A Minimax Lower Bound for EEG Source Localization

Praveen Venkatesh

Pulkit Grover

(vpraveen@cmu.edu)

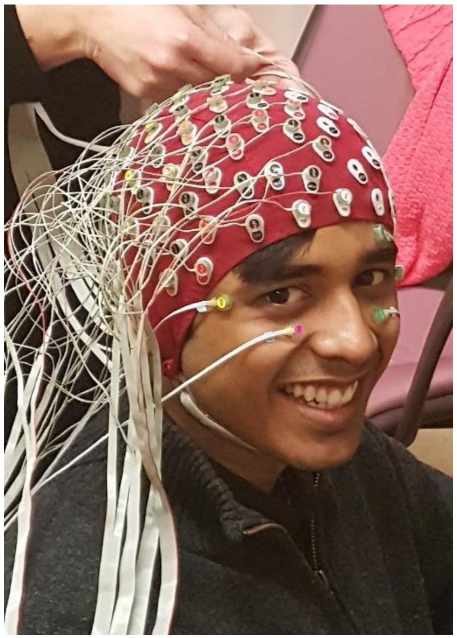
Carnegie Mellon University

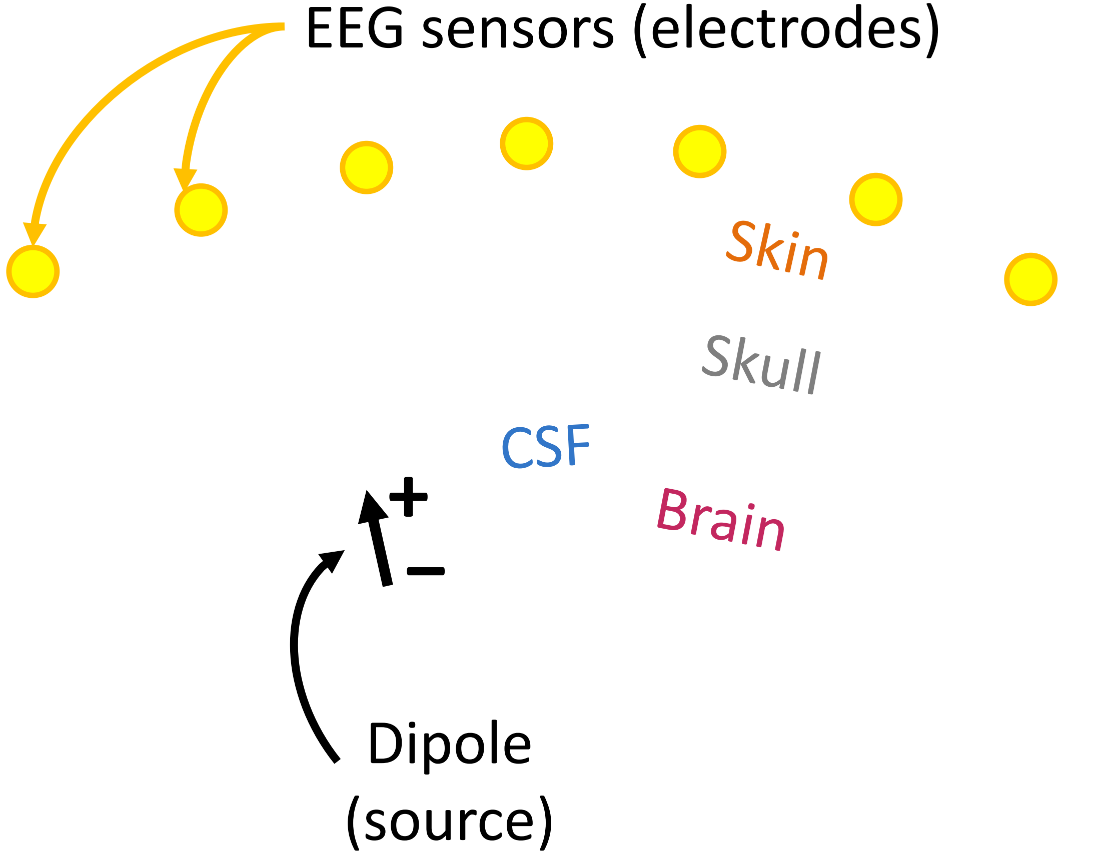
Electrical & Computer ENGINEERING

This work was supported in part by the **Dowd fellowship** from the College of Engineering at CMU, and in part by **SONIC**, one of the six SRC STARnet Centers, sponsored by MARCO and DARPA. The authors would like to thank Phillip and Marsha Dowd for their support and encouragement.

What is Electroencephalography (EEG)?

Measures brain activity (scalp potentials)





“Sources”: modeled as dipoles

Source Localization

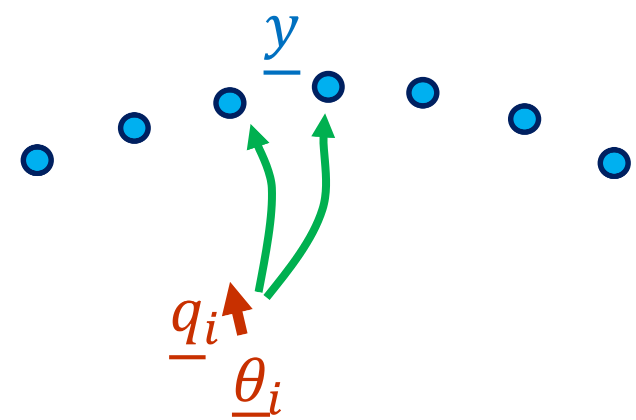
Estimate locations & dipole moments from sensor values

"Forward" problem:

Sensor values → $\underline{y} = \sum_{i=1}^{K} \boldsymbol{L}(\underline{\theta}_i) \cdot \underline{q}_i$

