USC Viterbi
School of Engineering

Data-driven sensor scheduling

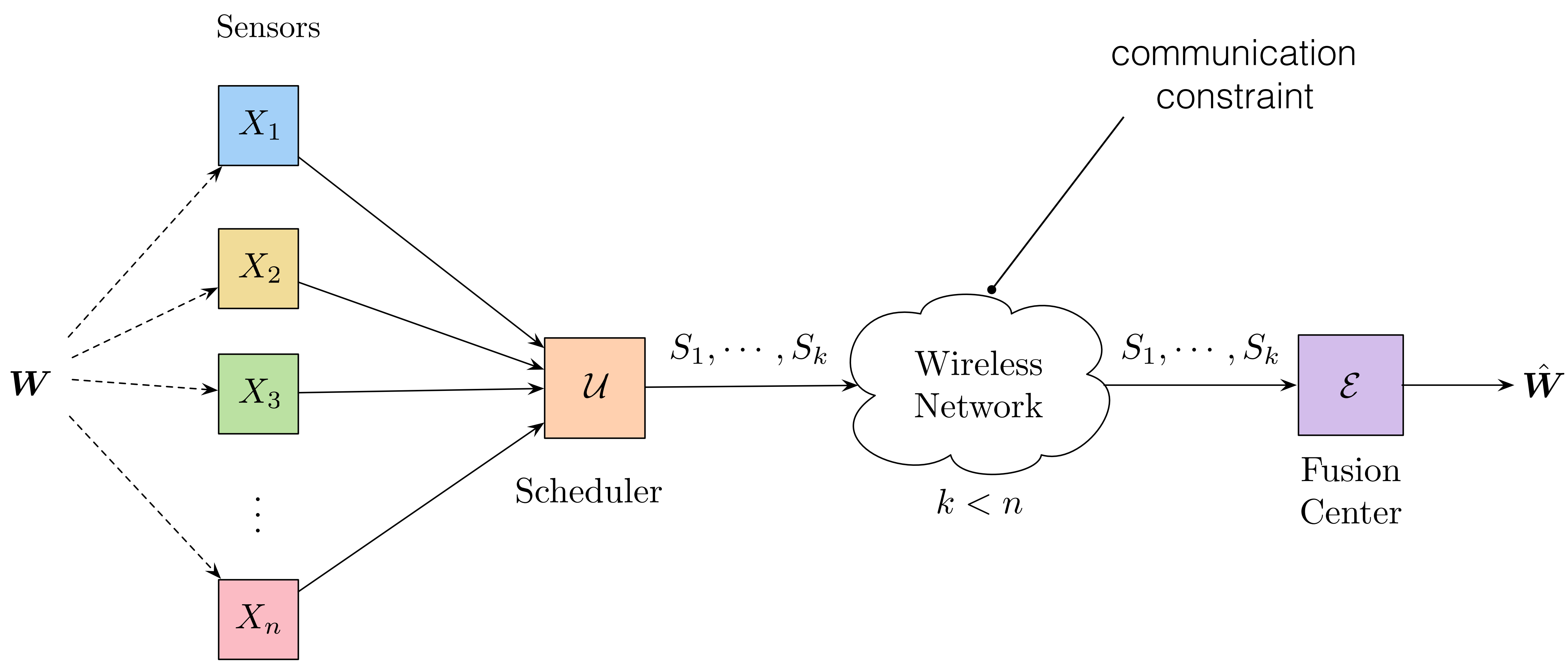
Marcos M. Vasconcelos
mvasconc@usc.edu

Urbashi Mitra
ubli@usc.edu

Dept. of Electrical Engineering
University of Southern California

Asilomar Conference on Signals, Systems, and Computers
November 3rd-6th, 2019 - Pacific Grove, CA

Sensor scheduling problem

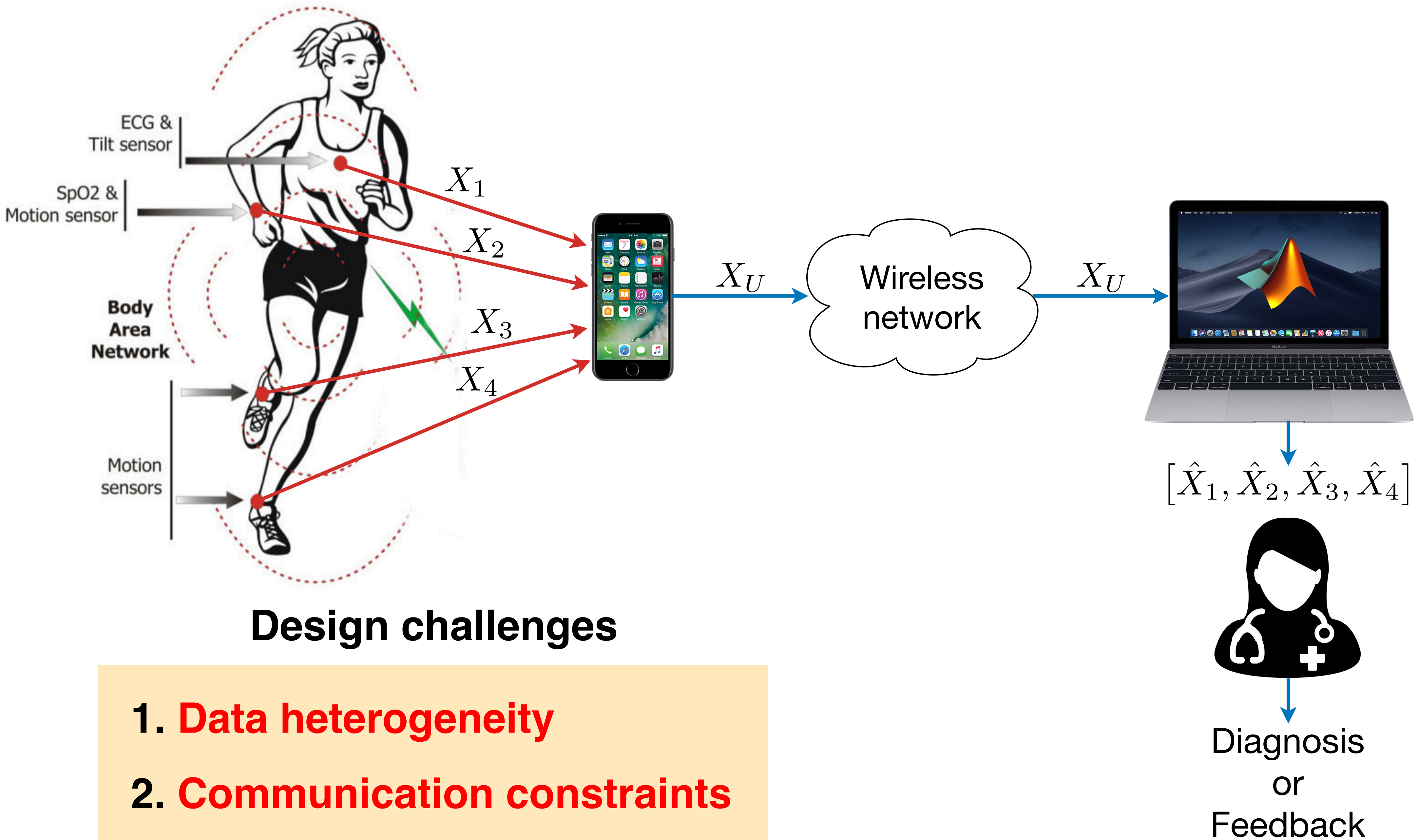


Choose k out of n sensors such that
the expected distortion between $\boldsymbol{W}$ and $\hat{\boldsymbol{W}}$ is minimized

