Recitation 3-TRANSFORMATIONS (09/24/04)
by Larissa Welti-Santos

References:
Strang, Gilbert. Linear Algebra and Its Applications, 1988, Harcourt.
Fukunaga, Keinosuke. Introduction to Statistical Pattern Recognition, 1990, Academic Press.

OUTLINE
EV & EW
EV & EW + covariance matrix
EV & EV + covariance matrix + transformations

EV & EW

Remember the characteristic equation $(\mathbf{M} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$?

λ will be an eigenvalue (EW) of $\mathbf{M}$ with a corresponding eigenvector $\mathbf{x}$ (EV). To find λ we compute $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$.

Things we ask ourselves (and by the way, understanding them is useful)

- Why do we care about $\det(\mathbf{M} - \lambda\mathbf{I})$?
  We care about it because λ will be an EW of $\mathbf{M}$ with nonzero EV if and only if $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$. Which means that $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$ is one necessary and sufficient condition for λ to be an EW.

- Ok, but what makes eigenvectors so unique?
  Well, given a matrix $\mathbf{M}$ the problem is to find those special vectors $\mathbf{x}$ on which $\mathbf{M}$ acts as simple multiplication (i.e. $\mathbf{Mx} = \lambda\mathbf{x}$), which means that $\mathbf{Mx}$ points in the same direction as $\mathbf{x}$.
  The determination of the eigenvectors and eigenvalues of a system is equivalent to matrix diagonalization. Therefore, if we can find a linear transformation to diagonalize the covariance matrix we can obtain uncorrelated random variables in general and independent random variables for normal distributions.

Aha! We know that a couple of things are easier and have nice properties with uncorrelated random variables and with independent random variables. In general, if we have a sample that comes from a normal distribution, variables are correlated and it often helps to view the data in a different way, for example in a new coordinate system.

EV & EW + covariance matrix

Let's start with understanding normal distributions.

Normal distributions have many nice and important properties such as

- Two uncorrelated normally distributed random variables are independent -- remember that this is generally not true for other random variables.

- A normal distribution is completely defined by its mean vector and its covariance matrix -- that is why we are concerned about how the mean and the variance changes under specific transformations.

- The marginal densities and the conditional densities are also normal

Why do we care about covariance matrices?

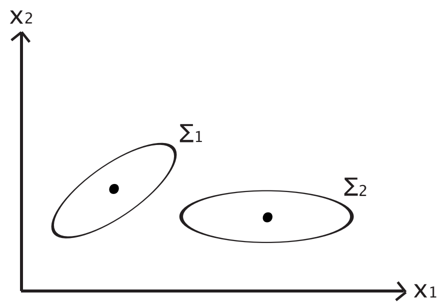
Since a diagonal covariance matrix means uncorrelated variables and particularly independent variables for a normal distribution, a "transformation" will allow us to always -- always happens just because of a property of the normal distribution (check Problem Set 2) -- find a set of axes such that random variables are independent in the new coordinate system.

EV & EW + covariance matrix + transformations

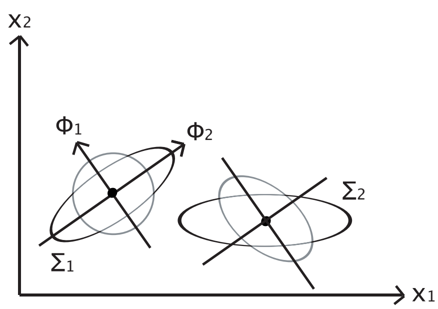
So, what are those transformations that will yield a "nice looking" normal distribution?

1. Linear transformation : to zero-mean – this is specially helpful because the covariance will be equal to the correlation.
2. Orthonormal transformation : to decorrelate -- orthonormal matrices have energy preserving characteristics.

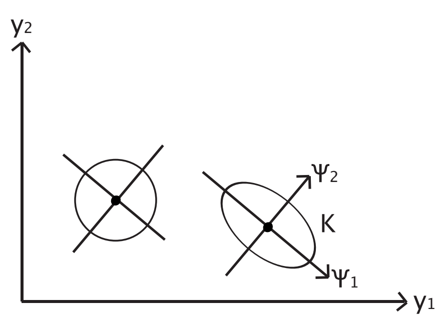
3. Whitening transformation : to scale the EV in proportion to $1/\sqrt{\lambda_i}$ (i.e. make the covariance matrix equal to the identity matrix (**I**)). We do it because after applying the whitening transformation, the covariance matrix is invariant under any orthonormal transformation.



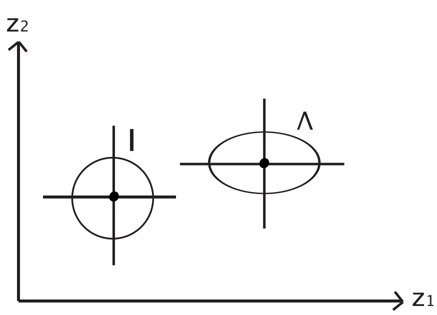
We start with two symmetric matrices Σ_1 and Σ_2.



1. We get the eigenvector and eigenvalue matrices of Σ_1, Θ and Φ respectively;
2. We whiten Σ_1 by $Y=\Theta^{\frac{1}{2}}\Phi^T\mathbf{x}$;
3. Σ_1 and Σ_2 are transformed to I and K . In general K is not a diagonal matrix.



4. We get Ψ and Λ, the eigenvector and eigenvalue matrices of K.



5. We apply an orthonormal transformation to diagonalize K by $Z=\Psi^T\mathbf{y}$;
6. I is invariant under this transformation.
7. Now both matrices are diagonalized.