WHY IS DIMENSIONAL REDUCTION USEFUL?

Example 1: PSTH

choice of $\Delta t$ determines space in which data represented
WHY IS DIMENSIONAL REDUCTION USEFUL?: Example 2: 2AFC
EXAMPLE 3: VARIATIONS IN SINGLE PHOTON RESPONSE

Can we compactly describe how single photon responses vary? Start: 120 dimensions (time bins) - but clearly not independent.
INGREDIENTS IN DIMENSIONAL REDUCTION

1. Define measure of interesting statistical structure in data set (e.g. variance for PCA)
2. Measure this structure in high dimensional space
3. Identify directions in this space that capture the most structure
4. Limit data to space characterized by those directions or axes
CORRELATION AND COVARIANCE

Correlation

\[ r = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sqrt{\sigma_x^2 \sigma_y^2}} \]

\( r = 0.5 \)
\( c = 0.35 \)

\( r = 0 \)
\( c = 0 \)

\( r = -0.5 \)
\( c = -0.35 \)

Covariance

\[ c = \langle xy \rangle - \langle x \rangle \langle y \rangle \]

Both correlation and covariance measure dependence of \( x \) on \( y \), but correlation normalized and covariance is not.
THE COVARIANCE MATRIX

\[ C = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{xy} & c_{yy} \end{bmatrix} \]

- **x and y independent**
  \[ C = \begin{bmatrix} 0.35 & 0 \\ 0 & 1 \end{bmatrix} \]
  covariance diagonal

- **x and y not independent**
  \[ C = \begin{bmatrix} 0.35 & -0.35 \\ -0.35 & 1.35 \end{bmatrix} \]
  covariance has nonzero off diagonal elements
THE MULTIVARIATE GAUSSIAN DISTRIBUTION

The covariance matrix \( C \) and the mean \( \bar{X} \) completely specify a Gaussian distribution.

\[
P(X) = Z \exp \left( -\frac{1}{2} (X - \bar{X}) C^{-1} (X - X)^T \right)
\]

The covariance matrix \( C \) and the mean \( \bar{X} \) completely specify a Gaussian
covariance can be used to generate coordinate system in which data symmetrical

- axes uncorrelated (not necessarily independent though)
- axes normalized by standard deviation

\[
x' = (\sqrt{C})^{-1} x
\]  

\[
x = \sqrt{C} x'
\]
WHY $\sqrt{C}$?

Start with simpler case: x, y uncorrelated

$$C = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$

$$C^{-1} = \begin{bmatrix} 1/\sigma_x^2 & 0 \\ 0 & 1/\sigma_y^2 \end{bmatrix}$$

What do we do more generally?
EIGENVECTORS DIAGONALIZE COVARIANCE MATRIX

x-y coordinates

\[ C = \begin{bmatrix} 0.09 & -0.14 \\ -0.14 & 1.18 \end{bmatrix} \]

e_1-e_2 coordinates

\[ C = \begin{bmatrix} 0.07 & 0 \\ 0 & 1.2 \end{bmatrix} \]

covariance matrix is diagonal in eigenvector coordinates
EIGENVECTOR DECOMPOSITION

• The eigenvalue equation:
  \[ M e_n = \lambda_n e_n \]

• Can rewrite any symmetric matrix (e.g. the covariance matrix)
  \[ M = E^{-1} \Lambda E, \]
  where the rows of \( E \) are the eigenvectors \( e \) and \( \Lambda \) is a diagonal matrix of eigenvalues:

  \[
  \Lambda = \begin{bmatrix}
  \lambda_1 & 0 & 0 \\
  0 & \lambda_2 & 0 \\
  0 & 0 & \lambda_3 \\
  \end{bmatrix}
  \]

• Eigenvectors of a symmetric matrix form axes of an orthogonal coordinate system

• \( E \) projects target vector into eigenvector space, \( \Lambda \) scales components of vector, \( E^{-1} \) projects result back into original space.
INGREDIENTS IN DIMENSIONAL REDUCTION

1. Define measure of interesting statistical structure in data set (e.g. variance for PCA)

2. Measure this structure in high dimensional space

3. Identify directions in this space that capture the most structure

4. Limit data to space characterized by those directions or axes
Consider covariance matrix in eigenvector space

- from definition of eigensystem:

\[ C e_n = \lambda_n e_n \rightarrow C = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \]