Athans - *Automatica* 1972
Joshi & Boyd - *IEEE TSP* 2009
Mo, Ambrosino & Sinopoli - *Automatica* 2011
Moon & Basar - *IEEE TSP* 2017

Body area networks

System coupling sensors on people and wireless networks



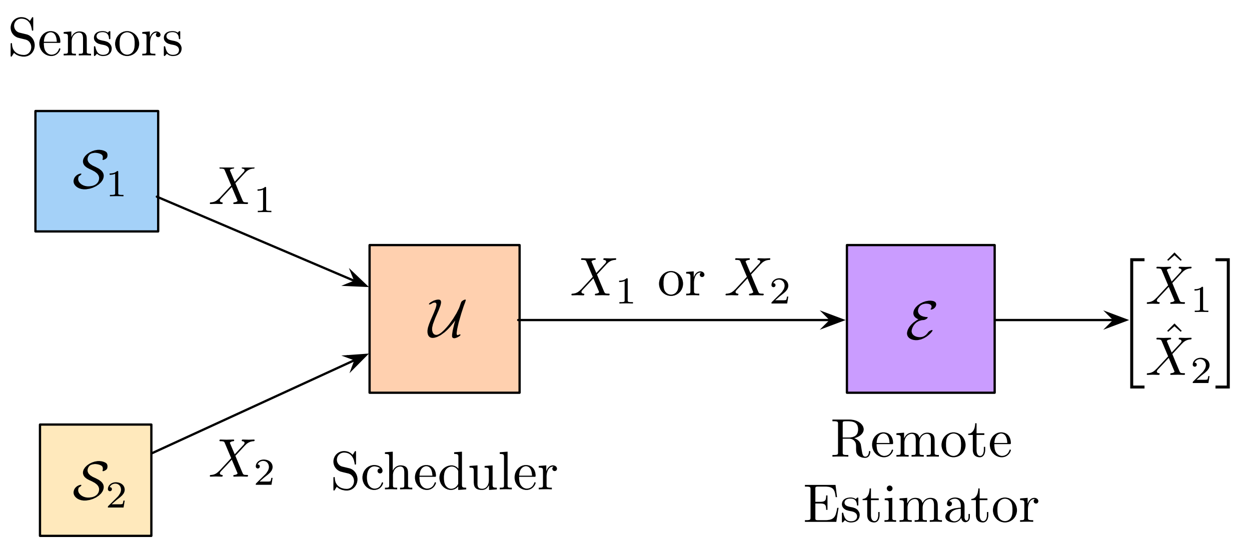
Design challenges

1. **Data heterogeneity**
2. **Communication constraints**
3. **Energy constraints**

Simplest case: two sensors

Observations

$(X_1, X_2) \sim f(x_1, x_2)$



Decision variable

$$U \in \{1, 2\}$$

Transmit

$$S = (1, X_1)$$

Transmit

$$S = (2, X_2)$$

Scheduling policy

$$U = \mathcal{U}(X_1, X_2)$$

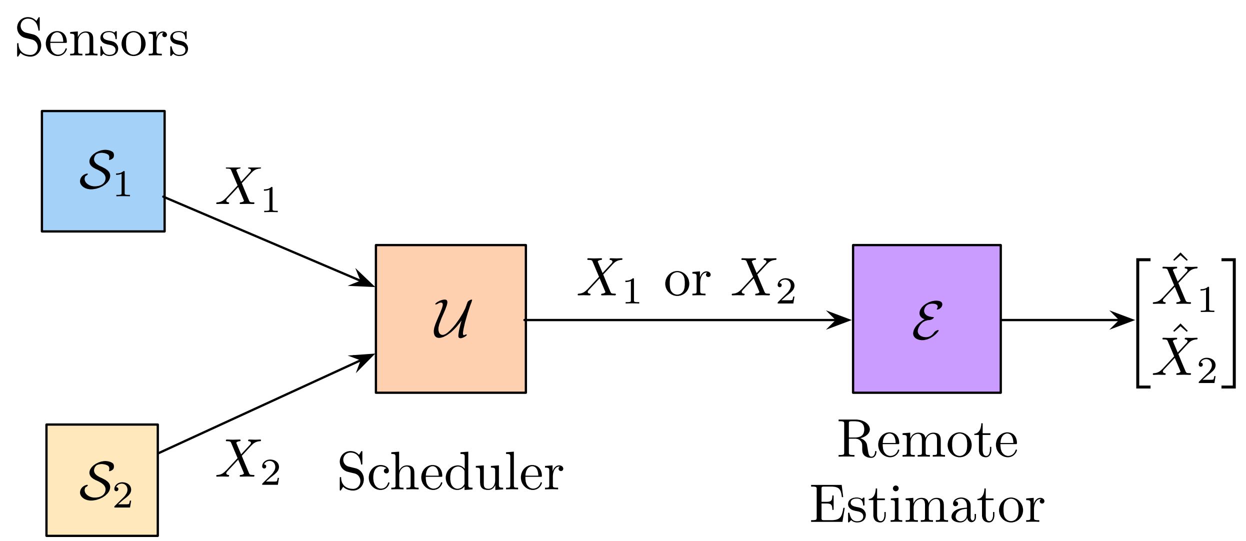
Estimation policy

$$\begin{bmatrix}\hat{X}_1\\ \hat{X}_2\end{bmatrix} = \mathcal{E}(Y)$$

Simplest case: two sensors

Observations

$(X_1, X_2) \sim f(x_1, x_2)$

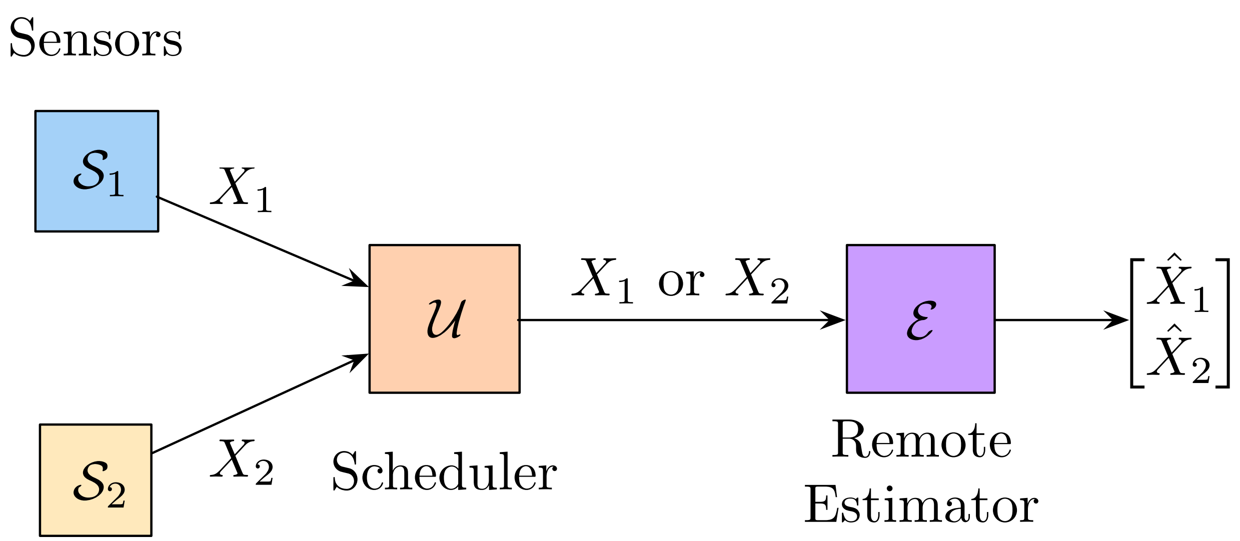


Decision variable	**Scheduling policy**	**Estimation policy**
$U \in \{1, 2\}$	$U = \mathcal{U}(X_1, X_2)$	$\begin{bmatrix}\hat{X}_1\\ \hat{X}_2\end{bmatrix} = \mathcal{E}(Y)$

