Principal Component Analysis

Linear Least Squares Approximation

Pierre Alliez
Inria Sophia Antipolis

Definition (point set case)

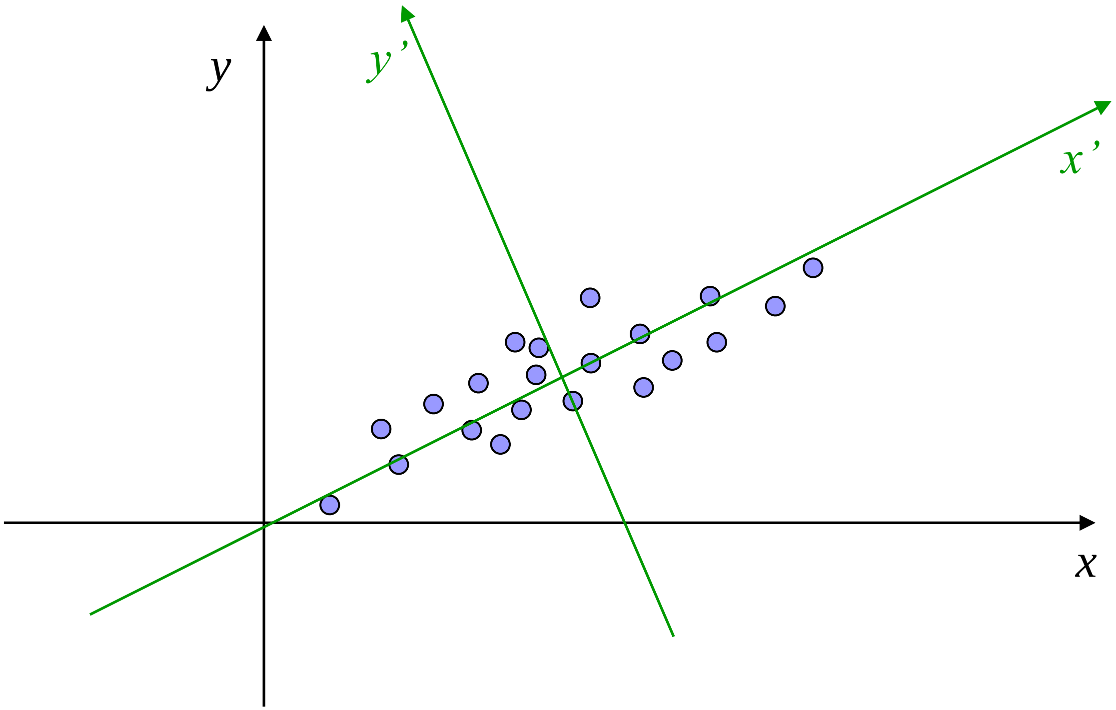
- Given a point set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in R^d$, linear least squares fitting amounts to find the linear sub-space of R^d which **minimizes the sum of squared distances** from the points to their projection onto this linear sub-space.

Definition (point set case)

- This problem is equivalent to search for the linear sub-space which **maximizes the variance of projected points**, the latter being obtained by eigen decomposition of the covariance (scatter) matrix.
- Eigenvectors corresponding to large eigenvalues are the directions in which the data has **strong component**, or equivalently large variance. If eigenvalues are the same there is no preferable sub-space.

PCA – the general idea

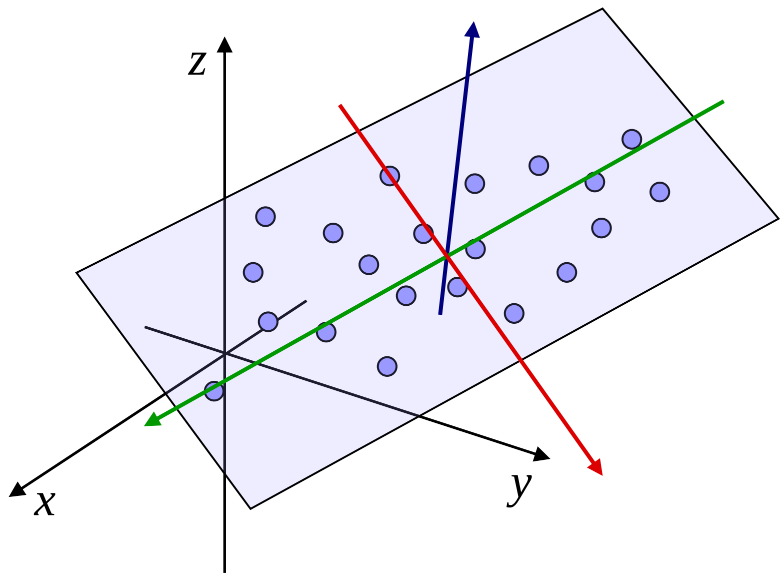
- PCA finds an orthogonal basis that best represents given data set.