Location

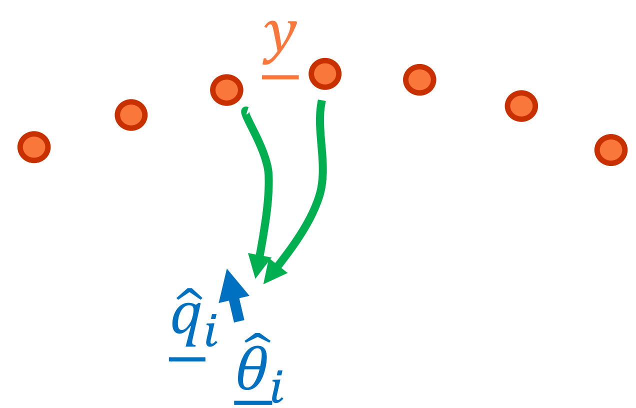
Dipole moment



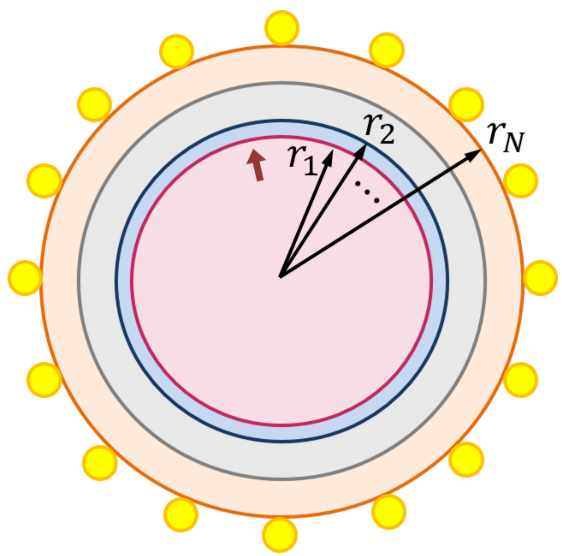
Source localization is the "inverse" problem, e.g.:

$$\left(\underline{\hat{q}}, \underline{\hat{\theta}}\right) = \arg\min_{\underline{\theta},\underline{q}} \left\|\underline{y} - \sum_{i=1}^{K} \boldsymbol{L}(\underline{\theta}_i) \cdot \underline{q}_i\right\|^2$$

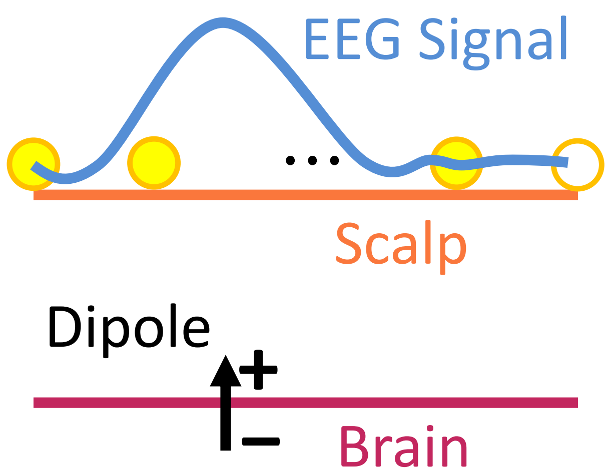
"Equivalent dipole fitting"



Simplifying the brain model

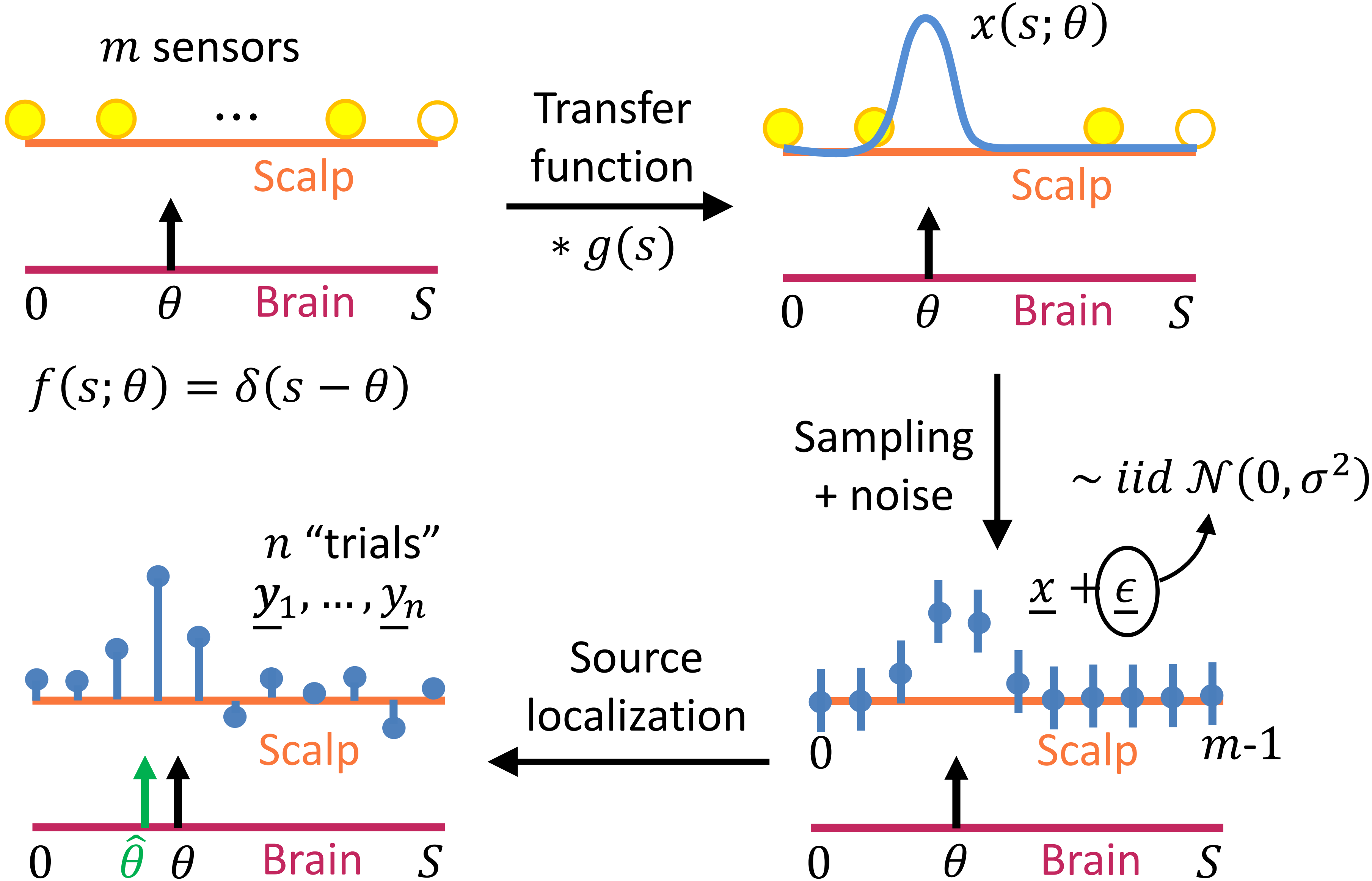


1. Spherical head model
   - ✘ Shifts are rotations
   - ✘ Spherical Harmonics
   - ✘ Uniform sampling



2. Linear head model
   - "Circular" domain
   - Known transfer function

Problem setup



Previous work on fundamental limits

- Source localization literature
  - Mosher et. al. (1993): Cramer Rao lower bounds
  - Gro ethod
- Mini blems
  - Ibra
  - Efromovich, '97
  - Cavalier and Tsybakov, '02

Did not address:

How does error scale with number of sensors?