Problem

$$\min_{(\mathcal{U},\mathcal{E})\in\mathbb{U}\times\mathbb{E}} \quad \mathcal{J}(\mathcal{U},\mathcal{E}) = \mathbf{E}\left[(X_1 - \hat{X}_1)^2 + (X_2 - \hat{X}_2)^2\right]$$

Simplest case: two sensors



$$\min_{(\mathcal{U},\mathcal{E})\in\mathbb{U}\times\mathbb{E}} \quad \mathcal{J}(\mathcal{U},\mathcal{E}) = \mathbf{E}\left[(X_1-\hat{X}_1)^2+(X_2-\hat{X}_2)^2\right]$$

“Dimensionality reduction”

or

“Subset selection”

Roweis and Saul - *Science* 2000

Notions of optimality

Global optimality

$$\mathcal{J}(\mathcal{U}^\star, \mathcal{E}^\star) \le \mathcal{J}(\mathcal{U}, \mathcal{E}), \quad (\mathcal{U}, \mathcal{E}) \in \mathbb{U} \times \mathbb{E}$$

$$\Longrightarrow$$

$$\not\Longleftarrow$$

Person-by-person optimality

$$\mathcal{J}(\mathcal{U}^\star, \mathcal{E}^\star) \le \mathcal{J}(\mathcal{U}, \mathcal{E}^\star), \quad \mathcal{U} \in \mathbb{U}$$
$$\mathcal{J}(\mathcal{U}^\star, \mathcal{E}^\star) \le \mathcal{J}(\mathcal{U}^\star, \mathcal{E}), \quad \mathcal{E} \in \mathbb{E}$$

Unfortunately, finding **globally optimal** solutions is **very difficult**

Finding **person-by-person optimal** solutions is **often much easier***

***depending on the statistics of the data**

Observation-driven scheduling

Arbitrary joint density

$$(X_1, X_2) \sim f(x_1, x_2)$$

$$\min_{(\mathcal{U},\mathcal{E})\in\mathbb{U}\times\mathbb{E}} \quad \mathcal{J}(\mathcal{U},\mathcal{E}) = \mathbf{E}\left[(X_1 - \hat{X}_1)^2 + (X_2 - \hat{X}_2)^2\right]$$

The optimal estimator is the conditional mean

$$\mathcal{E}^\star_{\mathcal{U}}(y) = \mathbf{E}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \;\middle|\; Y = y\right]$$

Optimal estimator

$$(X_1, X_2) \sim f(x_1, x_2)$$

$$\min_{(\mathcal{U},\mathcal{E}) \in \mathbb{U}\times\mathbb{E}} \quad \mathcal{J}(\mathcal{U},\mathcal{E}) = \mathbf{E}\left[(X_1 - \hat{X}_1)^2 + (X_2 - \hat{X}_2)^2\right]$$

The optimal estimator is the conditional mean

$$\mathcal{E}^\star_{\mathcal{U}}(1, x_1) = \begin{pmatrix} x_1 \\ \mathbf{E}\left[X_2 \mid U = 1, X_1 = x_1\right] \end{pmatrix}$$

$$\mathcal{E}^\star_{\mathcal{U}}(2, x_2) = \begin{pmatrix} \mathbf{E}\left[X_1 \mid U = 2, X_2 = x_2\right] \\ x_2 \end{pmatrix}$$

Generalized Nearest Neighbor Condition

Assuming that

$$\mathcal{E}(1,x_1) = \begin{bmatrix} x_1 \\ \eta_2(x_1) \end{bmatrix} \qquad \mathcal{E}(2,x_2) = \begin{bmatrix} \eta_1(x_2) \\ x_2 \end{bmatrix}$$

The cost becomes

$$\mathcal{J}(\mathcal{U},\mathcal{E}) = \int_{\mathbb{R}^2} \big(x_2 - \eta_2(x_1)\big)^2 \mathbf{1}\big(\mathcal{U}(x_1,x_2) = 1\big) f(x_1,x_2)dx_1dx_2 + \int_{\mathbb{R}^2} \big(x_1 - \eta_1(x_2)\big)^2 \mathbf{1}\big(\mathcal{U}(x_1,x_2) = 2\big) f(x_1,x_2)dx_1dx_2$$

$$\mathcal{U}^\star_{\mathcal{E}}(x_1,x_2) = 1 \Longleftrightarrow \big(x_1 - \eta_1(x_2)\big)^2 \geq \big(x_2 - \eta_2(x_1)\big)^2$$

True for any joint PDF!

Vasconcelos & Mitra (*IEEE TCNS* 2019)

Optimization problem

$$(X_1, X_2) \sim f(x_1, x_2)$$

Generalized nearest neighbor condition

$$\mathcal{U}^{\star}_{\mathcal{E}}(x_1, x_2) = 1 \Longleftrightarrow \big(x_1 - \eta_1(x_2)\big)^2 \geq \big(x_2 - \eta_2(x_1)\big)^2$$

Infinite dimensional **nonconvex** optimization

Problem 1

$$\mathcal{J}(\eta_1, \eta_2) = \mathbf{E}\Big[\min\big\{\big(X_1 - \eta_1(X_2)\big)^2, \big(X_1 - \eta_2(X_1)\big)^2\big\}\Big]$$

Arbitrary joint density

$$(X_1, X_2) \sim f(x_1, x_2)$$

Generalized nearest neighbor condition

$$\mathcal{U}^{\star}_{\mathcal{E}}(x_1, x_2) = 1 \iff \big(x_1 - \eta_1(x_2)\big)^2 \geq \big(x_2 - \eta_2(x_1)\big)^2$$

$$\eta_1(x) = w_1 x + w_2$$
$$\eta_2(x) = w_3 x + w_4$$

Affine estimators

Finite dimensional **nonconvex** optimization

Problem 2

$$\mathcal{J}(\mathbf{w}) = \mathbf{E}\left[\min\left\{\big(X_1 - (w_1 X_2 + w_2)\big)^2, \big(X_2 - (w_3 X_1 + w_4)\big)^2\right\}\right]$$

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 \end{bmatrix}$$

Arbitrary joint density

$$(X_1, X_2) \sim f(x_1, x_2)$$

Generalized nearest neighbor condition

$$\mathcal{U}^{\star}_{\mathcal{E}}(x_1, x_2) = 1 \iff \left(x_1 - \eta_1(x_2)\right)^2 \geq \left(x_2 - \eta_2(x_1)\right)^2$$

$$\eta_1(x) = w_1 x + w_2$$
$$\eta_2(x) = w_3 x + w_4$$

Affine estimators

Finite dimensional **difference-of-convex** optimization

$$\begin{aligned}\mathcal{J}(\mathbf{w}) = \mathbf{E}\left[\left(X_1 - (w_1 X_2 + w_2)\right)^2 + \left(X_2 - (w_3 X_1 + w_4)\right)^2\right] \\ - \mathbf{E}\left[\max\left\{\left(X_1 - (w_1 X_2 + w_2)\right)^2, \left(X_2 - (w_3 X_1 + w_4)\right)^2\right\}\right]\end{aligned}$$

Difference-of-Convex (DoC) decomposition

$$\mathcal{J}(\mathbf{w}) = \mathbf{E}\left[\min\left\{\left(X_1 - (w_1 X_2 + w_2)\right)^2, \left(X_2 - (w_3 X_1 + w_4)\right)^2\right\}\right]$$

$$\mathcal{J}(\mathbf{w}) = \mathcal{F}(\mathbf{w}) - \mathcal{G}(\mathbf{w})$$

$$\mathcal{F}(\mathbf{w}) = \mathbf{E}\left[\left(X_1 - (w_1 X_2 + w_2)\right)^2 + \left(X_2 - (w_3 X_1 + w_4)\right)^2\right]$$

convex

$$\mathcal{G}(\mathbf{w}) = \mathbf{E}\left[\max\left\{\left(X_1 - (w_1 X_2 + w_2)\right)^2, \left(X_2 - (w_3 X_1 + w_4)\right)^2\right\}\right]$$