- The sum of distances^2 from the x' axis is minimized.

PCA – the general idea

- PCA finds an orthogonal basis that best represents given data set.



3D point set

- PCA finds a best approximating plane (again, in terms of $\Sigma distances^2$)

Notations

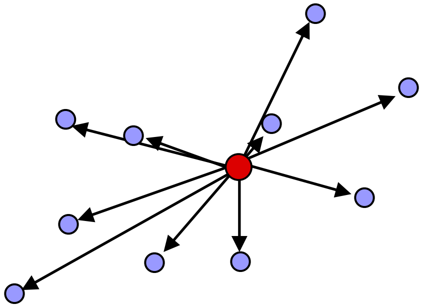
- Denote our data points by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathrm{n} \in R^d$

$$\mathbf{x_1} = \begin{pmatrix} x_1^1 \\ x_1^2 \\ \vdots \\ x_1^d \end{pmatrix}, \ \mathbf{x_2} = \begin{pmatrix} x_2^1 \\ x_2^2 \\ \vdots \\ x_2^d \end{pmatrix}, \ \ldots, \ \mathbf{x_d} = \begin{pmatrix} x_n^1 \\ x_n^2 \\ \vdots \\ x_n^d \end{pmatrix}$$

The origin of the new axes

- The origin is zero-order approximation of our data set (a point)
- It will be the center of mass:

$$\boxed{\mathbf{m} = \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i}$$



- It can be shown that:

$$\boxed{\mathbf{m} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\|\mathbf{x}_i - \mathbf{x}\right\|^2}$$

Scatter matrix

- Denote $\mathbf{y}_i = \mathbf{x}_i - \mathbf{m}, \quad i = 1, 2, \ldots, n$

$$\boxed{S = YY^T}$$

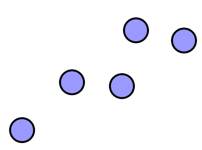
where Y is $d \times n$ matrix with $\mathbf{y}_k$ as columns ($k = 1, 2, \ldots, n$)

$$S = \underbrace{\begin{pmatrix} y_1^1 & y_2^1 & \cdots & y_n^1 \\ y_1^2 & y_2^2 & & y_n^2 \\ \vdots & \vdots & & \vdots \\ y_1^d & y_2^d & \cdots & y_n^d \end{pmatrix}}_{\mathrm{Y}} \underbrace{\begin{pmatrix} y_1^1 & y_1^2 & \cdots & y_1^d \\ y_2^1 & y_2^2 & \cdots & y_2^d \\ \vdots & & & \vdots \\ y_n^1 & y_n^2 & \cdots & y_n^d \end{pmatrix}}_{\mathrm{Y}^T}$$

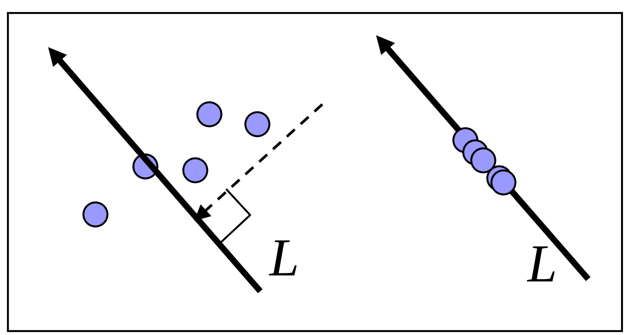
Variance of projected points

- In a way, S measures variance (= scatterness) of the data in different directions.
- Let’s look at a line L through the center of mass $\mathbf{m}$, and project our points $\mathbf{x}_i$ onto it. The variance of the projected points $\mathbf{x}'_i$ is:

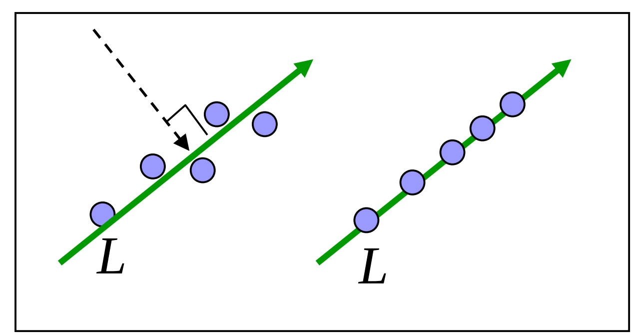
$$\mathrm{var}(L) = \frac{1}{n}\sum_{i=1}^{n} \| \mathbf{x}'_i - \mathbf{m} \|^2$$



Original set



Small variance

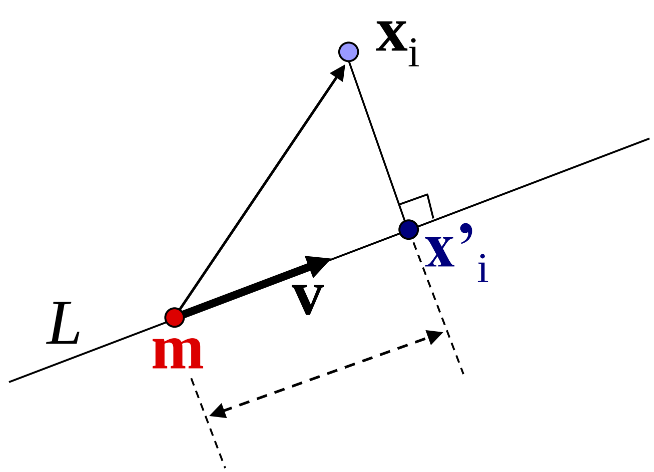


Large variance

Variance of projected points

- Given a direction $\mathbf{v}$, $\|\mathbf{v}\| = 1$, the projection of $\mathbf{x}_\mathrm{i}$ onto $L = \mathbf{m} + \mathbf{v}t$ is:

$$\boxed{\|\mathbf{x}_i' - \mathbf{m}\| = <\mathbf{v},\ \mathbf{x_i} - \mathbf{m}> \|\mathbf{v}\| = <\mathbf{v},\ \mathbf{y_i}> = \mathbf{v}^T\mathbf{y_i}}$$



Variance of projected points

- So,

$$\mathrm{var}(L) = \tfrac{1}{n}\sum_{i=1}^{n} \| \mathbf{x}_i' - \mathbf{m} \|^2 = \tfrac{1}{n}\sum_{i=1}^{n} (\mathbf{v}^\mathbf{T}\mathbf{y}_\mathbf{i})^2 = \tfrac{1}{n} \| \mathbf{v}^\mathbf{T} Y \|^2 =$$

$$= \tfrac{1}{n} \| Y^T \mathbf{v} \|^2 = \tfrac{1}{n} < Y^T\mathbf{v},\ Y^\mathbf{T}\mathbf{v} > = \tfrac{1}{n} \mathbf{v}^\mathbf{T} Y Y^T \mathbf{v} = \tfrac{1}{n} \mathbf{v}^\mathbf{T} S \mathbf{v} = \tfrac{1}{n} < S\mathbf{v},\ \mathbf{v} >$$

$$\sum_{i=1}^{n} (\mathbf{v}^T\mathbf{y}_\mathbf{i})^2 = \sum_{i=1}^{n} \left(\begin{pmatrix} v^1 & v^2 & \cdots & v^d \end{pmatrix} \begin{pmatrix} y_i^1 \\ y_i^2 \\ \vdots \\ y_i^d \end{pmatrix} \right)^2 = \left\| \begin{pmatrix} v^1 & v^2 & \cdots & v^d \end{pmatrix} \begin{pmatrix} y_1^1 & y_2^1 & \cdots & y_n^1 \\ y_1^2 & y_2^2 & & y_n^2 \\ \vdots & \vdots & & \vdots \\ y_1^d & y_2^d & \cdots & y_n^d \end{pmatrix} \right\|^2 = \| \mathbf{v}^\mathbf{T} Y \|^2$$

Directions of maximal variance

- So, we have: $\mathrm{var}(L) = <S\mathbf{v}, \mathbf{v}>$
- Theorem:

  Let $f : \{\mathbf{v} \in R^d \mid \|\mathbf{v}\| = 1\} \to R$,

  $f(\mathbf{v}) = <S\mathbf{v}, \mathbf{v}>$ (and S is a symmetric matrix).

  Then, the extrema of f are attained at the eigenvectors of S.

