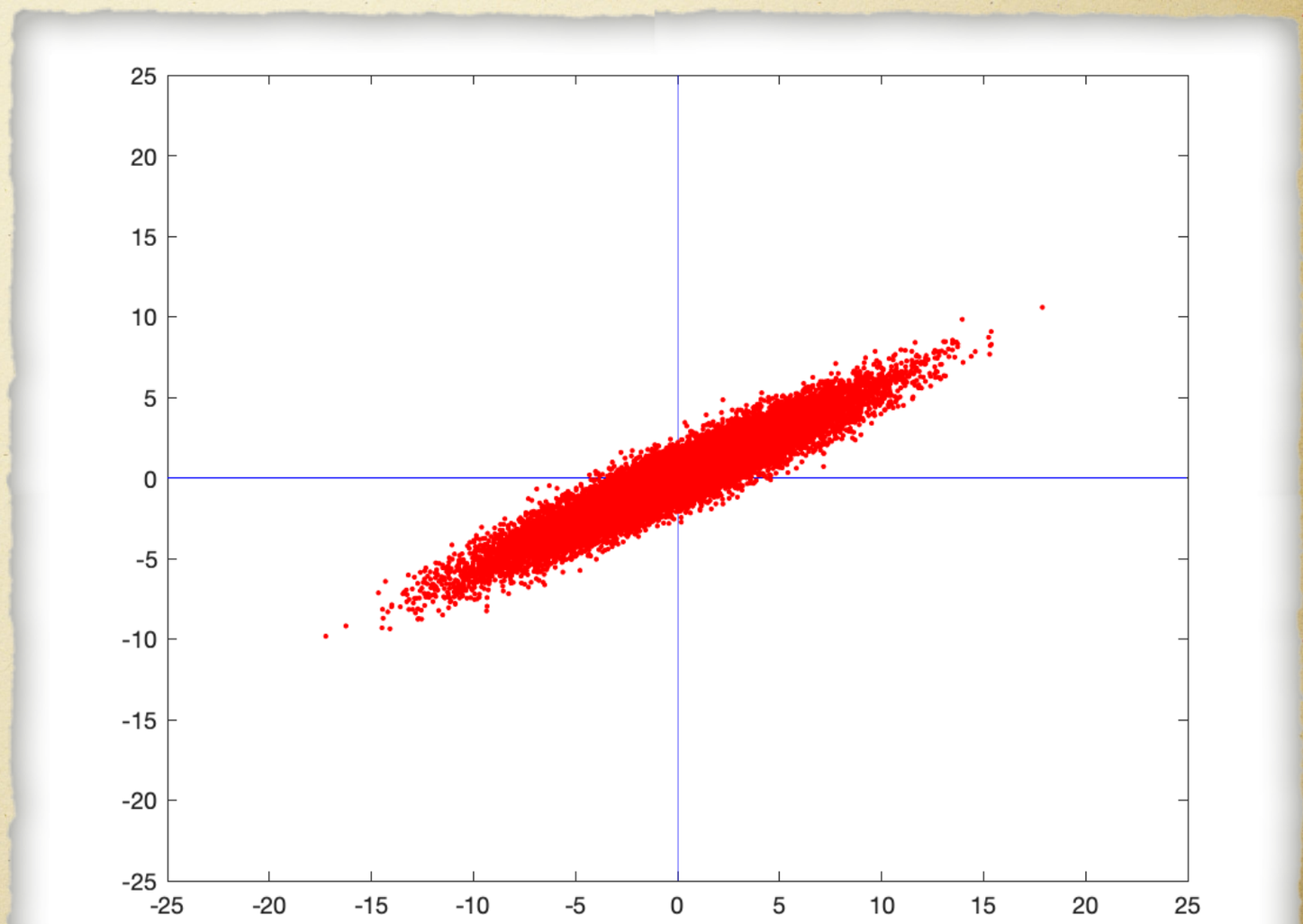


Mean and standard deviation



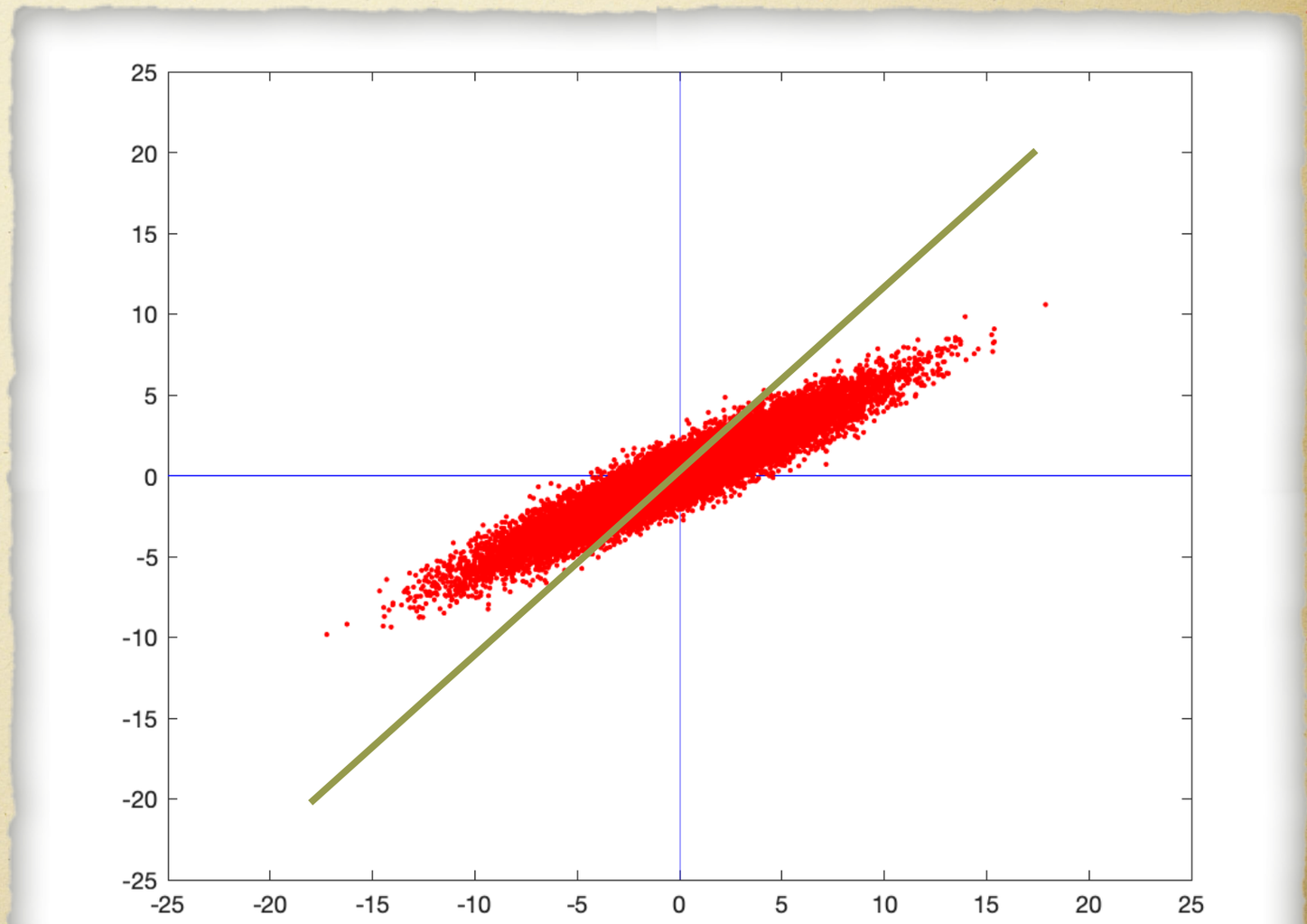
