ECE 417 Lecture 5: Eigenvectors

Mark Hasegawa-Johnson

9/12/2017

Content

- Linear transforms
- Eigenvectors
- Eigenvalues
- Symmetric matrices
- Symmetric positive definite matrices
- Covariance matrices
- Principal components

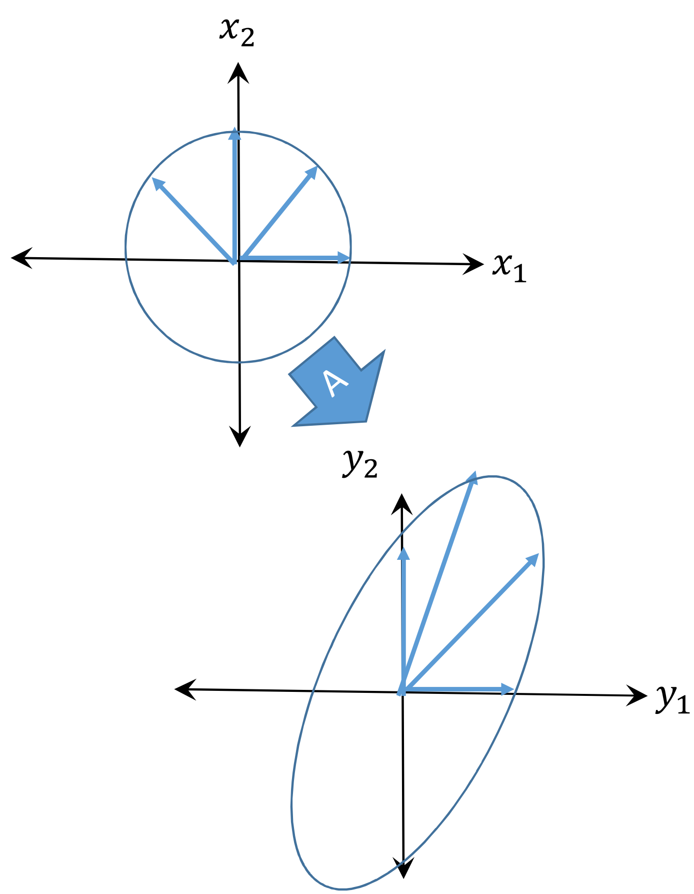
Linear Transforms

A linear transform $\vec{y} = A\vec{x}$ maps vector space $\vec{x}$ onto vector space $\vec{y}$. For example: the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ maps the vectors

$$\overrightarrow{x_1}, \overrightarrow{x_2}, \overrightarrow{x_3}, \overrightarrow{x_4} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

to the vectors

$$\overrightarrow{y_1}, \overrightarrow{y_2}, \overrightarrow{y_3}, \overrightarrow{y_4} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix}$$