Minimax Lower Bound: Le Cam's method

$$\theta \longrightarrow p_\theta \longrightarrow \underline{y} \longrightarrow \hat{\theta}$$

Dipole location	Data-generating distribution	Observed sensor values	Location estimate

Loss function:

$$\Phi\left(\rho\left(\hat{\theta}, \theta\right)\right) = \left\|\hat{\theta} - \theta\right\|_2^2$$

Minimax risk:

$$\inf_{\underline{\hat{\theta}}} \sup_{\underline{\theta} \in \Theta} \mathbb{E}_{\underline{y}}\left[\Phi\left(\rho\left(\hat{\theta}(\underline{y}), \theta\right)\right)\right]$$

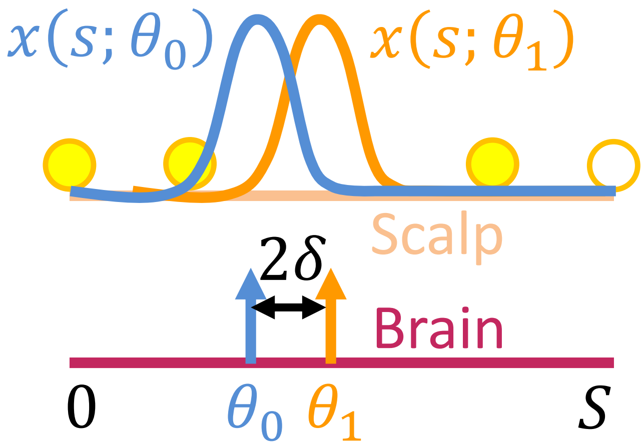
Best possible estimator

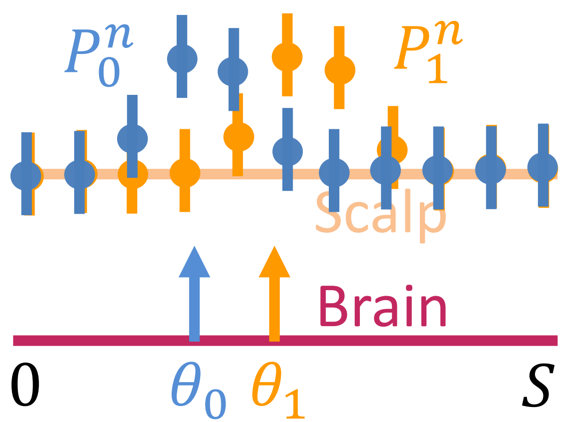
Worst-case parameters

Choose the estimator with the *best* worst-case performance

Le Cam's method

$$\Phi\left(\rho(\hat{\theta},\theta)\right) = \|\hat{\theta}-\theta\|_2^2$$





$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}_{\underline{y}}\left[\Phi\left(\rho(\hat{\theta}(\underline{y}),\theta)\right)\right]$$

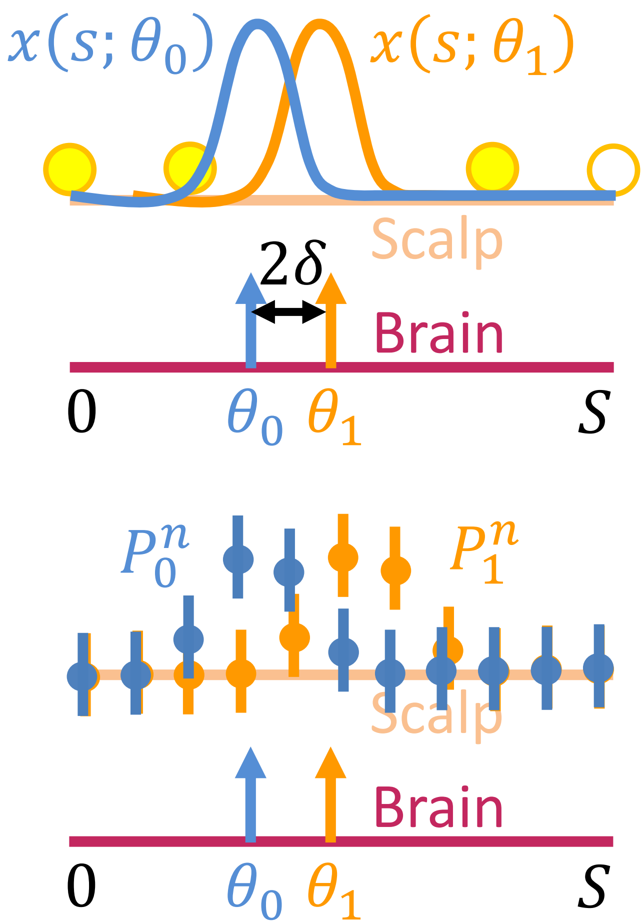
$$\geq \Phi(\delta) \inf_{\hat{V}} \mathbb{P}[\theta_{\hat{V}} \neq \theta_V]$$

$$= \frac{\delta^2}{2}[1 - \|P_1^n - P_0^n\|_{TV}]$$

$$\geq \frac{\delta^2}{2}\left[1 - \sqrt{\frac{n}{2} D_{KL}(P_0 \parallel P_1)}\right]$$

$$\geq \frac{\delta^2}{2}\left[1 - \sqrt{\frac{n}{4\sigma^2}\|\underline{x}(\theta_0) - \underline{x}(\theta_1)\|^2}\right]$$

Le Cam's method



$$\inf_{\hat{\theta}} \sup_{\theta\in\Theta} \mathbb{E}_{\underline{y}}\left[\left\|\hat{\theta}(\underline{y}) - \theta\right\|^2\right]$$