Multivariate Gaussians

- Standard deviation forms ellipses



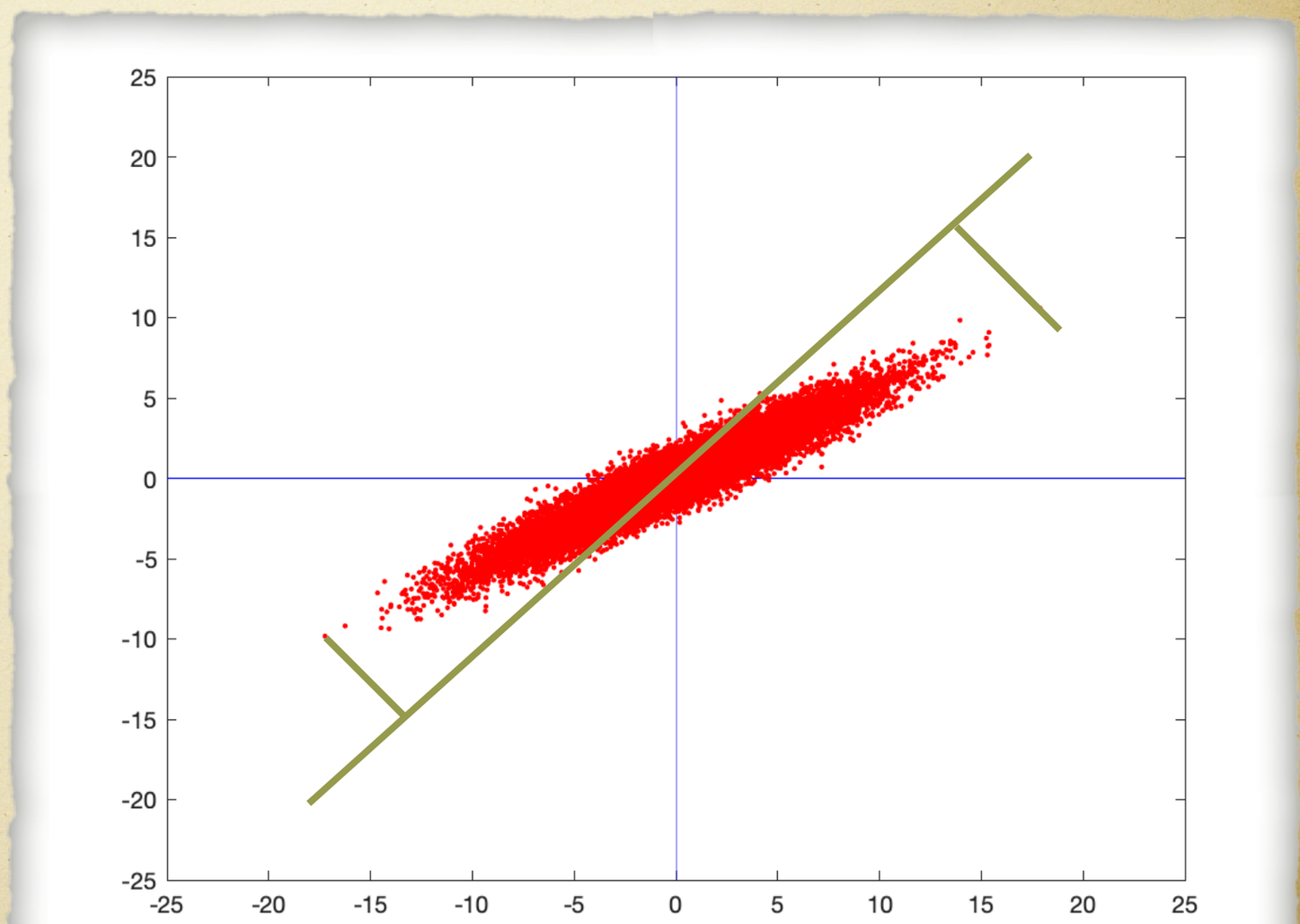