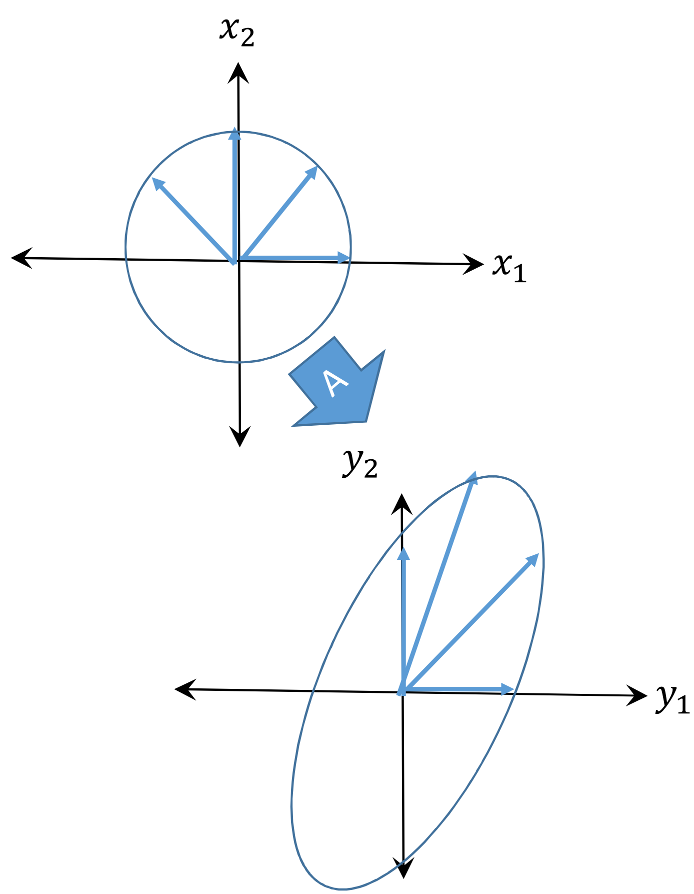
Linear Transforms

A linear transform $\vec{y} = A\vec{x}$ maps vector space $\vec{x}$ onto vector space $\vec{y}$. For example: the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ maps the vectors

$$X = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

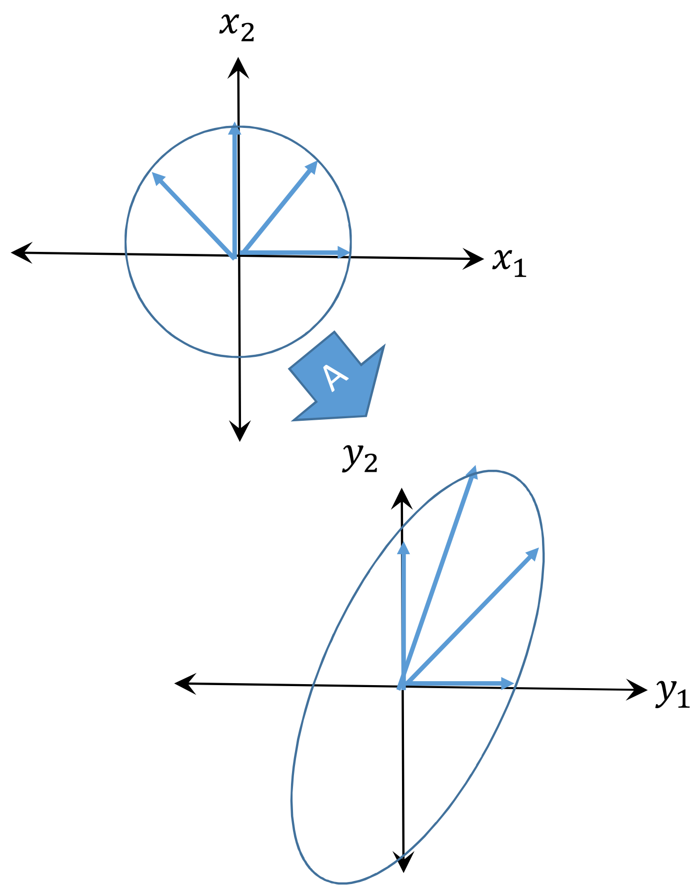
to the vectors

$$Y = \begin{bmatrix} 1 & \sqrt{2} & 1 & 0 \\ 0 & \sqrt{2} & 2 & \sqrt{2} \end{bmatrix}$$



Linear Transforms

A linear transform $\vec{y} = A\vec{x}$ maps vector space $\vec{x}$ onto vector space $\vec{y}$. The determinant of A tells you how much the area of a unit circle is changed under the transformation. For example: if $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$, then the unit circle in $\vec{x}$ (which has an area of π) is mapped to an ellipse with an area of $\pi|A| = 2\pi$.