Convex-concave procedure

Heuristics to find local minimizers

$$\mathcal{J}(\mathbf{w}) = \mathcal{F}(\mathbf{w}) - \mathcal{G}(\mathbf{w})$$

$\mathbf{w}^{(k)}$

$$\mathcal{G}_{\text{affine}}(\mathbf{w}; \mathbf{w}^{(k)}) = \mathcal{G}(\mathbf{w}^{(k)}) + g(\mathbf{w}^{(k)})^{\mathsf{T}}(\mathbf{w} - \mathbf{w}^{(k)})$$

subgradient of $\mathcal{G}$

$$\mathbf{w}^{(k+1)} = \arg\min_{\mathbf{w}} \left\{ \mathcal{F}(\mathbf{w}) - \mathcal{G}_{\text{affine}}(\mathbf{w}; \mathbf{w}^{(k)}) \right\}$$

converges to a local minimum

convex

Yuille & Rangarajan - NIPS (2002)
Lipp & Boyd - Optim. Eng. (2016)

Convex-concave procedure

$$\mathbf{A}\mathbf{w}^{(k+1)} = g(\mathbf{w}^{(k)}) + \mathbf{b}$$

$$\mathbf{A} = 2 \cdot \begin{bmatrix} \mathbf{E}[X_2^2] & \mathbf{E}[X_2] & 0 & 0 \\ \mathbf{E}[X_2] & 1 & 0 & 0 \\ 0 & 0 & \mathbf{E}[X_1^2] & \mathbf{E}[X_1] \\ 0 & 0 & \mathbf{E}[X_1] & 1 \end{bmatrix}$$

$$\mathbf{b} = 2 \cdot \begin{bmatrix} \mathbf{E}[X_1 X_2] \\ \mathbf{E}[X_1] \\ \mathbf{E}[X_1 X_2] \\ \mathbf{E}[X_2] \end{bmatrix}$$

Step size

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \mathbf{A}^{-1} j(\mathbf{w}^{(k)})$$

↓

$$\lambda = \|A^{-1}\|_F$$

$$\mathbf{A} = 2 \cdot \begin{bmatrix} \mathbf{E}[X_2^2] & \mathbf{E}[X_2] & 0 & 0 \\ \mathbf{E}[X_2] & 1 & 0 & 0 \\ 0 & 0 & \mathbf{E}[X_1^2] & \mathbf{E}[X_1] \\ 0 & 0 & \mathbf{E}[X_1] & 1 \end{bmatrix}$$

$$\lambda = \frac{1}{2} \left(\sum_{i \in \{1,2\}} \frac{\sqrt{1 + 2\big(\mathbf{E}[X_i]\big)^2 + \big(\mathbf{E}[X_i^2]\big)^2}}{\mathrm{Var}(X_i)} \right)$$

Convex-concave procedure

$$\mathbf{A}\mathbf{w}^{(k+1)} = g(\mathbf{w}^{(k)}) + \mathbf{b}$$

computational bottleneck

$$g(\mathbf{w}) = -2 \cdot \mathbf{E} \begin{bmatrix} X_2(X_1 - w_1X_2 - w_2)\mathbf{1}\big(|X_1 - w_1X_2 - w_2| \geq |X_2 - w_3X_1 - w_4|\big) \\ (X_1 - w_1X_2 - w_2)\mathbf{1}\big(|X_1 - w_1X_2 - w_2| \geq |X_2 - w_3X_1 - w_4|\big) \\ X_1(X_2 - w_3X_1 - w_4)\mathbf{1}\big(|X_1 - w_1X_2 - w_2| < |X_2 - w_3X_1 - w_4|\big) \\ (X_2 - w_3X_1 - w_4)\mathbf{1}\big(|X_1 - w_1X_2 - w_2| < |X_2 - w_3X_1 - w_4|\big) \end{bmatrix}$$

we will deal with this later…

Relationship with subgradient methods

$$\text{minimize} \quad \mathcal{J}(\mathbf{w})$$

$$\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \alpha_k j(\mathbf{w}^{(k)})$$

subgradient of $\mathcal{J}$

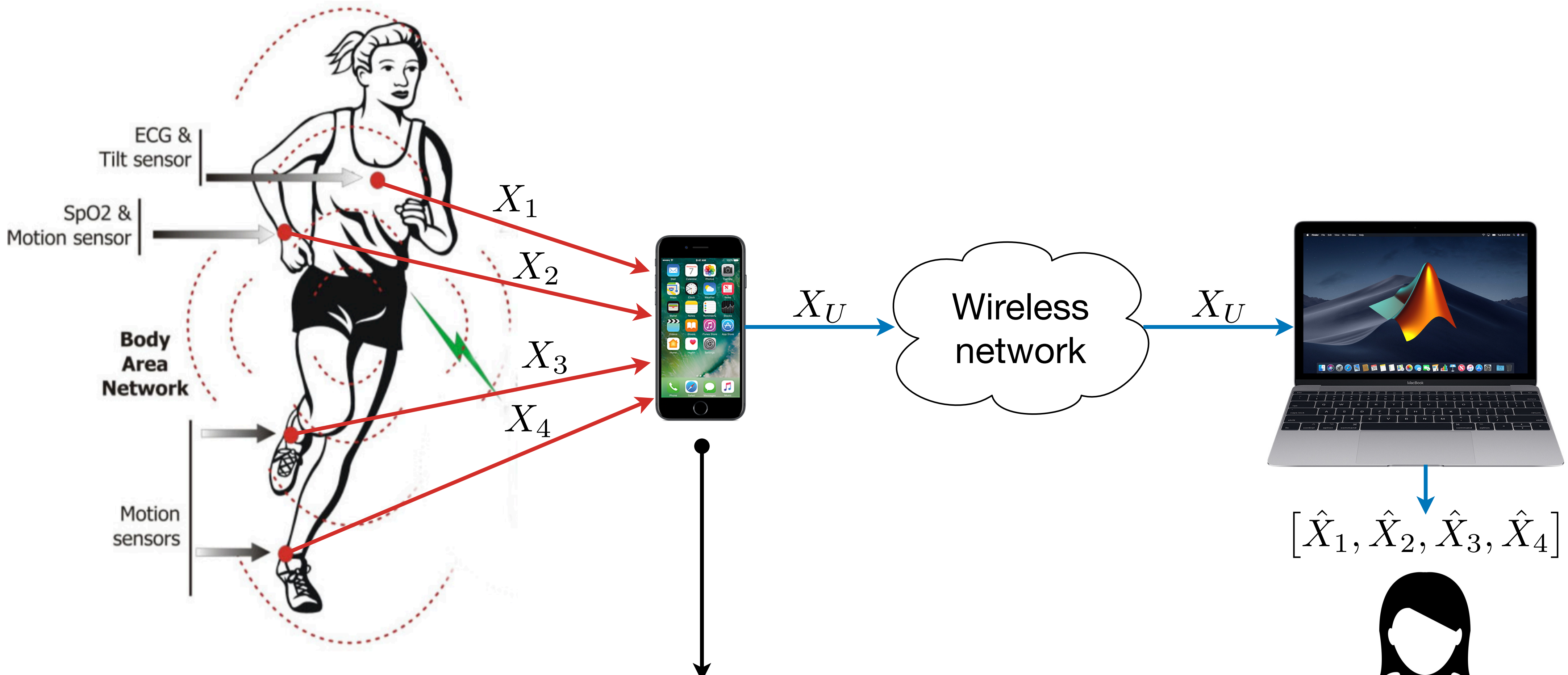
technical conditions needed for convergence

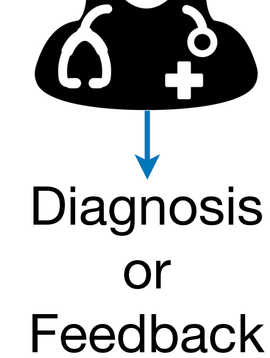
$$\mathbf{A}\mathbf{w}^{(k+1)} = g(\mathbf{w}^{(k)}) + \mathbf{b} \quad \longleftrightarrow \quad \mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} - \mathbf{A}^{-1} j(\mathbf{w}^{(k)})$$

guaranteed to converge!