Geometric View

- Standard deviation forms ellipses
- Fit a line



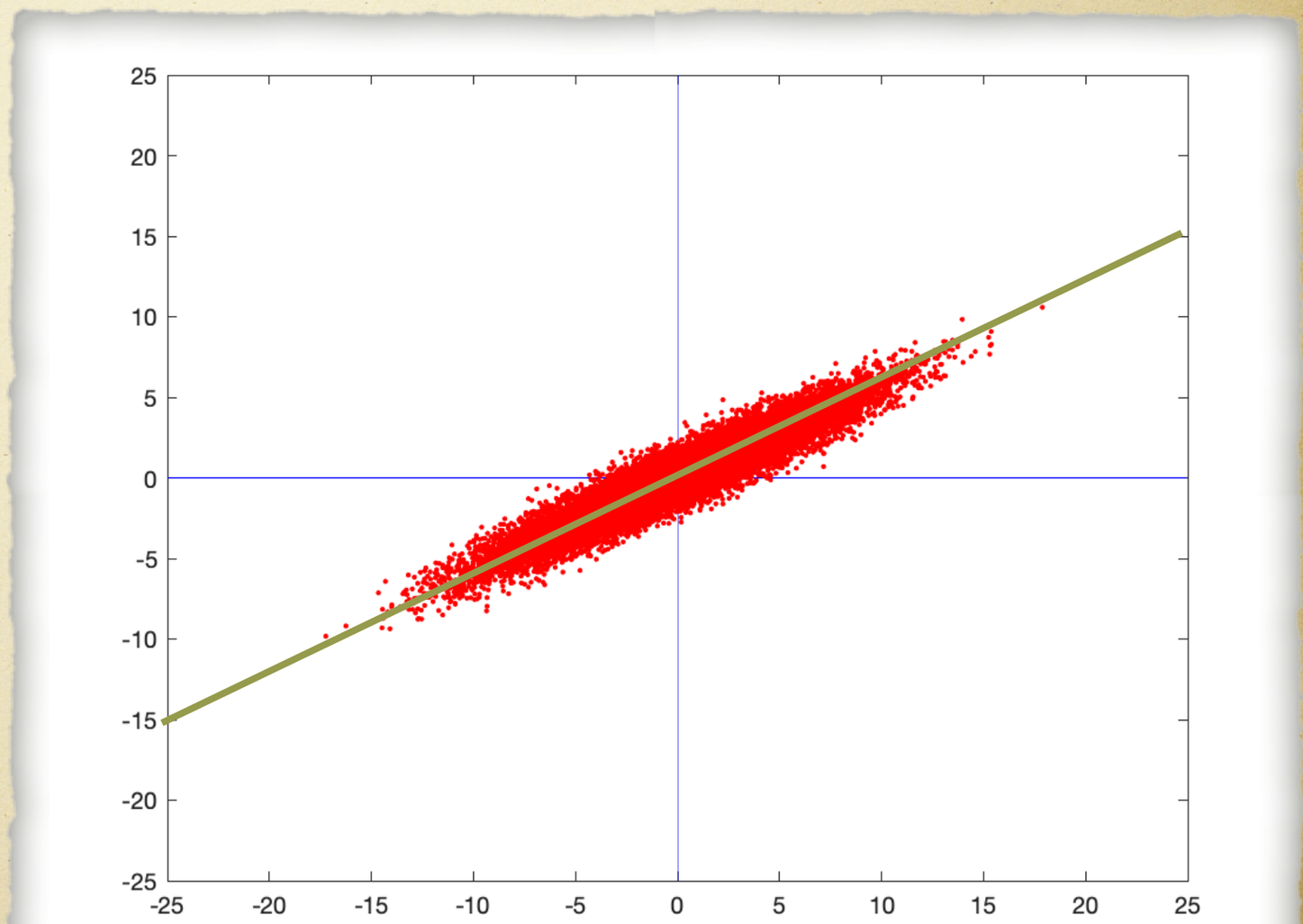
Geometric View

- Standard deviation forms ellipses
- Fit a line
  - Min residual
  - Max variance