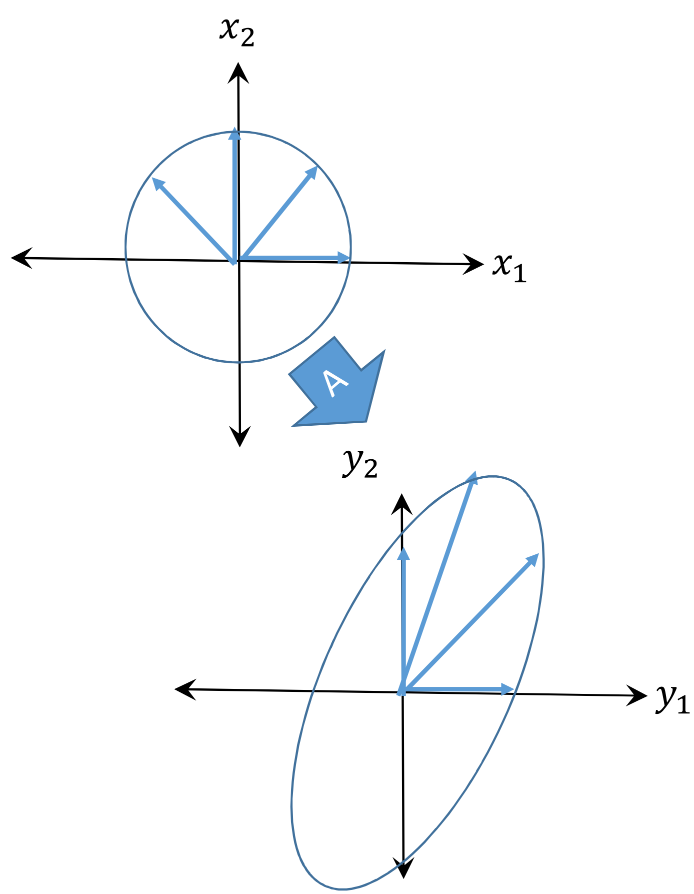
Eigenvectors

- For a D-dimensional square matrix, there may be up to D different directions $\vec{x} = \overrightarrow{v_d}$ such that, for some scalar λ_d,

$$A\overrightarrow{v_d} = \lambda_d \overrightarrow{v_d}$$

- For example: if $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$, then the eigenvectors and eigenvalues are

$$\overrightarrow{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \overrightarrow{v_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \lambda_1 = 1, \lambda_2 = 2$$

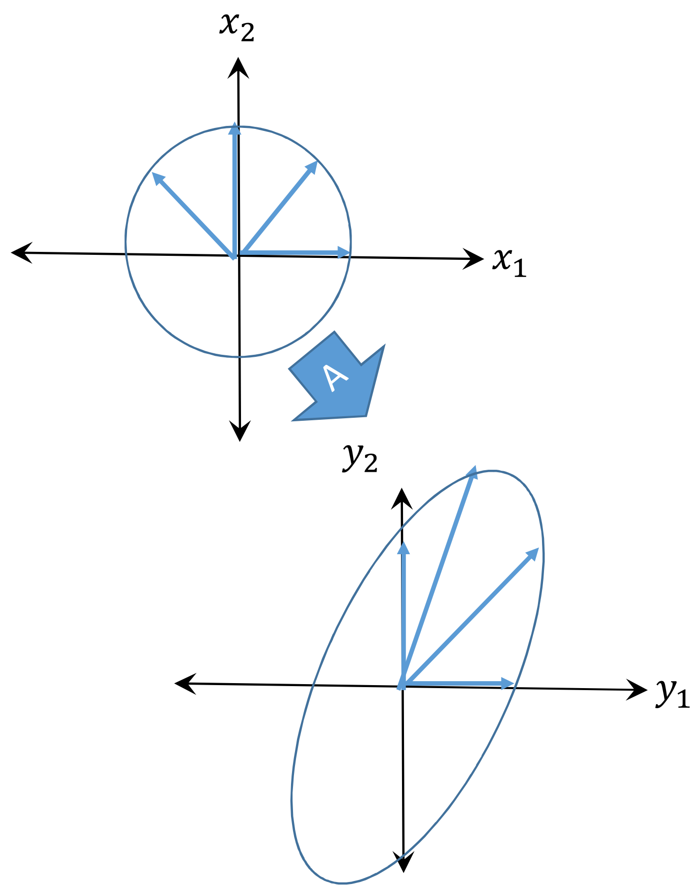


Eigenvectors

- An eigenvector is a direction, not just a vector. That means that if you multiply an eigenvector by any scalar, you get the same eigenvector: if $A\overrightarrow{v_d} = \lambda_d \overrightarrow{v_d}$, then it's also true that $cA\overrightarrow{v_d} = c\lambda_d \overrightarrow{v_d}$
- For example: the following are all the same eigenvector

$$\overrightarrow{v_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \sqrt{2}\overrightarrow{v_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, -\overrightarrow{v_2} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

