Lecture 2

Minimum Norm Solutions

Solving Linear Systems

Consider solving a linear system of equations in the form:

$$Xw = y$$

$X \in \mathbb{R}^{nxd}$

Data vector
n data points
d dimensions

$\mathbf{w}$: weights/coefficients
$\mathbf{y}$: observations/labels

With training data $\mathbf{X}_{\mathbf{train}}$ and $\mathbf{y}_{\mathbf{train}}$

You want to come up with a linear equation of the form $\hat{\mathbf{y}} = \mathbf{x}^{\mathbf{T}}\mathbf{w}$
Such that the error or residual

$$||Xw - y||_2^2$$

is minimized

4 Scenerios for Linear Equations

There are 4 scenarios in which we can determine a solution for w:

1. X is square ($n = d$) and full rank
2. X is tall or over determined ($n > d$) and full rank.
3. X is wide or under determined ($n < d$) and full rank.
4. X is rectangular ($n \neq d$) and not full rank.

Underdetermined Equations and Minimum norm

For an underdetermined system of linear equations (scenario 3), our problem $\mathbf{X}\mathbf{w} = \mathbf{y}$ has infinite solutions.

In such a scenario, we are interested in the minimum norm solution.
As such, we can re-formulate our problem as:

$$\min_{w} \|w\| \text{ s.t. } Xw = y$$

Generally, these type of problems are useful for constrained optimization and control problems - which you may see in your engineering courses.

Moore Penrose pseudo-inverse

When solving linear least squares problems in the form $Xw = y$ (X data matrix, w coefficients, y observations to predict). Often, X is not square or directly invertible.

The Moore-Penrose pseudoinverse X^+ gives us a generalization of the inverse matrix that can help us solve linear least squares solutions, $w = X^+y$.

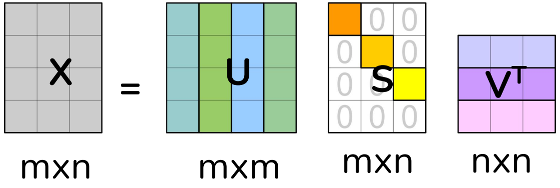
A computationally simple and accurate of computing the pseudoinverse involves using the singular value decomposition. If $X = USV^T$, then the pseudo-inverse would be

$$X^+ = VS^+U^T$$

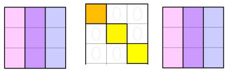
where S^+ of a diagonal matrix of singular values S is obtained by taking the reciprocal of the nonzero diagonal entries, then transposing that matrix $(S^{-1})^T$.

SVD to Pseudo-Inverse

Overdetermined Matrix X

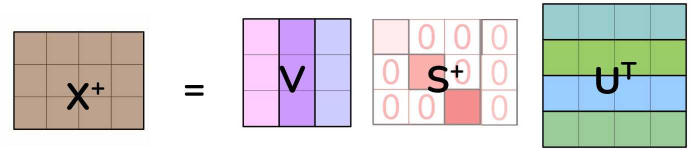


$X^TX = (USV^T)^T(USV^T) = VS^2V^T$



$(X^TX)^{-1} = VS^{-2}V^T$

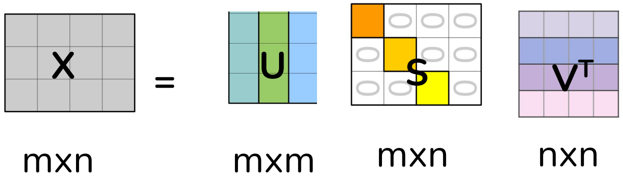
$X^+ = (X^TX)^{-1}X^T = VS^{-2}V^T(VS^TU^T) = VS^+U^T$



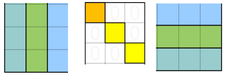
$X^+X = VV^T = I$ so $(X^TX)^{-1}X^T$ is left inverse of X.

$X(X^TX)^{-1}X^T$ gives projection onto Range(X)

Underdetermined Matrix X

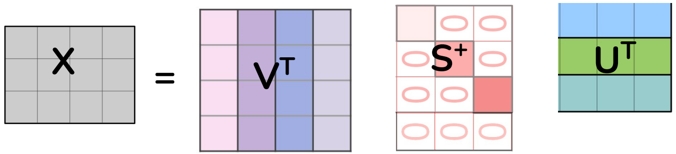


$XX^T = (USV^T)(USV^T)^T = US^2U^T$



$(XX^T)^{-1} = US^{-2}U^T$

$X^+ = X^T(XX^T)^{-1} = (VS^TU^T)US^{-2}U^T = VS^+U^T$



$XX^+ = UU^T = I$ so $X^T(XX^T)^{-1}$ is right inverse of X.