However…

“In practice we don’t know the joint distribution”






```
> data
          X1           X2          X3           X4
1  0.2215982 -0.040513953 -0.11320355 -0.928056469
2  0.2612376 -0.001308288 -0.10454418 -0.977229008
3  0.2789176 -0.016137590 -0.11060182 -0.995759902
4  0.2773308 -0.017383819 -0.11114810 -0.283740259
5  0.2891883 -0.009918505 -0.10756619  0.030035338
6  0.2554617 -0.023953149 -0.09730200 -0.354708025
7  0.2813734 -0.018158740 -0.10724561 -0.974059465
8  0.2770874 -0.015687994 -0.10921827 -0.986822280
9  0.2779115 -0.018420827 -0.10590854 -0.987271889
10 0.2764266 -0.018594920 -0.10550036 -0.423642838
11 0.2776153 -0.022661416 -0.11681294  0.046366681
```

Data-driven scheduling

Unknown density

$$(X_1, X_2) \sim \,?$$

Cannot compute expectations

Data: $\mathcal{D} = \Big\{ \big(x_1(k), x_2(k)\big),\ k = 1, \cdots, N \Big\}$

Replace expectations by the **empirical mean**

$$\hat{\mathcal{F}}(\mathbf{w}) = \frac{1}{N} \sum_{(x_1, x_2) \in \mathcal{D}} \big(x_1 - w_1 x_2 - w_2\big)^2 + \big(x_2 - w_3 x_1 - w_4\big)^2$$

$$\hat{\mathcal{G}}(\mathbf{w}) = \frac{1}{N} \sum_{(x_1, x_2) \in \mathcal{D}} \max\Big\{ \big(x_1 - w_1 x_2 - w_2\big)^2, \big(x_2 - w_3 x_1 - w_4\big)^2 \Big\}$$

Approximate convex-concave procedure

$$\hat{\mathcal{J}}(\mathbf{w}) = \hat{\mathcal{F}}(\mathbf{w}) - \hat{\mathcal{G}}(\mathbf{w})$$

empirical risk

$\mathbf{w}^{(k)}$

$$\hat{\mathcal{G}}_{\text{affine}}(\mathbf{w}; \mathbf{w}^{(k)}) = \hat{\mathcal{G}}(\mathbf{w}^{(k)}) + \hat{g}(\mathbf{w}^{(k)})^{\mathsf{T}}(\mathbf{w} - \mathbf{w}^{(k)})$$

subgradient of $\hat{\mathcal{G}}$

$$\mathbf{w}^{(k+1)} = \arg\min_{\mathbf{w}} \left\{ \hat{\mathcal{F}}(\mathbf{w}^{(k)}) - \hat{\mathcal{G}}_{\text{affine}}(\mathbf{w}, \mathbf{w}^{(k)}) \right\}$$

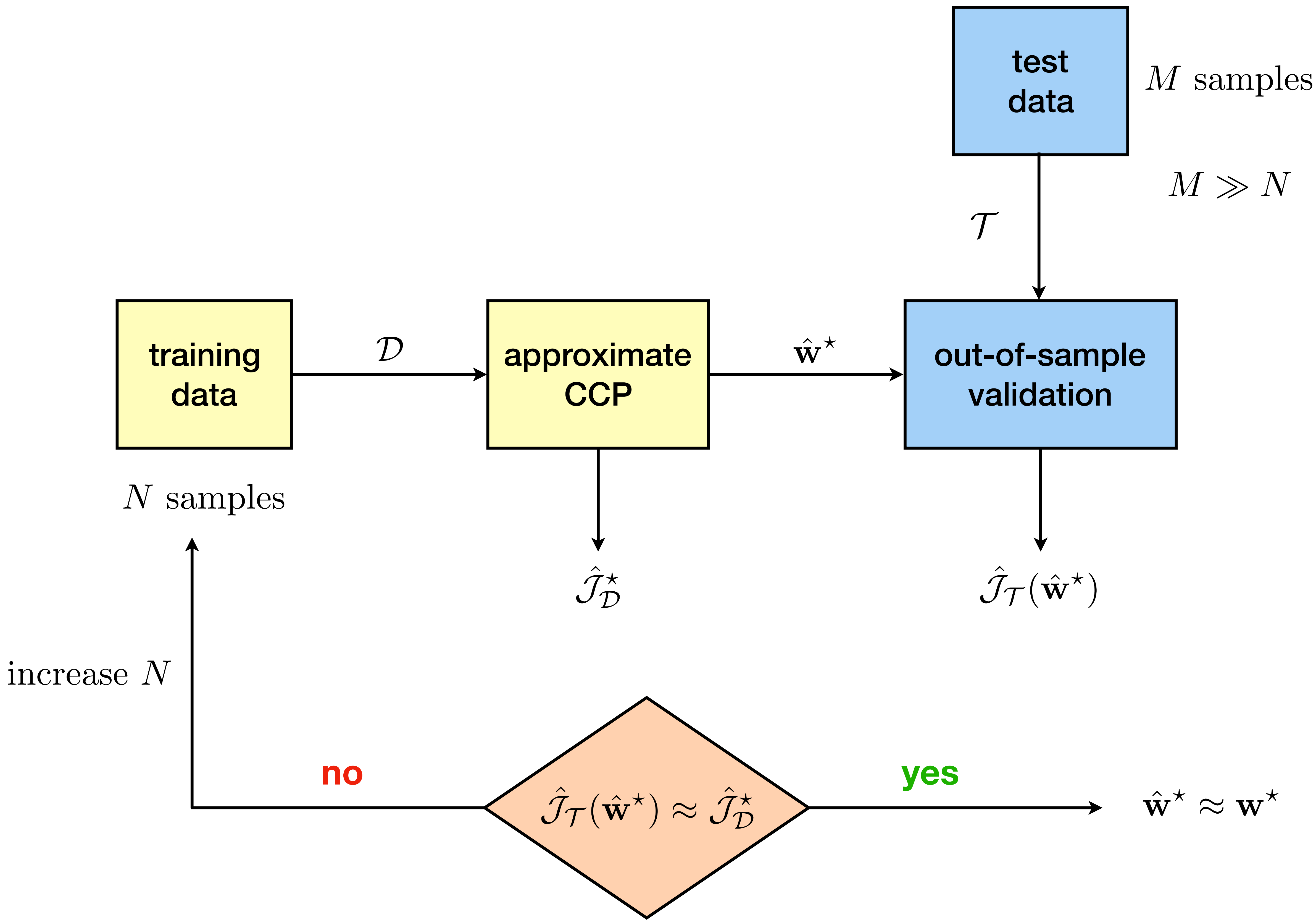
$$\mathbf{w}^{(k)} \longrightarrow \hat{\mathbf{w}}^{\star}$$