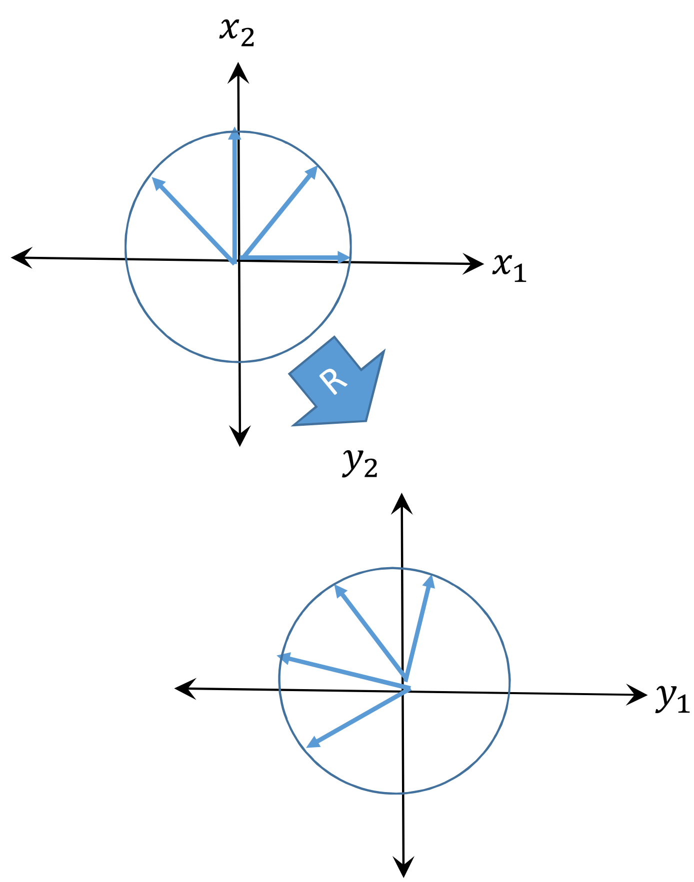
- Since scale doesn't matter, by convention, we normalize so that $\|\overrightarrow{v_d}\|_2 = 1$ and the first nonzero element is positive.



Eigenvectors

- Notice that only square matrices can have eigenvectors. For a non-square matrix, the equation $A\overrightarrow{v_d} = \lambda_d \overrightarrow{v_d}$ is impossible --- the dimension of the output is different from the dimension of the input.
- Not all matrices have eigenvectors! For example, a rotation matrix doesn't have any real-valued eigenvectors:

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



Eigenvalues

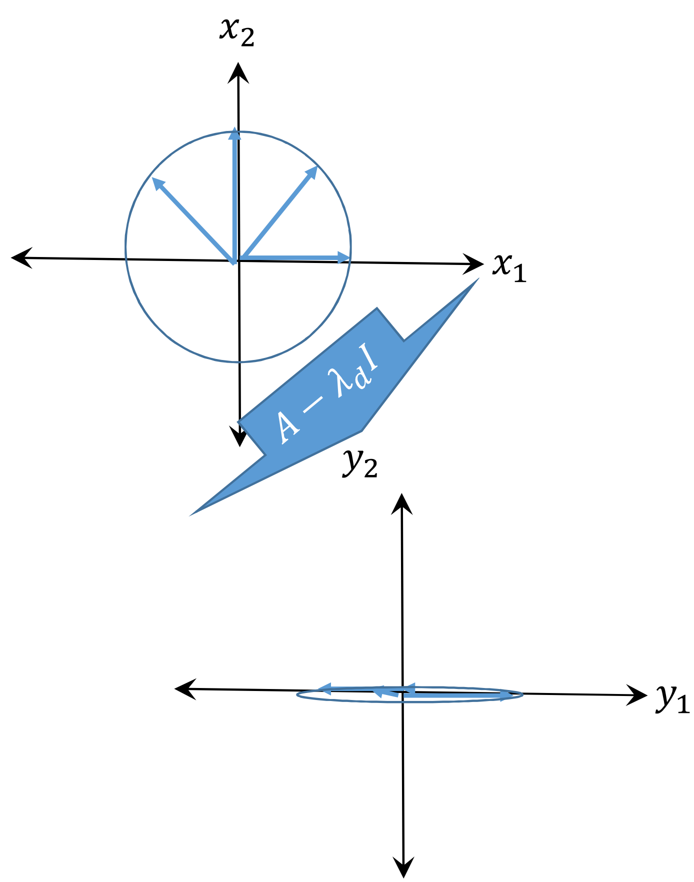
$$A\overrightarrow{v_d} = \lambda_d \overrightarrow{v_d}$$
$$A\overrightarrow{v_d} = \lambda_d I\ \overrightarrow{v_d}$$
$$A\overrightarrow{v_d} - \lambda_d I\ \overrightarrow{v_d} = \vec{0}$$
$$(A - \lambda_d I)\ \overrightarrow{v_d} = \vec{0}$$