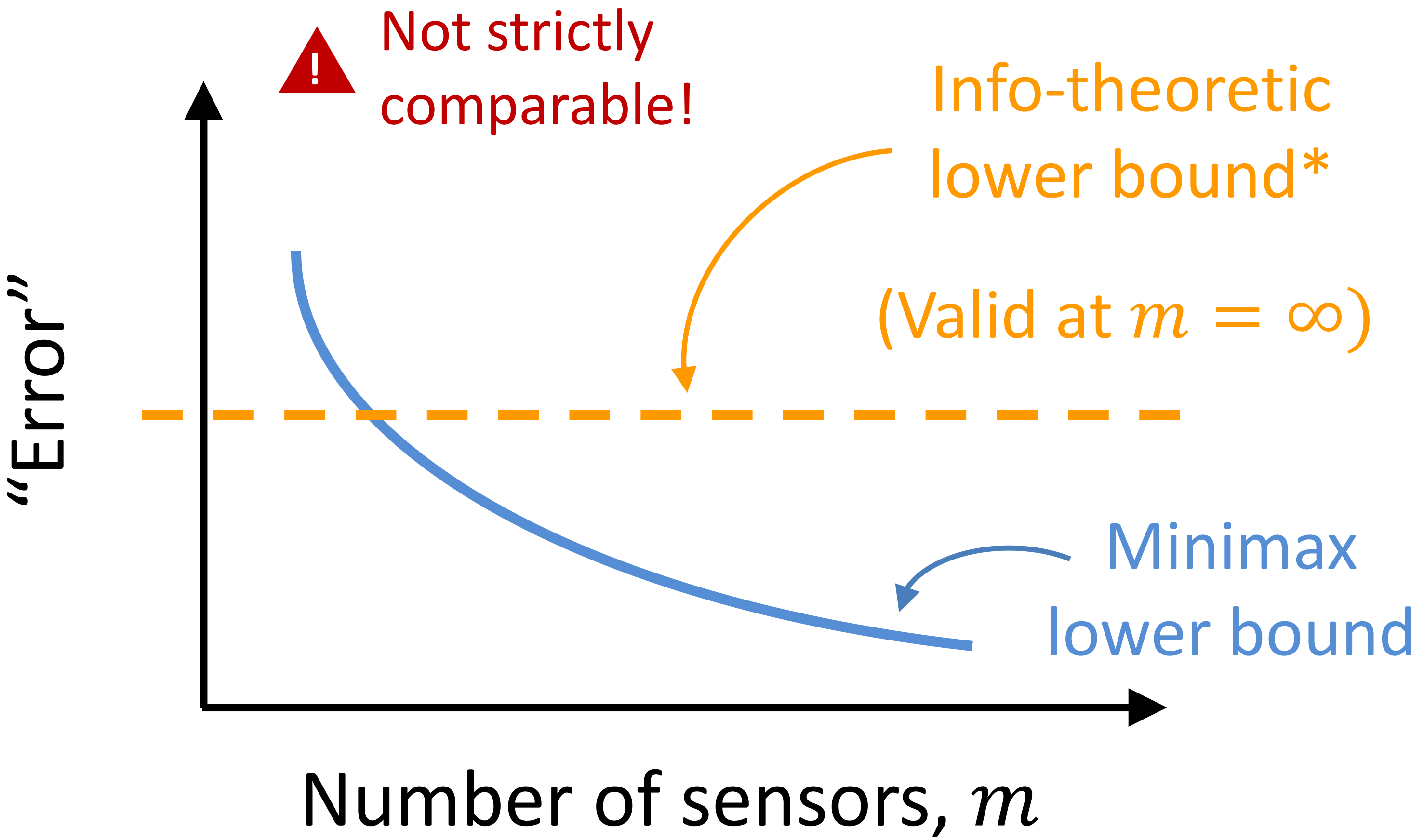
$$\geq \frac{\delta^2}{2}\left[1 - \sqrt{\frac{n}{4\sigma^2}\left\|\underline{x}(\theta_0) - \underline{x}(\theta_1)\right\|^2}\right]$$

$$\geq \frac{\delta^2}{2}\left[1 - \sqrt{\frac{n}{4\sigma^2}\kappa^2 4\delta^2 \cdot m}\right]$$

$$\approx \frac{1}{32}\,\frac{\sigma^2/n}{m\,\kappa^2}\,\frac{S}{w} \xrightarrow{m\to\infty} 0$$

So what's wrong with this?

This bound is loose!



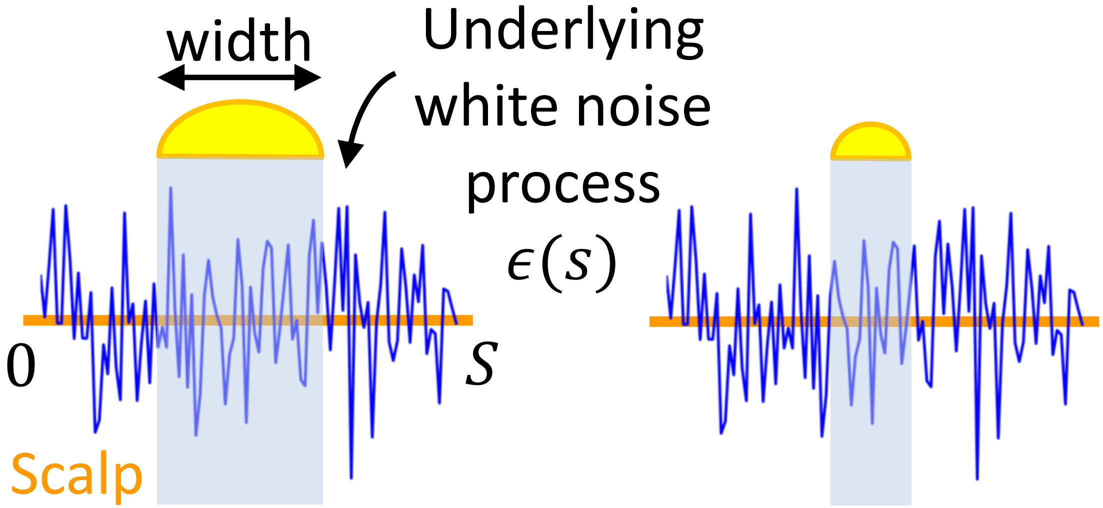
*Pulkit Grover, ISIT '16

Sensor model!