a local minimum of the empirical risk

$$\hat{g}(\mathbf{w}) = -2 \cdot \frac{1}{N} \sum_{(x_1, x_2) \in \mathcal{D}} \begin{bmatrix} x_2(x_1 - w_1 x_2 - w_2)\mathbf{1}\left(|x_1 - w_1 x_2 - w_2| \geq |x_2 - w_3 x_1 - w_4|\right) \\ (x_1 - w_1 x_2 - w_2)\mathbf{1}\left(|x_1 - w_1 x_2 - w_2| \geq |x_2 - w_3 x_1 - w_4|\right) \\ x_1(x_2 - w_3 x_2 - w_4)\mathbf{1}\left(|x_1 - w_1 x_2 - w_2| < |x_2 - w_3 x_1 - w_4|\right) \\ (x_2 - w_3 x_2 - w_4)\mathbf{1}\left(|x_1 - w_1 x_2 - w_2| < |x_2 - w_3 x_1 - w_4|\right) \end{bmatrix}$$

easy to compute

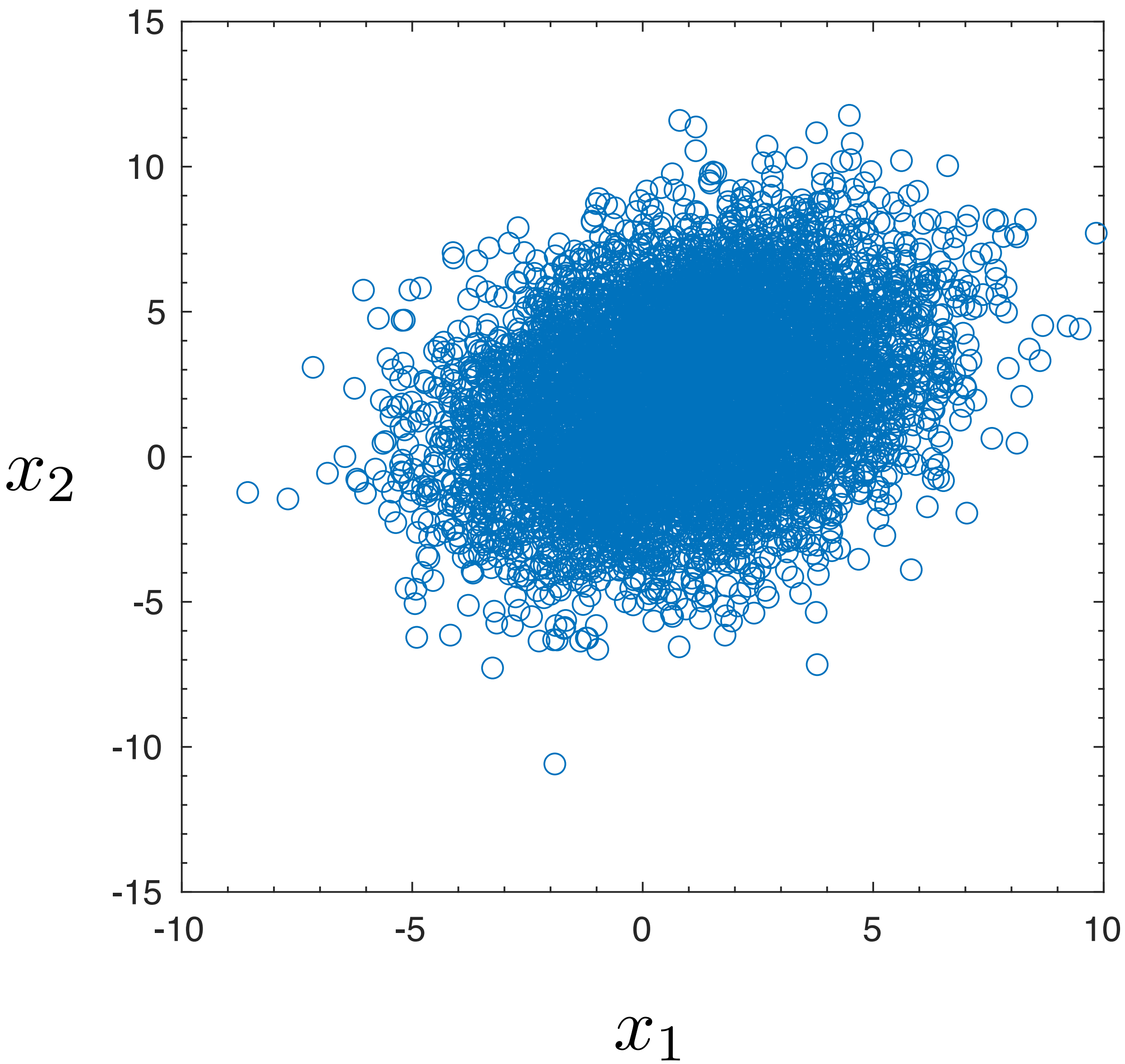
Learning the optimal scheduler



Gaussian data

$$(X_1, X_2) \sim \mathcal{N}\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}5 & 1.7748\\1.7748 & 7\end{bmatrix}\right)$$

$$\mathcal{D} = \Big\{\big(x_1(k), x_2(k)\big),\ k = 1, \cdots, N\Big\}$$



Exact CCP

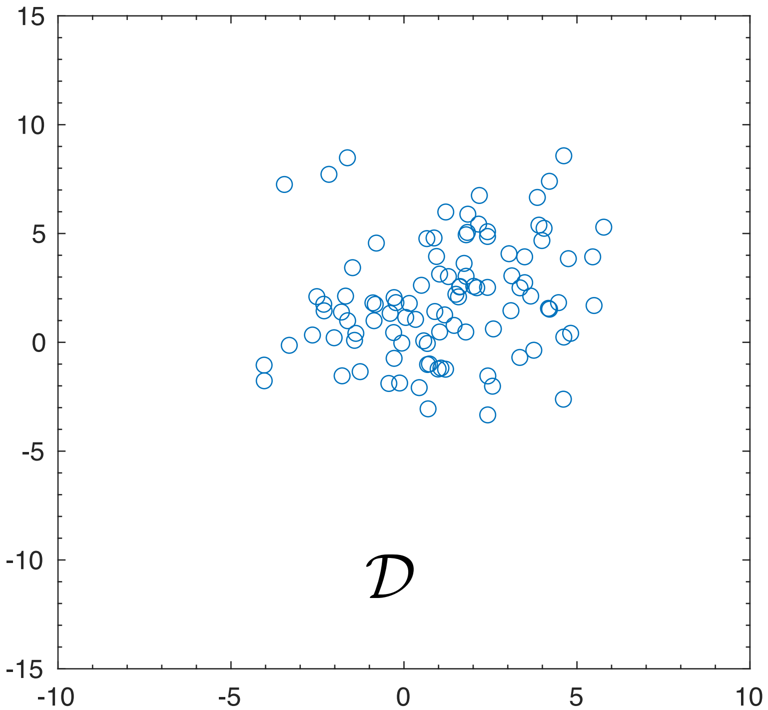
$$\mathcal{J}^\star = 1.9704$$

$$\mathbf{w}^\star = \begin{bmatrix}0.1691\\0.6638\\0.2043\\1.8152\end{bmatrix}$$