That means that when you use the linear transform $(A - \lambda_d I)$ to transform the unit circle, the result has zero area. Remember that the area of the output is $\pi|A - \lambda_d I|$. So that means that, for any eigenvalue λ_d, the determinant of the matrix difference is zero:

$$|A - \lambda_d I| = 0$$

Example:

$$A - \lambda_2 I = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} - 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$$



Eigenvalues

Let's talk about that equation, $|A - \lambda_d I| = 0$. Remember how the determinant is calculated, for example if

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \text{ then } |A - \lambda I| = 0 \text{ means that}$$

$$0 = |A - \lambda I| = \begin{vmatrix} a-\lambda & b & c \\ d & e-\lambda & f \\ g & h & i-\lambda \end{vmatrix} =$$

$$(a-\lambda)(e-\lambda)\,(i-\lambda) - b(d(i-\lambda) - gf) + c\big(dh - g(e-\lambda)\big)$$

- We assume that $a, b, c, d, e, f, g, h, i$ are all given in the problem statement. Only λ is unknown. So the equation $|A - \lambda I| = 0$ is a D'th order polynomial in one variable.
- The fundamental theorem of algebra says that a D'th order polynomial has D roots (counting repeated roots and complex roots).

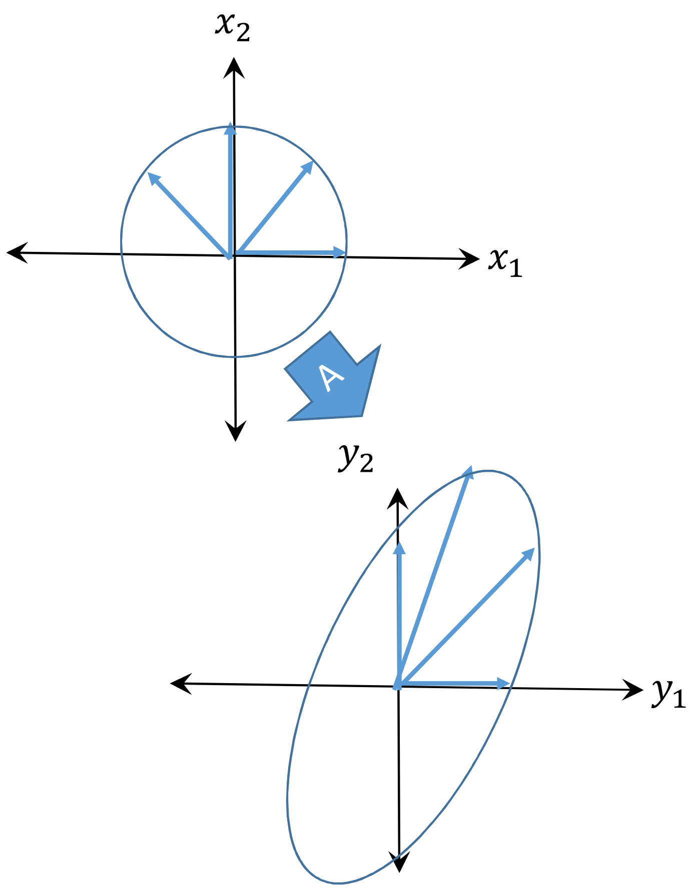
Eigenvalues

So a DxD matrix always has D eigenvalues (counting complex and repeated eigenvalues). This is true even if the matrix has no eigenvectors!! The eigenvalues are the D solutions of the polynomial equation