- So, eigenvectors of S are directions of maximal/minimal variance.

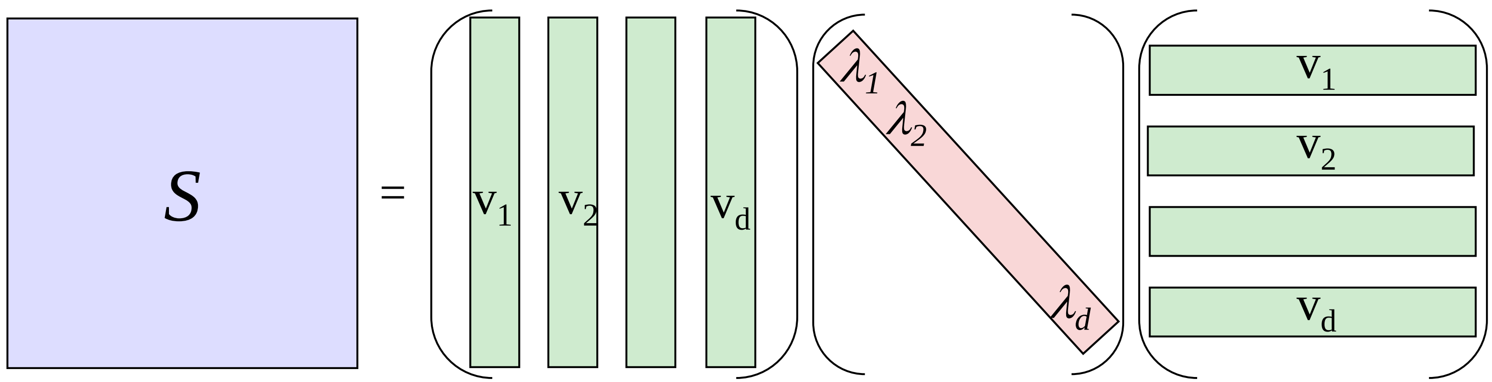
Summary so far

- We take the centered data points $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_\mathrm{n} \in R^d$
- Construct the scatter matrix $S = YY^T$
- S measures the variance of the data points
- Eigenvectors of S are directions of max/min variance.

Scatter matrix - eigendecomposition

- $\boldsymbol{S}$ is symmetric

$\Rightarrow$ $\boldsymbol{S}$ has eigendecomposition: $\boldsymbol{S} = V\Lambda V^{T}$

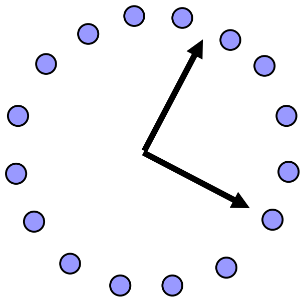


The eigenvectors form orthogonal basis

Principal components

- Eigenvectors that correspond to big eigenvalues are the directions in which the data has strong components (= large variance).
- If the eigenvalues are more or less the same – there is no preferable direction.

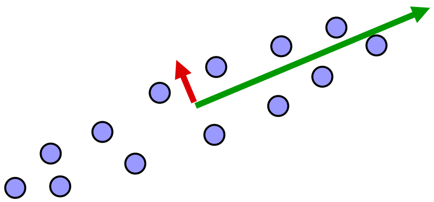
Principal components



- There's no preferable direction
- $\boldsymbol{S}$ looks like this:

$$\begin{pmatrix} \lambda & \\ & \lambda \end{pmatrix}$$

- Any vector is an eigenvector

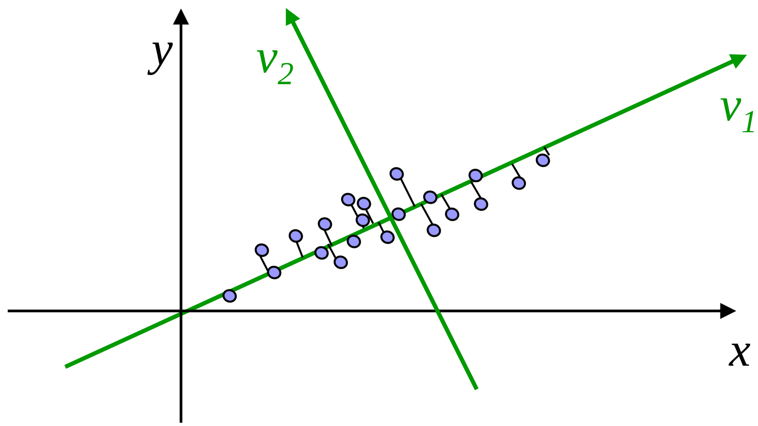


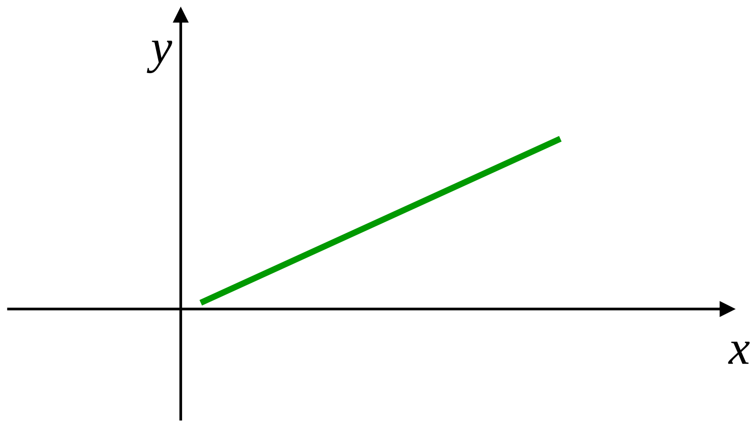
- There is a clear preferable direction
- $\boldsymbol{S}$ looks like this:

$$V\begin{pmatrix} \lambda & \\ & \mu \end{pmatrix}V^T$$

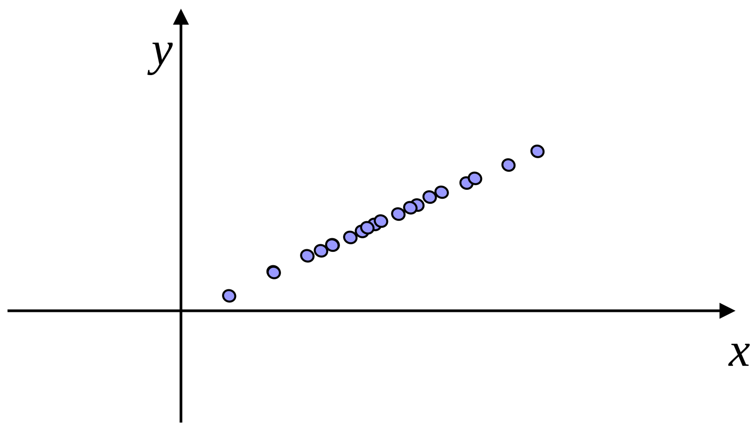
- μ is close to zero, much smaller than λ.

For approximation





This line segment approximates the original data set



The projected data set approximates the original data set

For approximation

- In general dimension d, the eigenvalues are sorted in descending order:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$$

- The eigenvectors are sorted accordingly.
- To get an approximation of dimension $d' < d$, we take the d' first eigenvectors and look at the subspace they span ($d' = 1$ is a line, $d' = 2$ is a plane…)

For approximation

- To get an approximating set, we project the original data points onto the chosen subspace:

$$\mathbf{x}_\mathrm{i} = \mathbf{m} + \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_{d'}\mathbf{v}_{\mathrm{d'}} + \ldots + \alpha_d\mathbf{v}_\mathrm{d}$$

Projection:

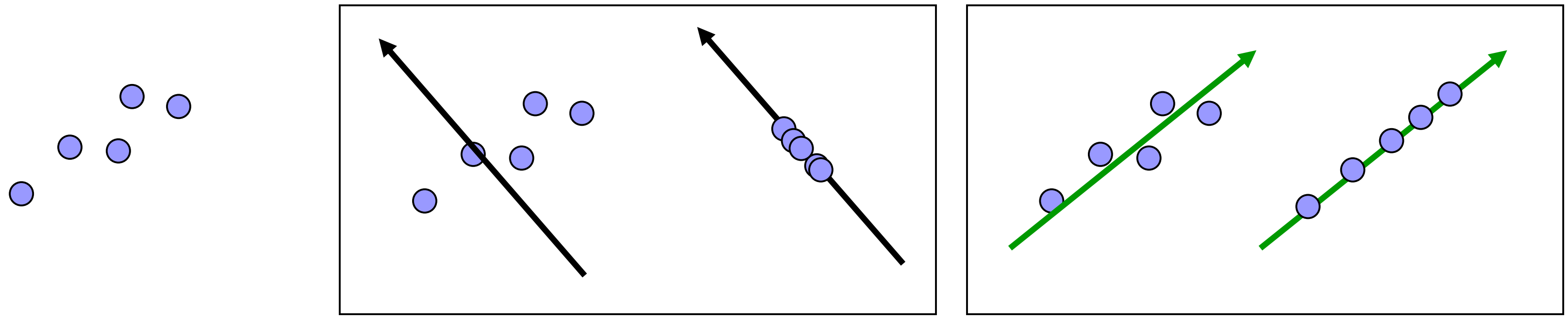
$$\mathbf{x}_\mathrm{i}' = \mathbf{m} + \underbrace{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_{d'}\mathbf{v}_{\mathrm{d'}}} + 0\cdot\mathbf{v}_{\mathrm{d'+1}} + \ldots + 0\cdot\mathbf{v}_\mathrm{d}$$