$I - X^T(XX^T)^{-1}X$ gives projection onto Null(X)

4 Scenerio: Block Form of Pseudoinverse

	row m, col n, rank r	$\boldsymbol{X}^{+}$	=	$\boldsymbol{V}$	$\boldsymbol{S}^{+}$	$\boldsymbol{U}$
1.	$m = n = r$	$\boldsymbol{X}^{-1}$	=	$\begin{bmatrix} \boldsymbol{V}_R \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}^{-1} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{U}_R^T \end{bmatrix}$
2.	$m > n, n = r$	$\boldsymbol{X}^{+}$ $\boldsymbol{X}^{-L}$	=	$\begin{bmatrix} \boldsymbol{V}_R \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}^{-1} & 0 \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{U}_R^T \\ \boldsymbol{U}_N^T \end{bmatrix}$
3.	$m = r, m < n$	$\boldsymbol{X}^{+}$ $\boldsymbol{X}^{-R}$	=	$\begin{bmatrix} \boldsymbol{V}_R & \boldsymbol{V}_N \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}^{-1} \\ 0 \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{U}_R^T \end{bmatrix}$
4.	$m \neq n \neq r$	$\boldsymbol{X}^{+}$	=	$\begin{bmatrix} \boldsymbol{V}_R & \boldsymbol{V}_N \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{U}_R^T \\ \boldsymbol{U}_N^T \end{bmatrix}$

Moore Penrose Inverse Gives Us Minimum Norm for Underdetermined Systems

All of our solutions to this problem will have the form $\{w|Xw = y\} = \{w_r + w_n|w_r \in R(X), w_n \in N(X)\}$, where $R(X)$ is the range of the columns of X and $N(X)$ is the null space of the columns of X.

The least norm solution will be:

$$w = X^T(XX^T)^{-1}y$$

To see that this is the least norm solution, let us consider another solution z, so we have $y = Xz = Xw$. Then $X(z - w) = 0$. Then we see

$$(z - w)^T w = (z - w)^T X^T (XX^T)^{-1} y = (X(z - w))^T (XX^T)^{-1} y = 0$$

So

$$\|z\|^2 = \|z + w - w\|^2 = \|w\|^2 + \|z - w\|^2 \geq \|w\|^2$$

Note that cross terms cancel. From this, we see that our least norm solution has the smallest norm.

Lagrange Multipliers

Our problem is formulated as:

$$\min_{w} \|w\|_2$$
$$\text{s.t.} Xw = y$$

Let us introduce Lagrange Multipliers:

$$L(w, \lambda) = w^T w + \lambda^T (Xw - y)$$

Taking derivatives:

$$\nabla_w L = 2w + X^T \lambda = 0 \Rightarrow w = -\frac{X^T \lambda}{2}$$

$$\nabla_\lambda L = Xw - y = 0$$

Lagrange Multipliers

Substituting w from the first equation into the 2nd equation, we get:

$$Xw - y = -X\frac{X^T\lambda}{2} - y = 0 \Rightarrow \lambda = -2(XX^T)^{-1}y$$

And therefore we plug back into the first equation to get:

$$w = \frac{X^T\lambda}{2} = (X^T)(XX^T)^{-1}y$$

which is our minimum norm solution.

General Norm Minimization

More generally, we may have a problem of the form

$$\min_{w} \|Xw - y\|$$
$$\text{subject to} \quad Cw = d$$

Here, least squares and least norm problems are a special case of the formulation above. The problem formulation is equivalent to saying:

$$\min_{w} \frac{1}{2}\|Xw - y\|^2$$
$$\text{subject to} \quad Cw = d$$

General Norm Minimization

The Lagrangian is:

$$\begin{aligned} L(x, \lambda) &= \frac{1}{2}\|Xw - y\|^2 + \lambda^T(Cw - d) \\ &= \frac{1}{2}w^T X^T X w - y^T X w + \frac{1}{2}y^T y + \lambda^T C w - \lambda^T d \end{aligned}$$