$$|A - \lambda_d I| = 0$$

Positive Definite Matrix

- A linear transform $\vec{y} = A\vec{x}$ is called "positive definite" (written $A \succ 0$) if, for any vector $\vec{x}$,
$$\vec{x}^T A\vec{x} > 0$$
- So, you can see that this means $\vec{x}^T\vec{y} > 0$.
- So this means that a matrix is positive definite if and only if the output of the transform, $\vec{y}$, is never rotated away from the input, $\vec{x}$, by 90 degrees or more! ← (useful geometric intuition)
- For example, the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ is positive-definite.



Symmetric matrices

We've been working with "right eigenvectors:"

$$A\overrightarrow{v_d} = \lambda_d \overrightarrow{v_d}$$

There may also be left eigenvectors, which are row vectors $\vec{u}_d^T$, and corresponding left eigenvalues μ_d:

$$\vec{u}_d^T A = \mu_d \vec{u}_d^T$$

If A is symmetric ($A = A^T$), then the left and right eigenvectors and eigenvalues are the same, because

$$\lambda_d \vec{v}_d^T = (\lambda_d \overrightarrow{v_d})^T = (A\overrightarrow{v_d})^T = \vec{v}_d^T A^T = \vec{v}_d^T A = \mu_d \vec{u}_d^T$$

Symmetric positive definite matrices

If A is symmetric ($A = A^T$), then you can do an interesting thing if you multiply the matrix by its eigenvectors both before and after:

$$\vec{v}_d^T A \vec{v}_d = \vec{v}_d^T (\lambda_d \vec{v}_d) = \lambda_d ||\vec{v}_d||_2^2 = \lambda_d$$

So if a symmetric matrix is positive definite, then all of its eigenvalues are positive real numbers. It turns out that the opposite is also true:

A symmetric matrix is positive definite if and only if all of its eigenvalues are positive.

Symmetric positive definite matrices

Symmetric positive definite matrices turn out to also have one more unbelievably useful property: their eigenvectors are orthogonal.

$$\vec{v}_i^T \vec{v}_j = 0 \text{ if } i \neq j$$

If $i = j$ then, by convention, we have

$$\vec{v}_i^T \vec{v}_i = \|\vec{v}\|_2^2 = 1$$

So suppose we create the matrix

$$V = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_D]$$

This is an orthonormal matrix:

$$V^T V = I$$

It turns out that, also, $VV^T = I$.

Symmetric positive definite matrices

If A is symmetric ($A = A^T$), then

$$\vec{v}_d^T A \vec{v}_d = \vec{v}_d^T (\lambda_d \vec{v}_d) = \lambda_d \|\vec{v}_d\|_2^2 = \lambda_d$$

…but also…

$$\vec{v}_i^T \vec{v}_j = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$$

That means we can write A as

$$A = \sum_{i=1}^{D} \lambda_i \vec{v}_i \vec{v}_i^T = V \Lambda V^T$$

Because

$$\vec{v}_j^T A \vec{v}_j = \sum_{i=1}^{D} \lambda_i \vec{v}_j^T \vec{v}_i \vec{v}_i^T \vec{v}_j = \lambda_j$$

Symmetric positive definite matrices

If A is symmetric and positive definite we can write

$$A = \sum_{i=1}^{D} \lambda_i \vec{v}_i \vec{v}_i^T = V \Lambda V^T$$

Equivalently

$$V^T A V = V^T V \Lambda V^T V = I \Lambda I = \Lambda$$

Covariance matrices

Suppose we have a dataset containing N independent sample vectors, $\vec{x}_n$. The true mean is approximately given by the sample mean,

$$\vec{\mu} = E[\vec{x}] \approx \frac{1}{N}\sum_{n=1}^{N} \vec{x}_n$$

Similarly, the true covariance matrix is approximately given by the sample covariance matrix,

$$\Sigma = E[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T] \approx \frac{1}{N}\sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T$$