…it takes a long time to compute

100 data points

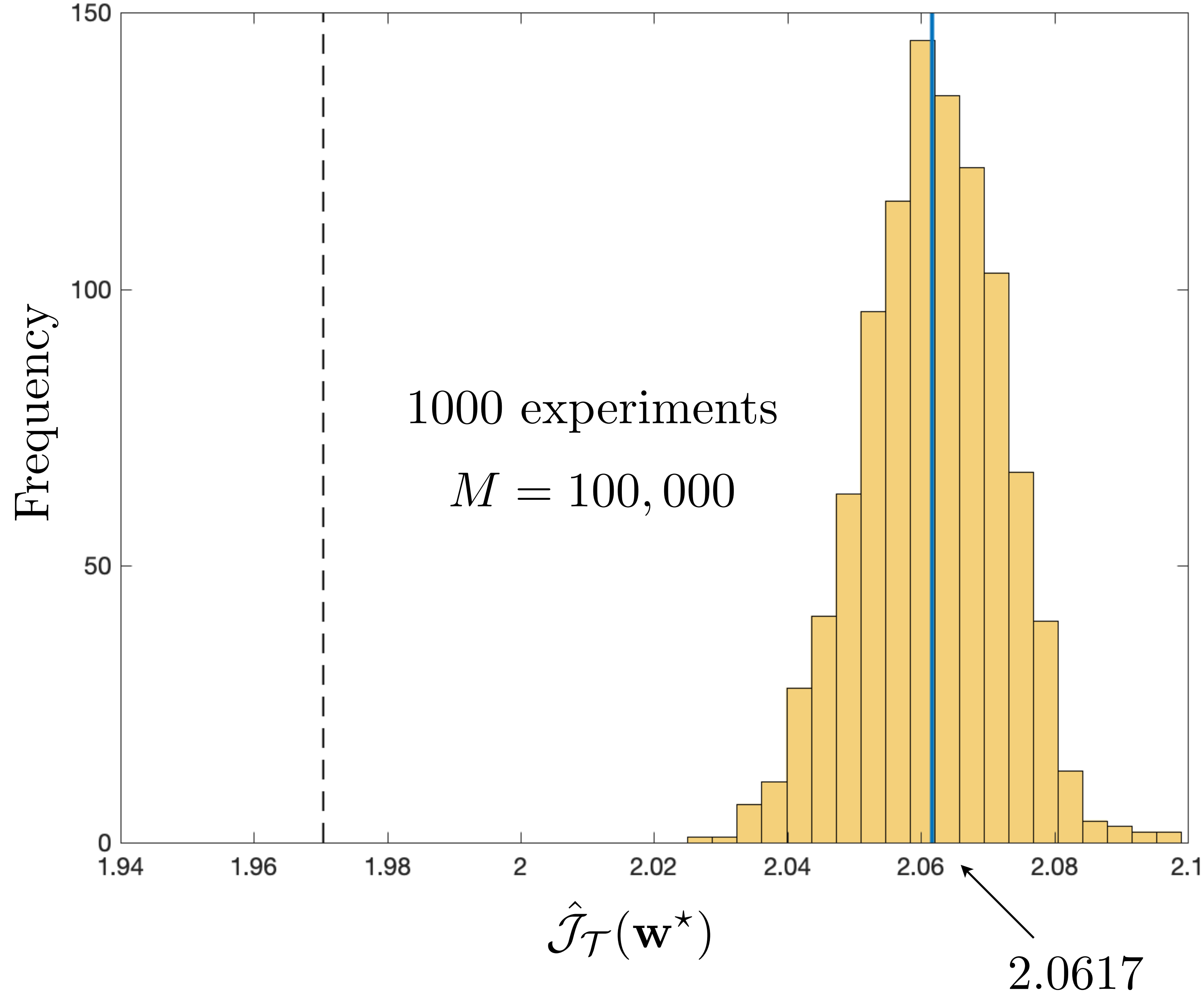
training



$$\mathbf{w}^{\star} = \begin{bmatrix} 0.0022 \\ 1.0100 \\ 0.2541 \\ 1.8854 \end{bmatrix}$$

$$\hat{\mathcal{J}}_{\mathcal{D}}(\hat{w}^{\star}) = 1.5979$$

out-of-sample validation

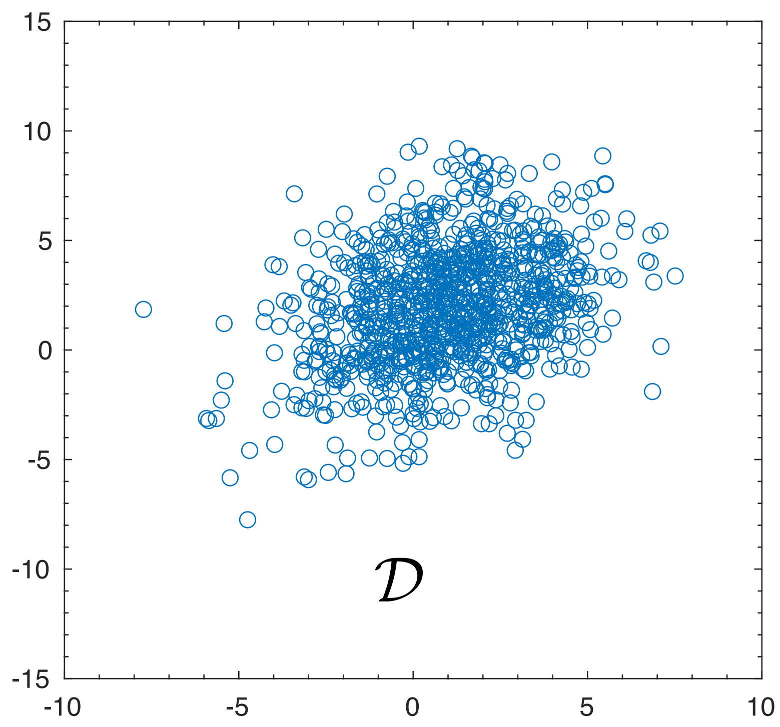


$$\left|\hat{\mathcal{J}}_{\mathcal{T}}(\hat{w}^{\star}) - \hat{\mathcal{J}}_{\mathcal{D}}(\hat{w}^{\star})\right| = 0.4638$$

not enough data!