Taking partial derivative to x and λ and setting to 0:

$$\begin{aligned} \nabla_w L &= X^T X w - X^T y + C^T \lambda = 0 \\ \nabla_\lambda L &= Cw - d = 0 \end{aligned}$$

or

$$\begin{bmatrix} X^T X & X^T \\ X & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} X^T y \\ d \end{bmatrix}$$

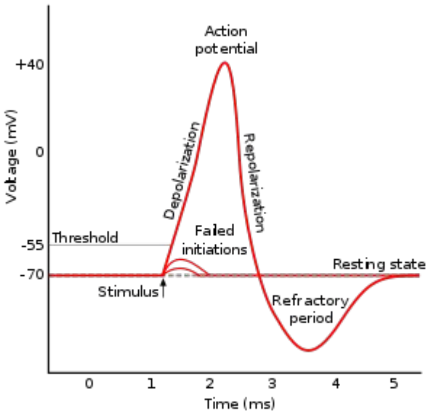
If the left-most block matrix is invertible, then we will have:

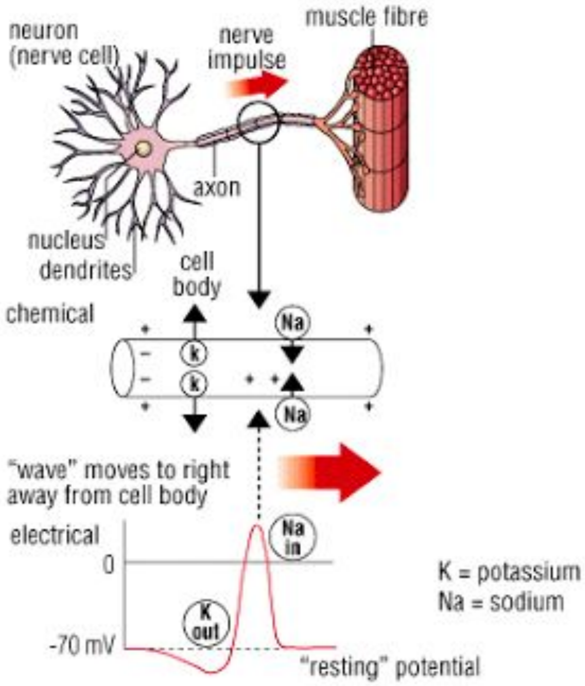
$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} X^T X & X^T \\ X & 0 \end{bmatrix}^{-1} \begin{bmatrix} X^T y \\ d \end{bmatrix}$$