Optimality of approximation

- The approximation is optimal in least-squares sense. It gives the minimal of:

$$\sum_{k=1}^{n} \left\| \mathbf{x}_k - \mathbf{x}'_k \right\|^2$$

- The projected points have maximal variance.



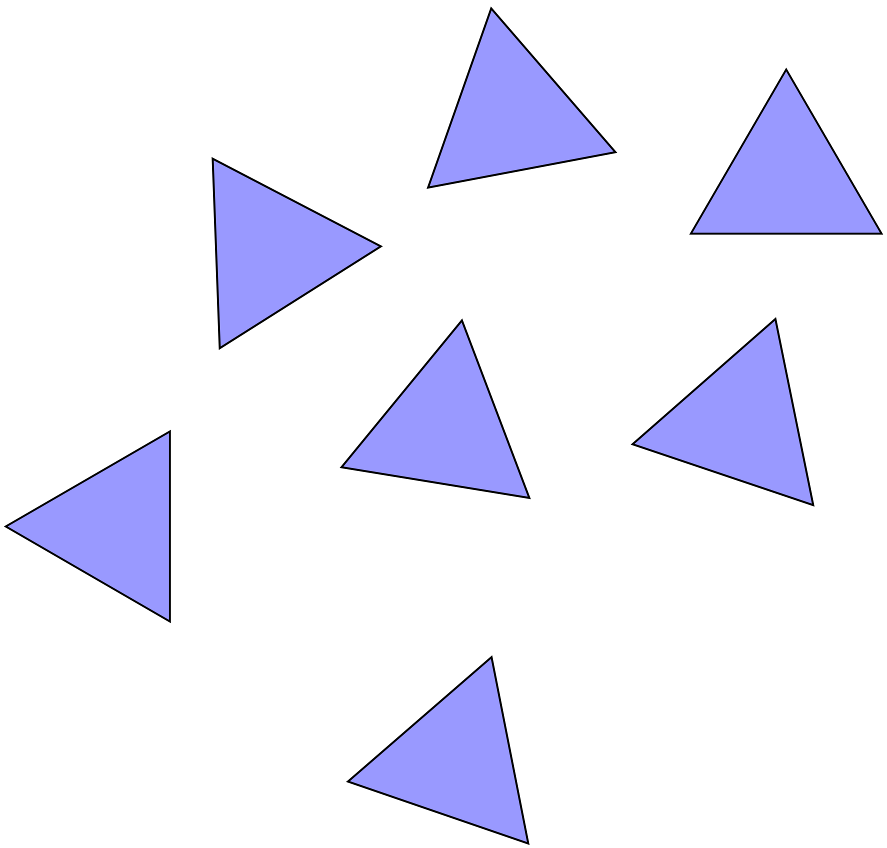
Original set   projection on arbitrary line   projection on $\mathbf{v}_1$ axis

PCA on Point Sets

$$S = \begin{pmatrix} y_1^1 & y_2^1 & \cdots & y_n^1 \\ y_1^2 & y_2^2 & & y_n^2 \\ \vdots & \vdots & & \vdots \\ y_1^d & y_2^d & \cdots & y_n^d \end{pmatrix} \begin{pmatrix} y_1^1 & y_1^2 & \cdots & y_1^d \\ y_2^1 & y_2^2 & \cdots & y_2^d \\ \vdots & & & \vdots \\ y_n^1 & y_n^2 & \cdots & y_n^d \end{pmatrix}$$

demo

PCA on Geometric Primitives?



$$C_i = \int_{s_i} x x^T dx$$

Coordinate relative to center of mass