- from definition of covariance matrix:

\[ C = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \]

- eigenvalues are equal to variance of projections along corresponding eigenvector

- total variance in data is given by sum of eigenvalues
EXAMPLE: VARIATIONS IN SINGLE PHOTON RESPONSE

- singles
- failures

2 pA

0.0  0.5  1.0
sec

current (pA)

0  0.5  1.0
0  0.5  1.0

0  0.5  1.0

0  0.5  1.0

variance (pA^2)
EXAMPLE: VARIATIONS IN SINGLE PHOTON RESPONSE

\[ \frac{\lambda_1}{\sum_n \lambda_n} = 0.75 \]
\[ \frac{\lambda_2}{\sum_n \lambda_n} = 0.15 \]
\[ \frac{\lambda_3}{\sum_n \lambda_n} = 0.07 \]

first 3 (of 120) modes capture 97% of variance
INGREDIENTS IN DIMENSIONAL REDUCTION

1. Define measure of interesting statistical structure in data set (e.g. variance for PCA)

2. Measure this structure in high dimensional space

3. Identify directions in this space that capture the most structure

4. Limit data to space characterized by those directions or axes
WHERE ARE WE?

- covariance matrix describes how different variables depend on each other (i.e. how much they covary)
- eigenvectors diagonalize covariance matrix
- eigenvalues describe variance of data projected along each eigenvector

ranking axes in order of importance using eigenvalues provides systematic way of eliminating ‘unimportant’ dimensions
SOME USES OF PCA

• identify consistent statistical structure in data set
  - spike sorting
  - constraints on mechanistic models for responses
  - stimulus features leading to spiking

• provide compact, efficient description of responses
  - discrimination with good generalization
  - effective generative models of responses
LIMITATIONS OF PCA

PCA considers only variance - depending on data higher order statistical properties may matter (ICA)

for this very non-Gaussian distribution eigenvectors of covariance do not capture ‘appropriate’ directions (though the directions captured may still be practically useful)

PCA considers only variance - depending on data higher order statistical properties may matter (ICA)
VARIATIONS IN SINGLE PHOTON RESPONSE, TAKE 2
(OR, THE DEVIL IS IN THE DETAILS)
WHAT COVARIANCE DO WE USE?

\[ C = \text{cov}(\text{Singles}) \]

\[ C = \text{cov}(\text{Singles}) - \text{cov}(\text{Failures}) \]

guarantee that eigenvalues positive applies only to true covariance matrix - not difference covariance
WHAT COVARIANCE DO WE USE?

\[ C = \text{cov}(\text{Singles}) \]

\[ C = \text{cov}(\text{Singles}) - \text{cov}(\text{Failures}) \]

- none of modes looks like average single photon response - because mean subtracted in constructing covariance
- subtracting covariance of failures eliminates structure prior to flash - a good thing!
HOW MANY COMPONENTS DO WE KEEP?

- Component variance
- 0 singles
- * failures

largest 3 components capture >97% of excess variance of singles
MODELING FLUCTUATIONS IN SINGLE PHOTON RESPONSES

\[ I(t) = \sum_{n} c_n e_n(t) \]

Expansion components

- \( e_0(t) \)
- \( e_1(t) \)

Expansion coefficients

- \( c_0 \)
- \( c_1 \)
A PARTICULARLY IMPORTANT EXAMPLE: THE FOURIER TRANSFORM

- Project \( x(t) \) into linear basis defined by \( \exp(i\omega_i t) \)

\[
x(t) = \sum c_i \exp(i\omega_i t)
\]

where the coefficients \( c_i \) are the projections

\[
c_i = \int dt \ x(t) \exp(i\omega_i t)
\]

- For translationally invariant data, the eigenvectors of the covariance matrix are Fourier components
- i.e. the Fourier modes \( \exp(i\omega_i t) \) diagonalize the time representation.

- The eigenvalues are \( c_i^2 \) (i.e. the power at \( \omega_i \))

- The variance is \( \sigma_x^2 = \sum \sigma_i^2 \)
EIGENVECTORS OF FAILURES (NOISE) ARE SINUSOIDS

\[ c_i = \int dt \, x(t) \exp(i\omega_i t) \]

\[ C = \begin{bmatrix} c_1^2 & 0 & \ldots & 0 \\ 0 & c_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & c_n^2 \end{bmatrix} \]