Noise $\sim iid\ \mathcal{N}(0, \sigma^2)$,
independent of m

SNR *decreases* as
# of sensors increases

$$y_k = \int_{\text{width}} \big(x(s;\theta) + \epsilon(s)\big)\, ds$$



More averaging

Less averaging

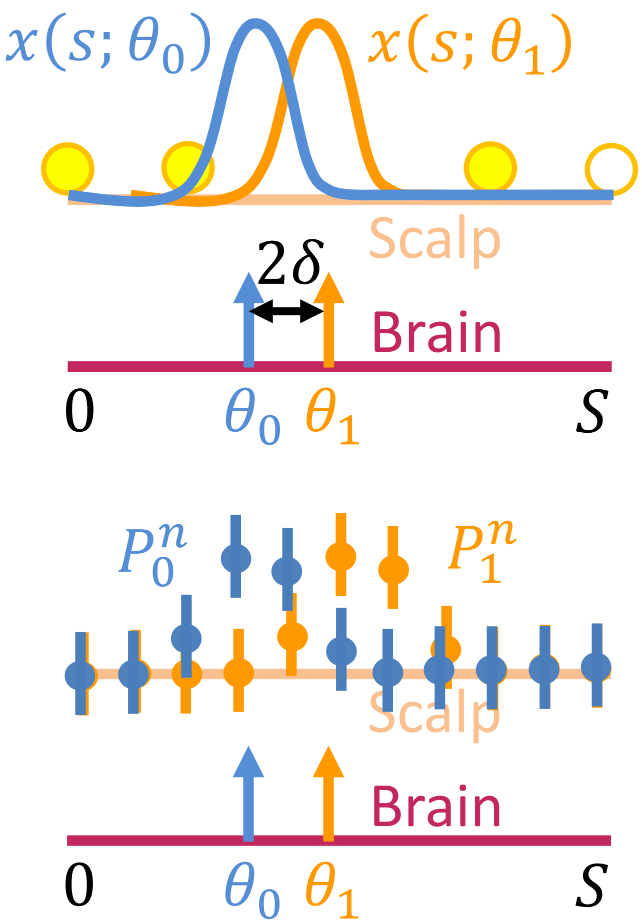
Noise var. ∝ 1 / width

width ∝ 1 / #sensors

Noise var. ∝ #sensors

SNR ∝ 1 / #sensors

Bounds for the “integrator sensor” model



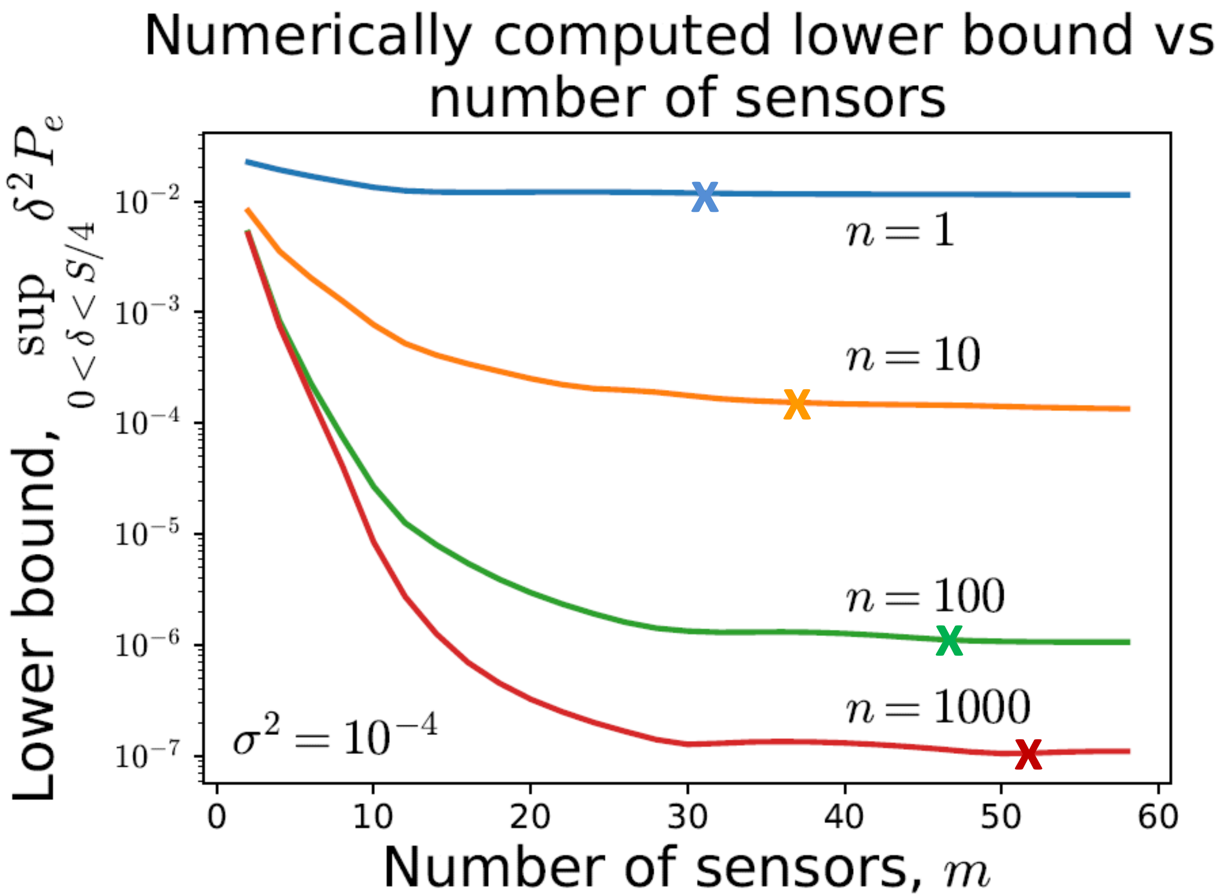
$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_{\underline{y}} \left[\left\| \hat{\theta}(\underline{y}) - \theta \right\|^2 \right]$$

$$\geq \Phi(\delta) \inf_{\hat{V}} \mathbb{P}\left[\hat{V} \neq V\right]$$

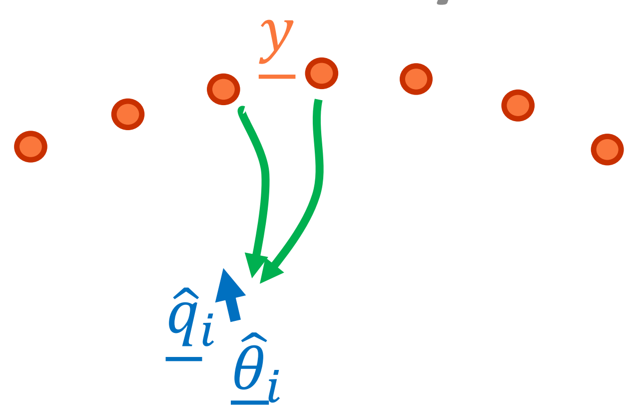
$$\geq \delta^2 Q\left(\frac{\left\| \underline{x}(\theta_0) - \underline{x}(\theta_1) \right\|}{\sigma(m)} \right)$$