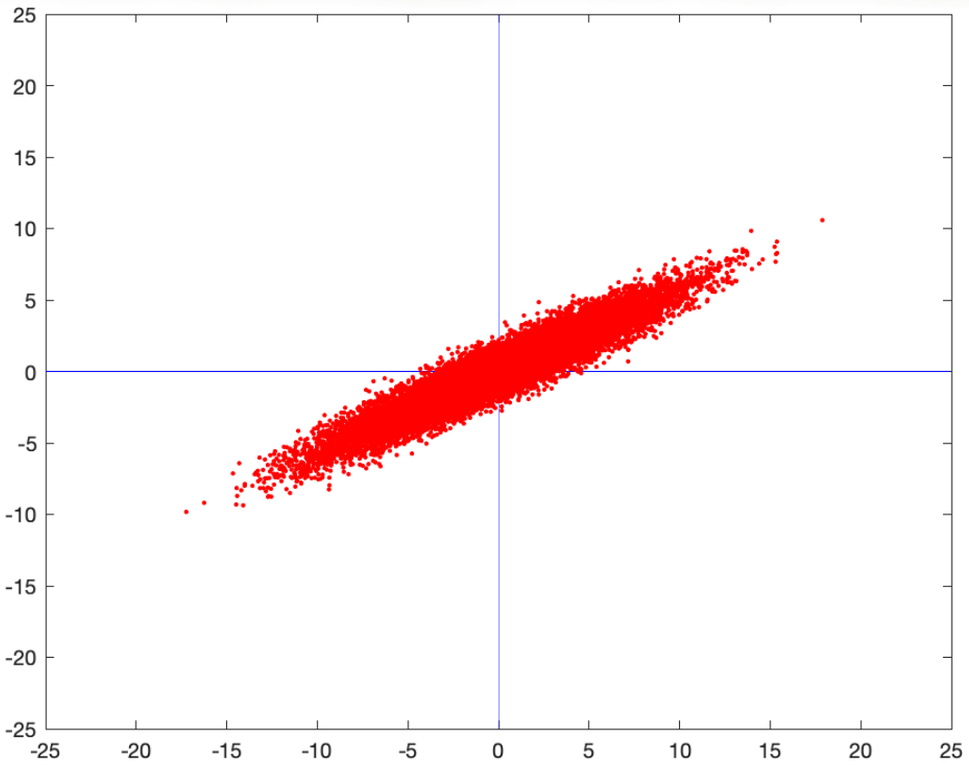


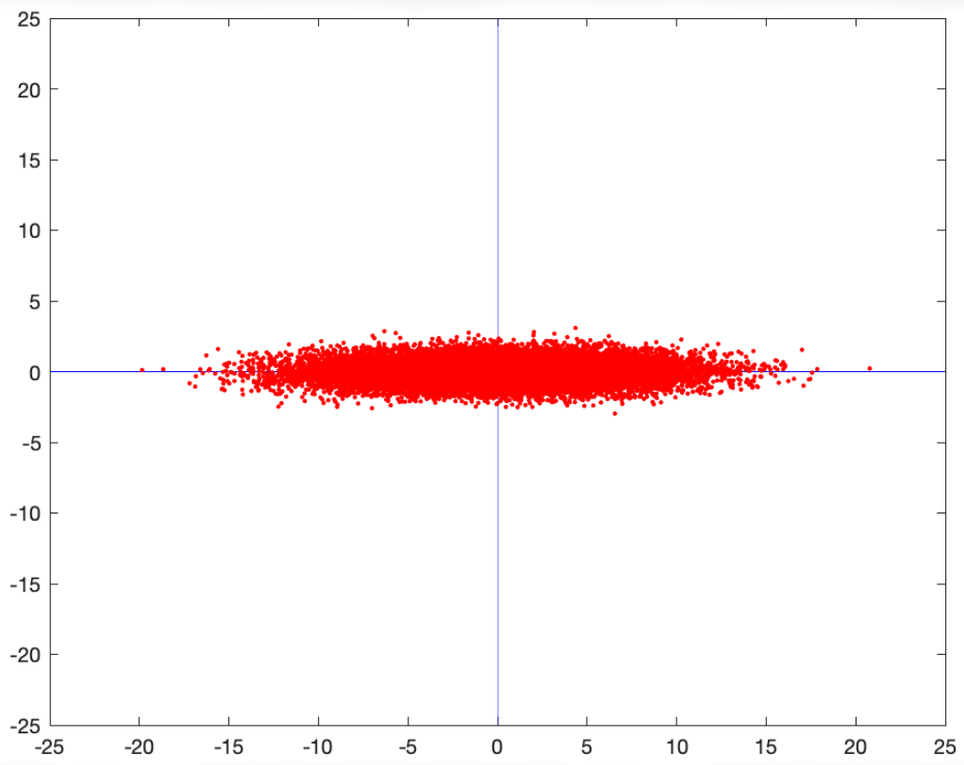
Geometric View

- Standard deviation forms ellipses
- Fit a line
  - Min residual
  - Max variance



PCA = Finding the Best Directions





Algebraic View

- A number of variables (sensors, items, etc)
- Set of data measurements
- Center data
- Calculate variance and covariance

	Var 1	Var 2	Var 3
Meas 1	A_{11}	A_{12}	A_{13}
Meas 2	A_{21}	A_{22}	A_{23}
Meas 3	A_{31}	A_{32}	A_{33}
Meas 4	A_{41}	A_{42}	A_{43}

Calculating Variance and Covariance

- Variance
  - $\sigma_i^2 = \frac{1}{n-1} A_i^T A_i$
- Covariance
  - $\sigma_{ij}^2 = \frac{1}{n-1} A_i^T A_j$
- Combined
  - $C = \frac{1}{n-1} A^T A$

Eigen Decomposition

- $Cx = \lambda x$

Eigen Decomposition

- $Cx = \lambda x$
- $CX = X\Lambda$
- $C = X\Lambda X^{-1}$
- $C = X\Lambda X^{T}$

Better way to calculate

- SVD
- $A = U\Sigma V^T$
- $C = \frac{1}{n-1} A^T A$
- $C = \frac{1}{n-1} V\Sigma^2 V^T$