Covariance matrices

Define the "sum-of-squares matrix" to be

$$S = \sum_{n=1}^{N} (\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T$$

So that the sample covariance is $\Sigma \approx S/N$. Suppose that we define the centered data matrix to be the following DxN matrix:

$$\tilde{X} = [\vec{x}_1 - \vec{\mu}, \vec{x}_2 - \vec{\mu}, \dots, \vec{x}_N - \vec{\mu}]$$

Then the sum-of-squares matrix is

$$S = \tilde{X}\tilde{X}^T = [\vec{x}_1 - \vec{\mu}, \dots, \vec{x}_N - \vec{\mu}] \begin{bmatrix} (\vec{x}_1 - \vec{\mu})^T \\ \dots \\ (\vec{x}_N - \vec{\mu})^T \end{bmatrix}$$

Covariance matrices

Well, a sum-of-squares matrix is obviously symmetric. It's also almost always positive definite:

$$\vec{x}^T S \vec{x} = [\vec{x}^T(\vec{x}_1 - \vec{\mu}), \dots, \vec{x}^T(\vec{x}_N - \vec{\mu})] \begin{bmatrix} (\vec{x}_1 - \vec{\mu})^T \vec{x} \\ \dots \\ (\vec{x}_N - \vec{\mu})^T \vec{x} \end{bmatrix}$$

That quantity is positive unless the new vector, $\vec{x}$, is orthogonal to $(\vec{x}_n - \vec{\mu})$ for every vector in the training database. As long as $N \geq D$, that's really, really unlikely.

Covariance matrices

So a sum-of-squares matrix can be written as

$$S = \sum_{i=1}^{D} \lambda_i \vec{v}_i \vec{v}_i^T = V \Lambda V^T$$

And the covariance can be written as

$$\Sigma = \frac{S}{N} = \frac{1}{N} \sum_{i=1}^{D} \lambda_i \vec{v}_i \vec{v}_i^T = V \left(\frac{\Lambda}{N} \right) V^T$$

Principal components

Suppose that

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \lambda_D \end{bmatrix}, V = [\vec{v}_1, \dots, \vec{v}_D]$$

are the eigenvalue and eigenvector matrices of S, respectively. Define the principal components of $\vec{x}_n$ to be $y_{dn} = \vec{v}_d^T(\vec{x}_n - \vec{\mu})$, or

$$\vec{y}_n = V^T(\vec{x}_n - \vec{\mu}) = \begin{bmatrix} {\vec{v}_1}^T(\vec{x}_n - \vec{\mu}) \\ \dots \\ {\vec{v}_D}^T(\vec{x}_n - \vec{\mu}) \end{bmatrix}$$

Principal components

Suppose that Λ and V are the eigenvalue and eigenvector matrices of S, respectively. Define the principal components to be $\vec{y}_n = V^T(\vec{x}_n - \vec{\mu})$.

Then the principal components y_{dn} are not correlated with each other, and the variance of each one is given by the corresponding eigenvalue of S.

$$E[\vec{y}\vec{y}^T] \approx \frac{1}{N}\sum_{n=1}^{N} \vec{y}_n{\vec{y}_n}^T = \frac{1}{N}\sum_{n=1}^{N} \begin{bmatrix} y_{1n} \\ \cdots \\ y_{Dn} \end{bmatrix} [y_{1n}, \ldots, y_{Dn}]$$

$$= \frac{1}{N}\sum_{n=1}^{N} V^T(\vec{x}_n - \vec{\mu})(\vec{x}_n - \vec{\mu})^T V$$

$$= V^T S V = \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ldots & 0 \\ 0 & 0 & \lambda_D \end{bmatrix}$$

Summary

- Principal component directions are the eigenvectors of the covariance matrix (or of the sum-of-squares matrix – same directions, because they are just scaled by N)
- Principal components are the projections of each training example onto the principal component directions
- Principal components are uncorrelated with each other: the covariance is zero
- The variance of each principal component is the corresponding eigenvalue of the covariance matrix

Implications

- The total energy in the signal, $E[\|\vec{x}-\vec{\mu}\|_2^2]$, is equal to the sum of the eigenvalues.
- If you want to keep only a small number of dimensions, but keep most of the energy, you can do it by keeping the principal components with the highest corresponding eigenvalues.