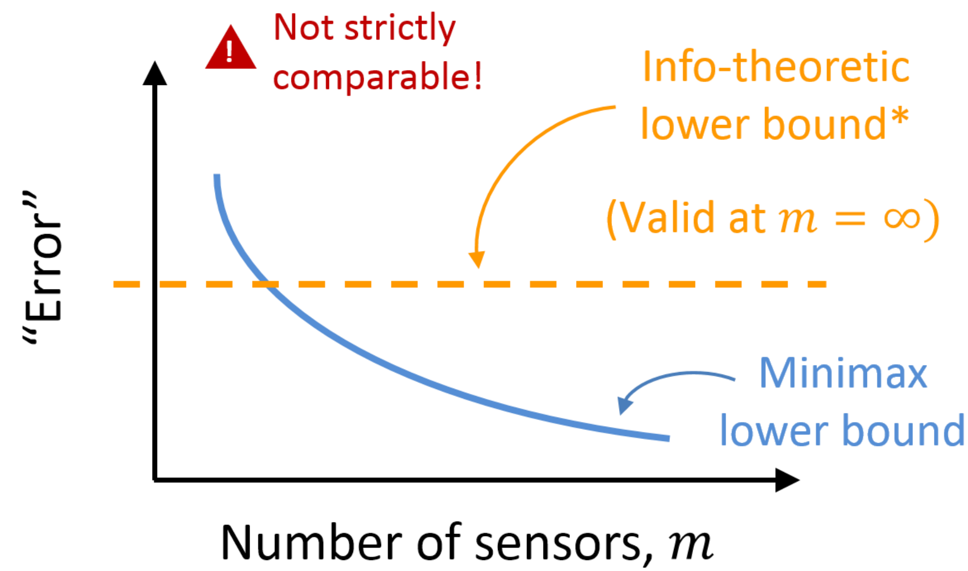
(For gaussian noise)

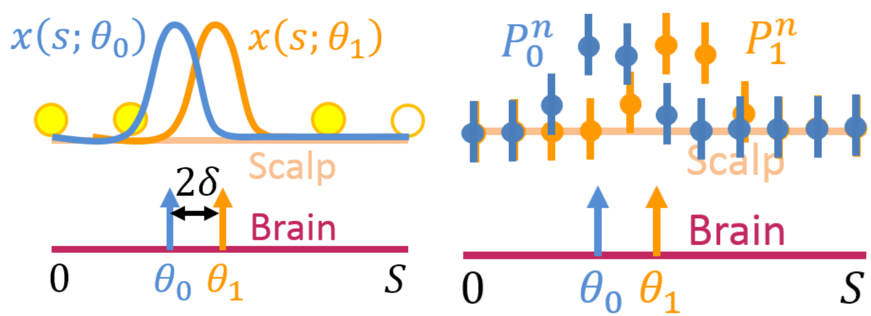
Bounds for the “integrator sensor” model

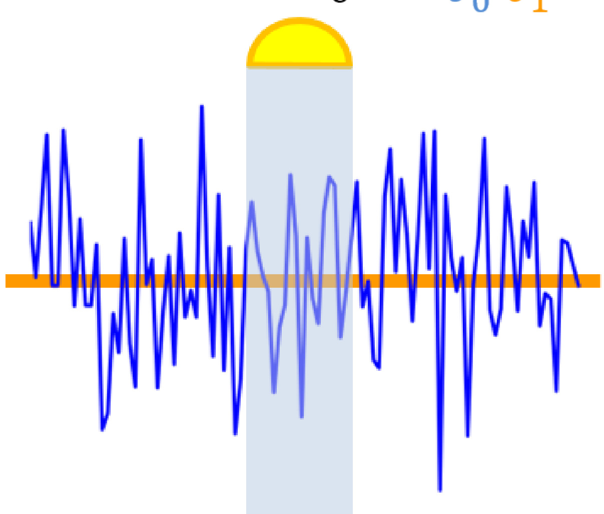


Summary

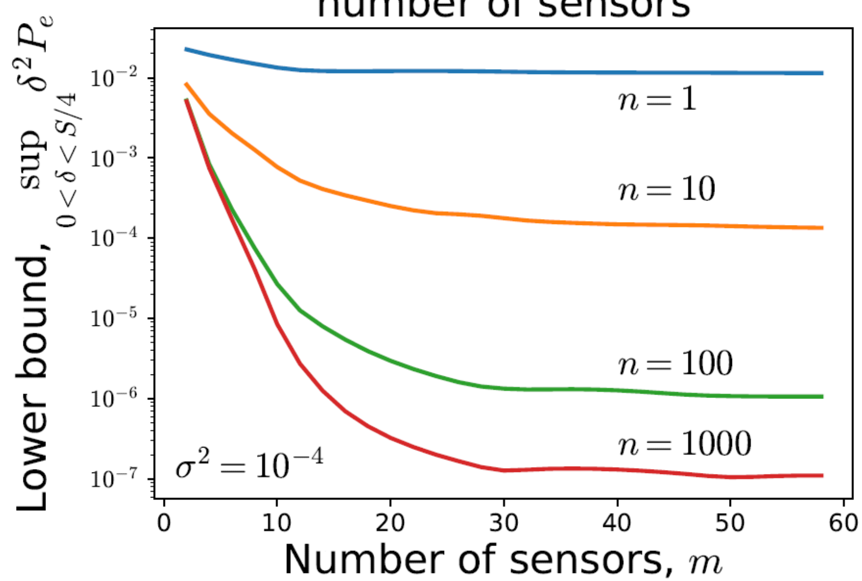




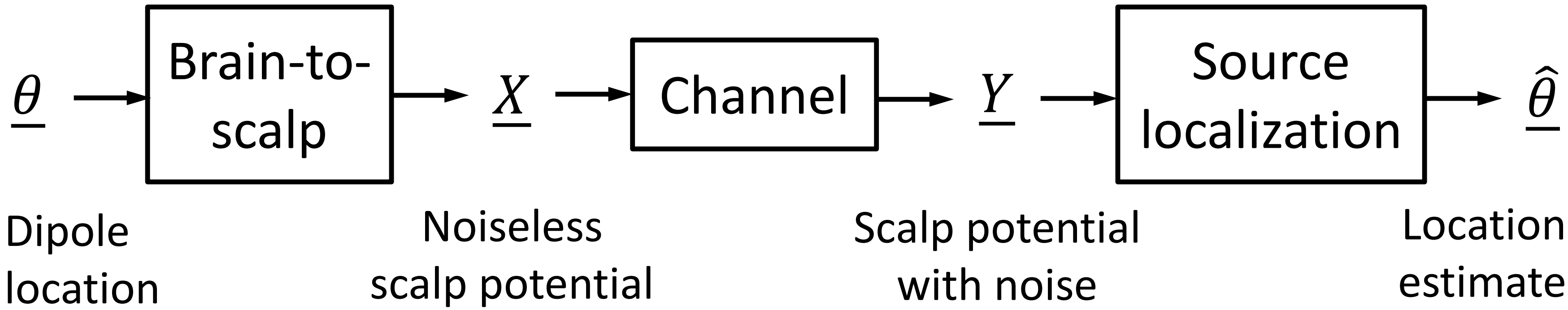




Numerically computed lower bound vs. number of sensors



Shortcomings of the info-theory bound



$$\mathbb{E}\left\|\underline{\hat{\theta}} - \underline{\theta}\right\|^2 \sim I\left(\underline{\theta}; \underline{\hat{\theta}}\right) \leq I\left(\underline{X}; \underline{Y}\right) \leq C \text{ (channel capacity)}$$