Introduction to Brain Machine Interface

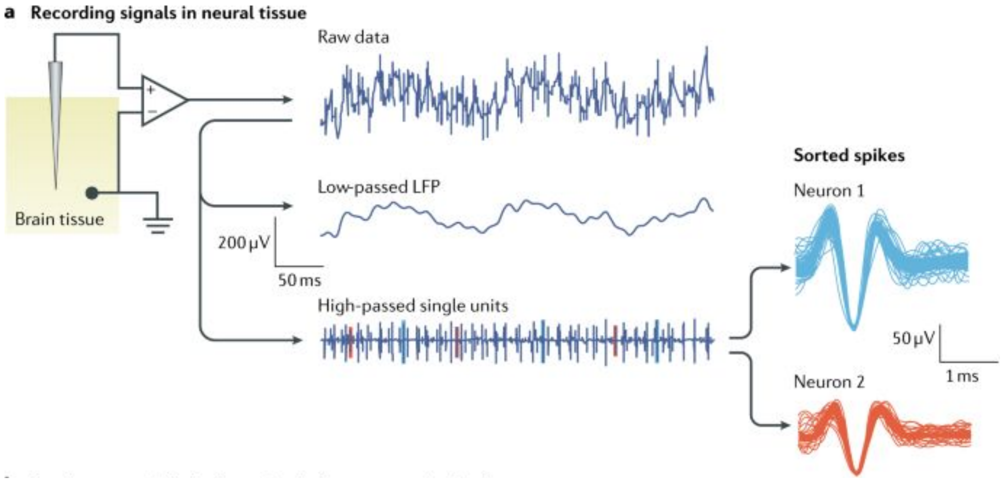
Basics of Electrophysiology

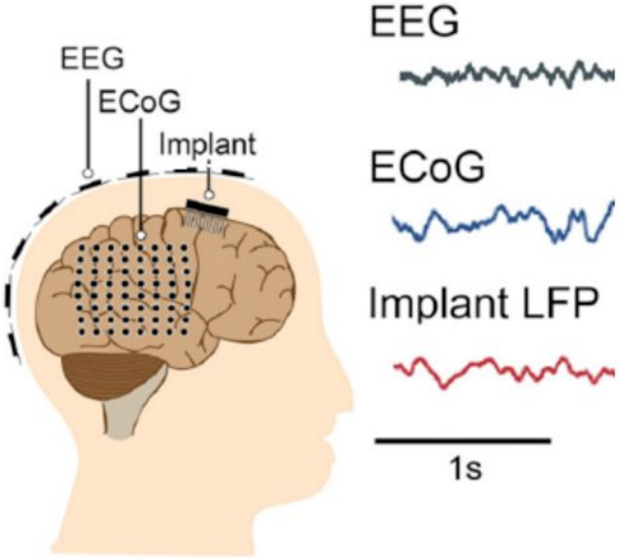
- Neurons conduct via action potentials
  - Electrical pulses due to electrochemical gradient across membrane
  - Causes neurotransmitter release
  - With enough neurotransmitter, subsequent neurons activate action potentials and pass on the signal