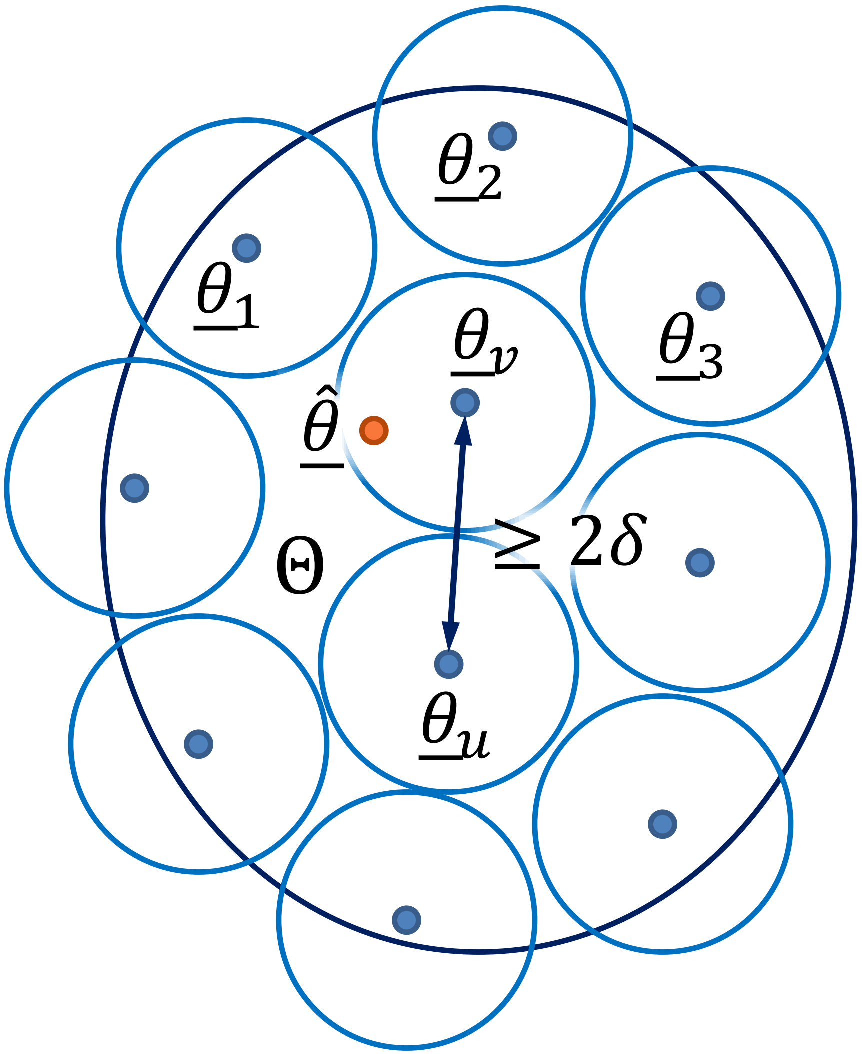
$\underline{X}$ and $\underline{Y}$ are "continuous-space" signals

$\Rightarrow$ Assumes an infinite number of sensors

Can severely underestimate the lower bound!

e.g. imagine if you had only 10 sensors

Lower bounding estimation error with hypothesis testing error

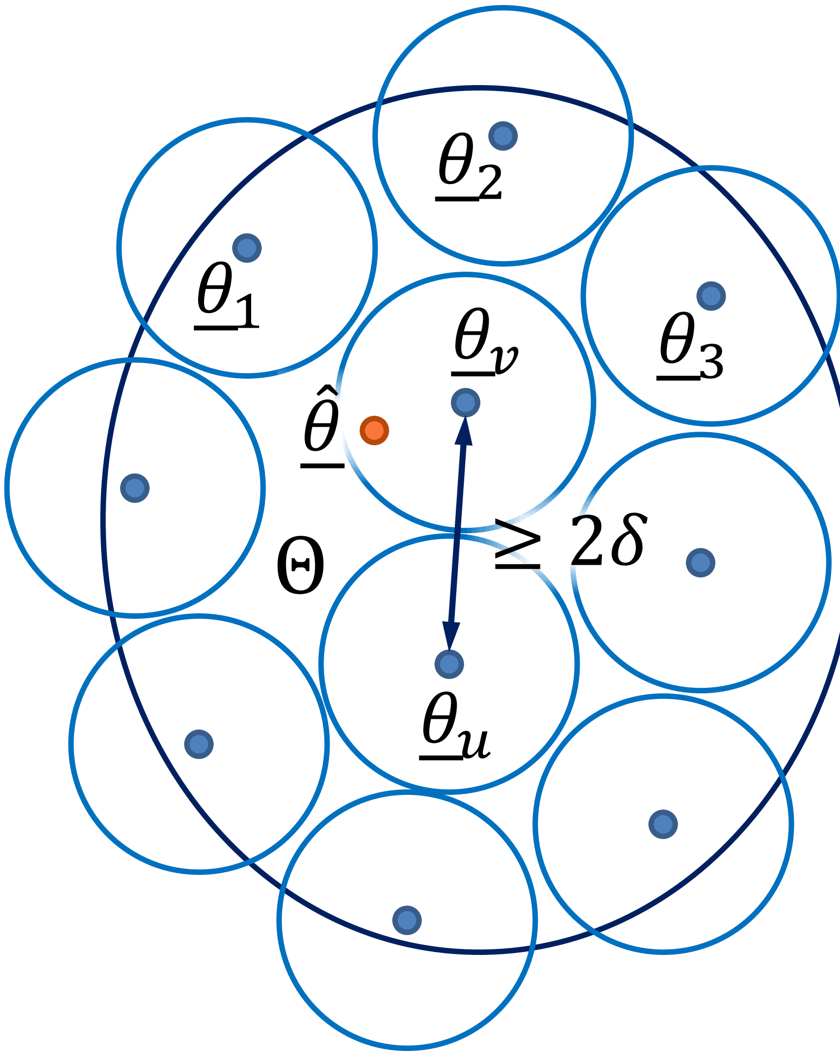


- Assume $\underline{\theta} \in \{\underline{\theta}_v\}_{v=1}^{\mathcal{V}}$
- V = unknown index
- $\hat{V}$ = Estimator of V