How signals are recorded: low level to high level

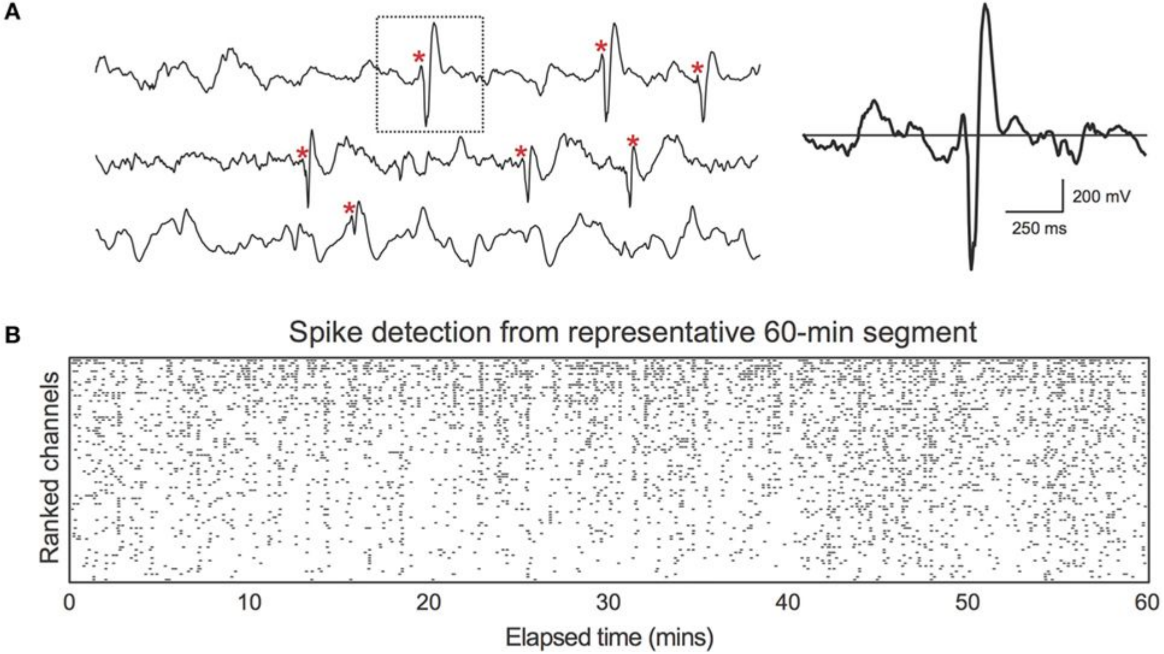




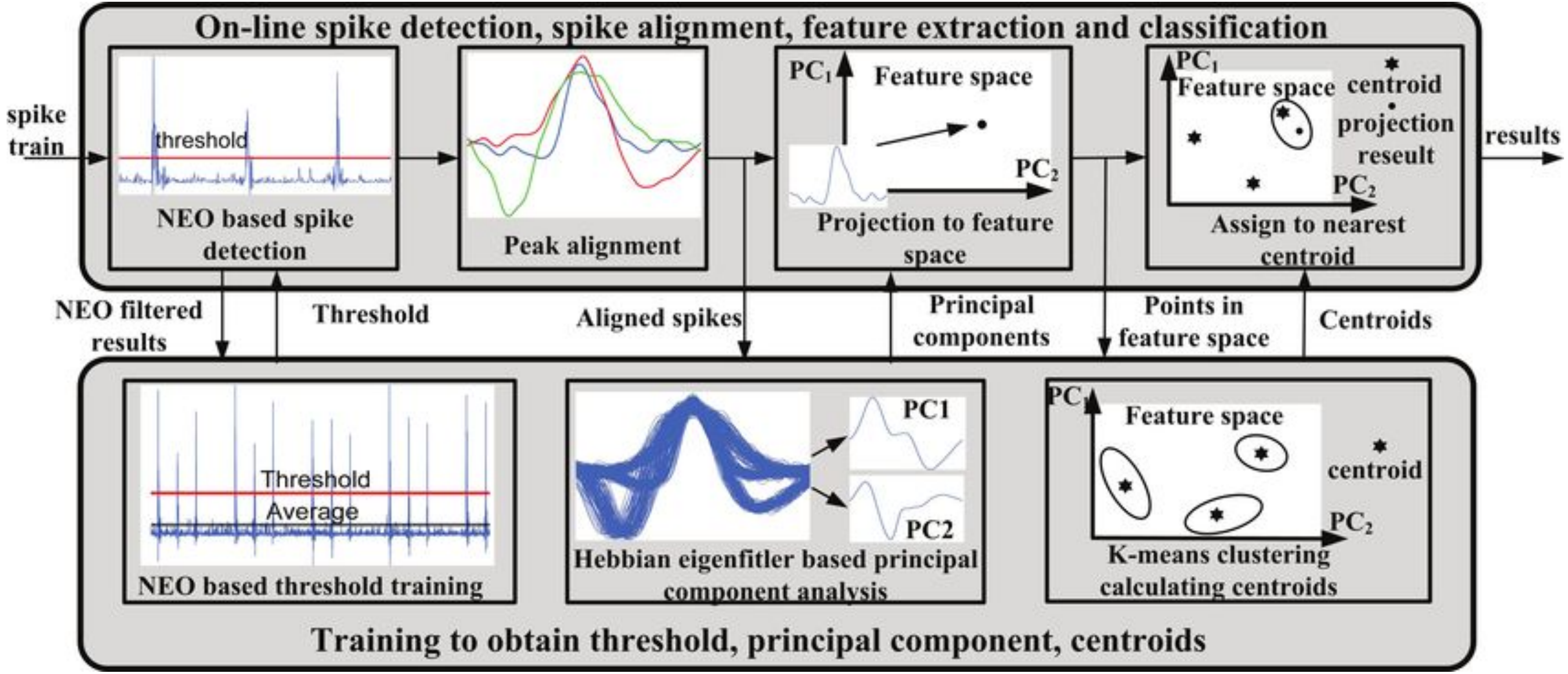
https://www.nature.com/articles/s41583-019-0140-6