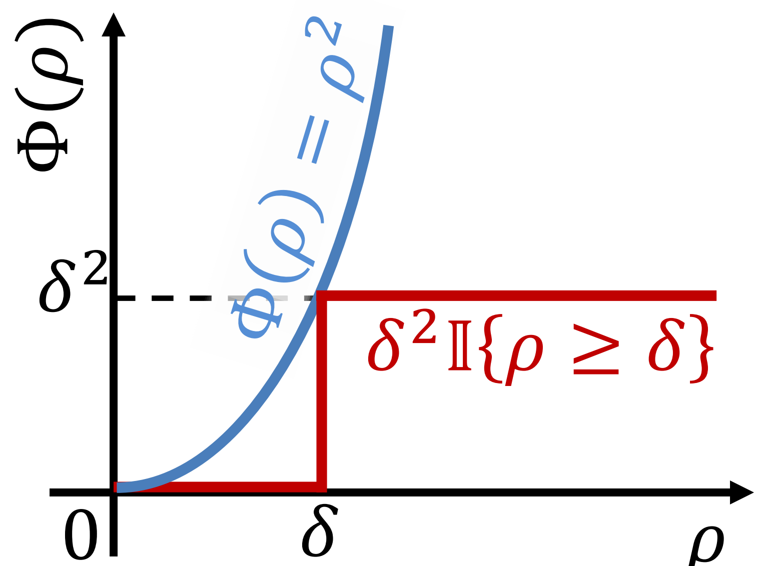
$$\hat{V}(Y) = \arg\min_{v \in \mathcal{V}} \rho\left(\underline{\hat{\theta}}(Y), \underline{\theta}_v\right)$$

Construct a "2δ-packing" of Θ

Relating error in estimation and hypothesis testing



$$\mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta}_v)\right)\right]$$
$$\geq \mathbb{E}\left[\Phi(\delta)\,\mathbb{I}\{\rho(\underline{\hat{\theta}}, \underline{\theta}_v) \geq \delta\}\right]$$
$$= \Phi(\delta)\,\mathbb{P}\left[\rho(\underline{\hat{\theta}}, \underline{\theta}_v) \geq \delta\right]$$
$$= \Phi(\delta)\,\mathbb{P}\left[\hat{V} \neq v \mid V = v\right]$$



From estimation to testing

Need to bound: $\inf_{\underline{\hat{\theta}}} \sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta})\right)\right]$

$$\sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta})\right)\right] \geq \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta_v})\right)\right]$$

$$\mathcal{V} \cdot \sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta})\right)\right] \geq \sum_{v=1}^{\mathcal{V}} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta_v})\right)\right]$$

$$\sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta})\right)\right] \geq \frac{1}{\mathcal{V}} \sum_{v=1}^{\mathcal{V}} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta_v})\right)\right]$$

Putting it together

$$\inf_{\underline{\hat{\theta}}} \sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta})\right)\right]$$

$$\geq \inf_{\underline{\hat{\theta}}} \frac{1}{\mathcal{V}} \sum_{v=1}^{\mathcal{V}} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta}_v)\right)\right]$$

$$\sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta})\right)\right] \geq \frac{1}{\mathcal{V}} \sum_{v=1}^{\mathcal{V}} \mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta}_v)\right)\right]$$

$$\geq \inf_{\underline{\hat{V}}} \frac{1}{\mathcal{V}} \sum_{v=1}^{\mathcal{V}} \Phi(\delta)\, \mathbb{P}\left[\hat{V} \neq v \mid V = v\right]$$

$$\mathbb{E}\left[\Phi\left(\rho(\underline{\hat{\theta}}, \underline{\theta}_v)\right)\right] \geq \Phi(\delta)\, \mathbb{P}\left[\hat{V} \neq v | V = v\right]$$

$$= \Phi(\delta) \inf_{\hat{V}} \mathbb{P}\left[\hat{V} \neq V\right]$$

Need to bound this next

V takes values $\{1, \dots, \mathcal{V}\}$ uniformly

Le Cam's method

Binary hypothesis: $V \in \{0,1\}$ (uniformly)

$$\mathbb{P}\left[\hat{V} \neq V\right] = \frac{1}{2} P_{V=0}\left(\hat{V} \neq 0\right) + \frac{1}{2} P_{V=1}\left(\hat{V} \neq 1\right)$$

Define $A = \left\{\underline{Y} : \hat{V}\left(\underline{Y}\right) = 1\right\}$, "acceptance region"

$$\frac{1}{2}\left[P_0\left(\hat{V} \neq 0\right) + P_1\left(\hat{V} \neq 1\right)\right] = \frac{1}{2}\left[P_0(A) + P_1(A^c)\right]$$

$$= \frac{1}{2}\left[P_0(A) + 1 - P_1(A)\right]$$

Taking infimum:

$$\inf_{\hat{V}} \mathbb{P}\left[\hat{V} \neq V\right] = \frac{1}{2} \inf_{A} \left\{1 - \left(P_1(A) - P_0(A)\right)\right\}$$

$$= \frac{1}{2}\left[1 - \sup_{A} \left\{P_1(A) - P_0(A)\right\}\right]$$

$$= \frac{1}{2}\left[1 - \|P_1 - P_0\|_{TV}\right]$$

Le Cam's method for source localization

$$\|P_1^n - P_0^n\|_{TV}^2 \leq \frac{1}{2} D_{KL}(P_1^n \| P_0^n) = \frac{n}{2} D_{KL}(P_1 \| P_0)$$

For normal distributions of equal variance,

$$D_{KL}(P_1 \| P_0) = \frac{1}{2\sigma^2} \left\| \boldsymbol{L}(\underline{\theta}_0)\underline{q} - \boldsymbol{L}(\underline{\theta}_1)\underline{q} \right\|^2 = \frac{1}{2\sigma^2} d(\underline{\theta}_0, \underline{\theta}_1)$$

$$\inf_{\hat{\underline{\theta}}} \sup_{\underline{\theta} \in \Theta} \mathbb{E}\left[\Phi\left(\rho(\hat{\underline{\theta}}, \underline{\theta})\right)\right] \geq \sup_{\underline{\theta}_0, \underline{\theta}_1} \frac{\delta^2}{2}\left[1 - \sqrt{\frac{n}{4\sigma^2} d(\underline{\theta}_0, \underline{\theta}_1)}\right]$$