Extract Spike Waveforms from neural recordings

- Filter Signal
- Detect spikes
- Waveform Extraction and Alignment
- Analysis



Application of PCA: Spike Sorting



Spike Waveform clustering

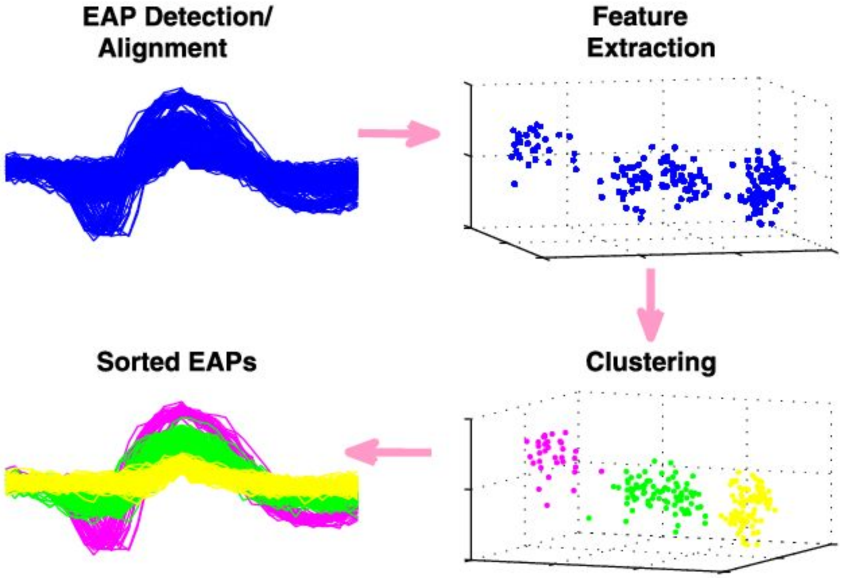
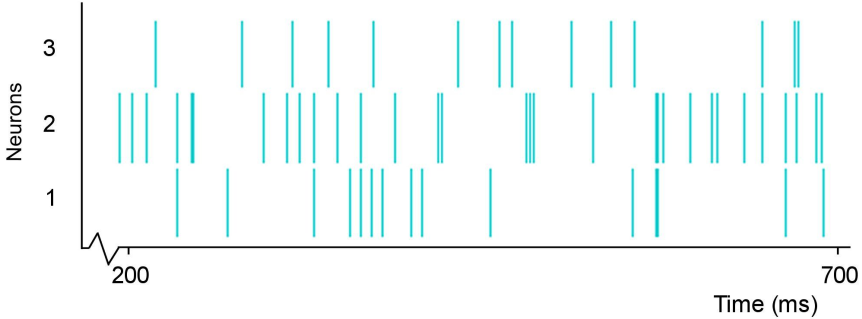
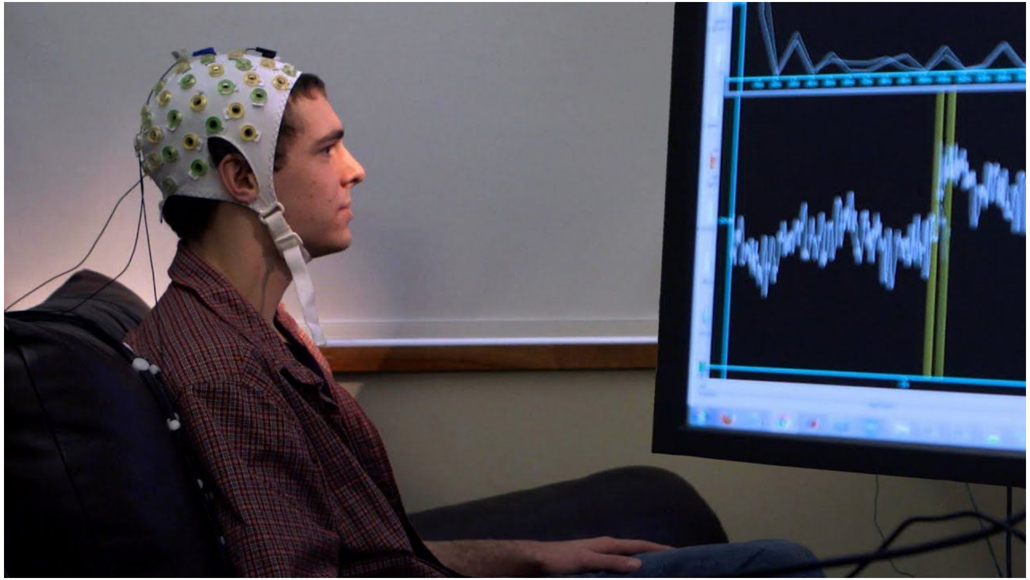


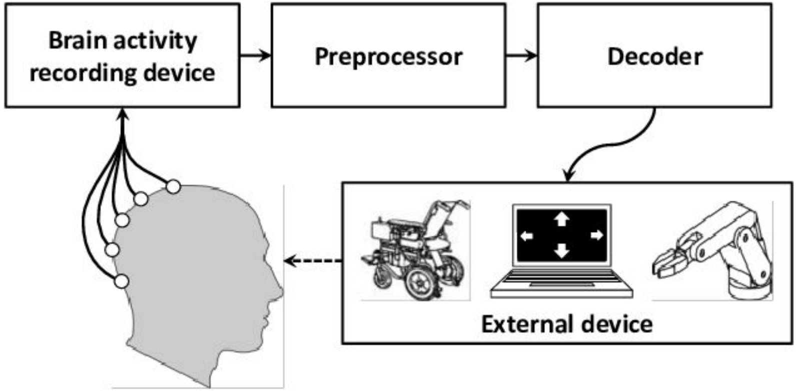
Fig. 1. Typical procedures for spike sorting.

Neuronal Spiking Over Time



Brain Computer Interface





Decode from neurons (spike class) and spike train (frequency)

Brain Computer Interface

Example classifying hand movement intention