1,000 data points

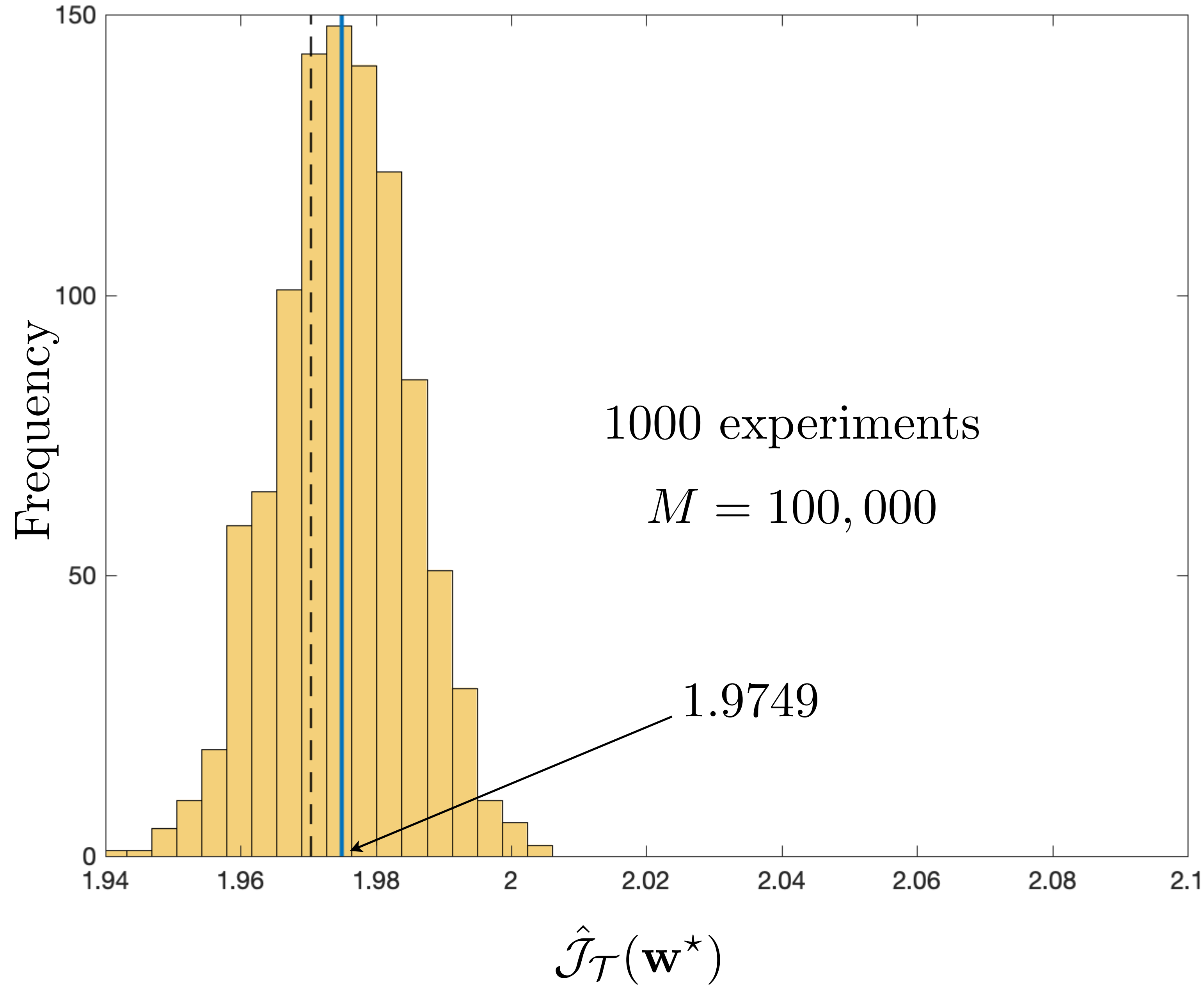
training



$$\mathbf{w}^\star = \begin{bmatrix} 0.1400 \\ 0.6405 \\ 0.2360 \\ 1.8333 \end{bmatrix}$$

$$\hat{\mathcal{J}}_{\mathcal{D}}(\hat{w}^\star) = 1.9205$$

out-of-sample validation

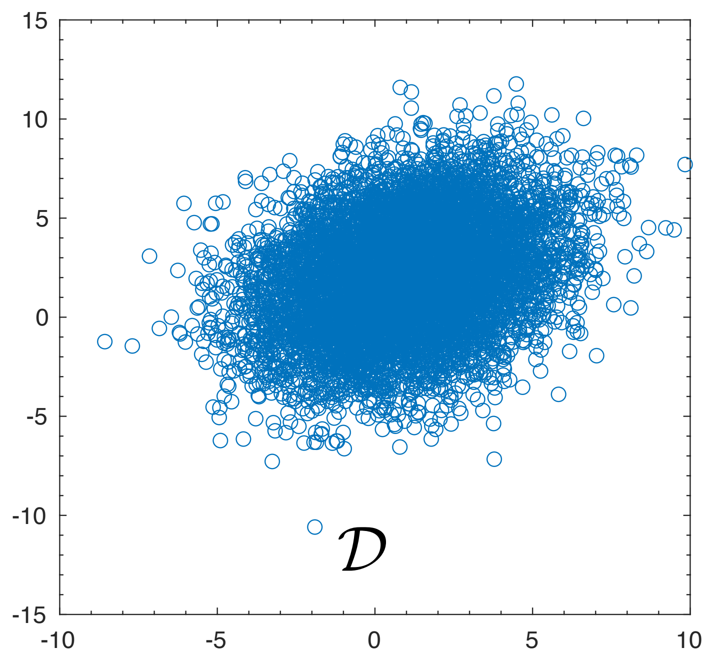


$$\left|\hat{\mathcal{J}}_{\mathcal{T}}(\hat{w}^\star) - \hat{\mathcal{J}}_{\mathcal{D}}(\hat{w}^\star)\right| = 0.0544$$

good, but not quite there yet

10,000 data points

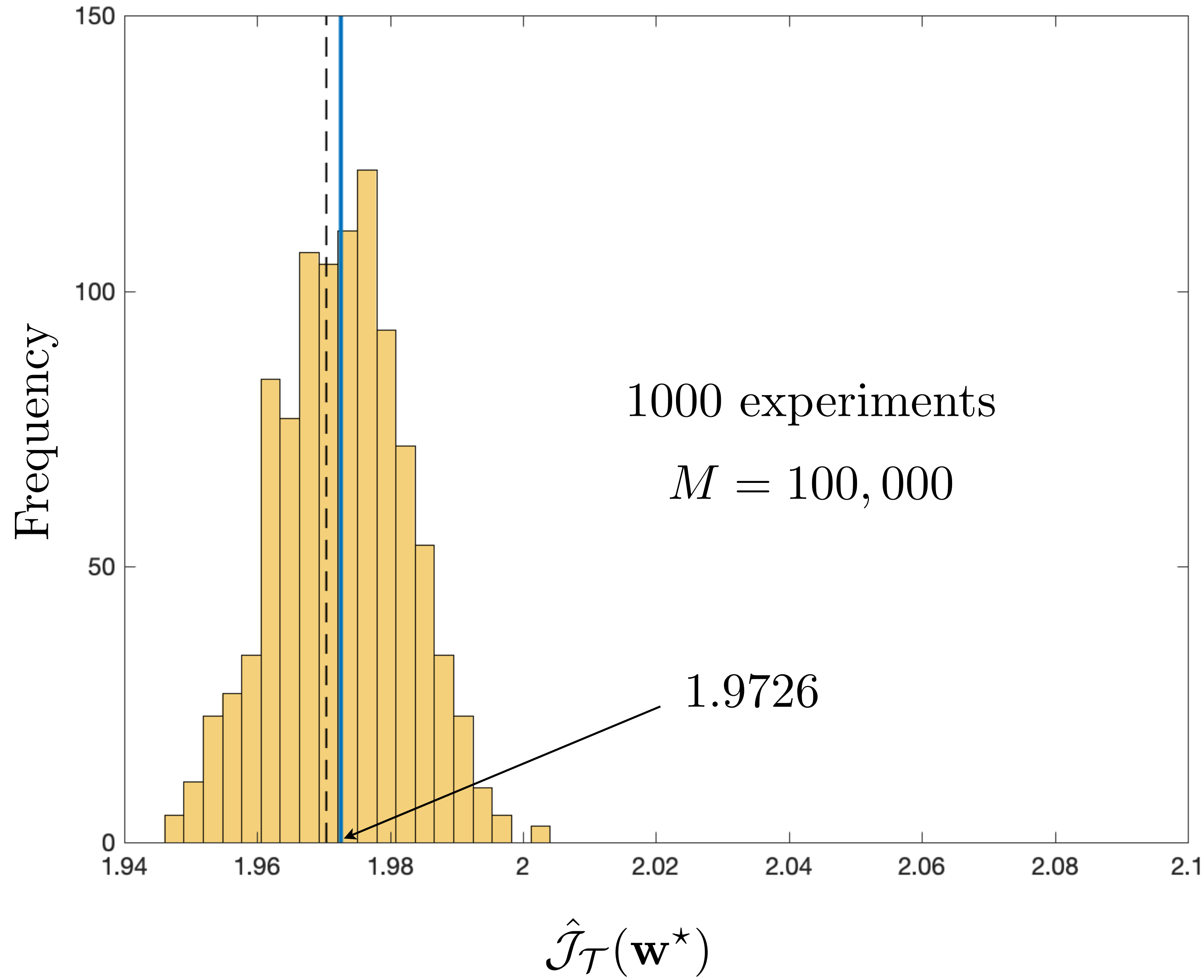
training



$$\mathbf{w}^\star = \begin{bmatrix} 0.1490 \\ 0.7457 \\ 0.2392 \\ 1.8130 \end{bmatrix}$$

$$\hat{\mathcal{J}}_{\mathcal{D}}(\hat{w}^\star) = 1.9714$$

out-of-sample validation

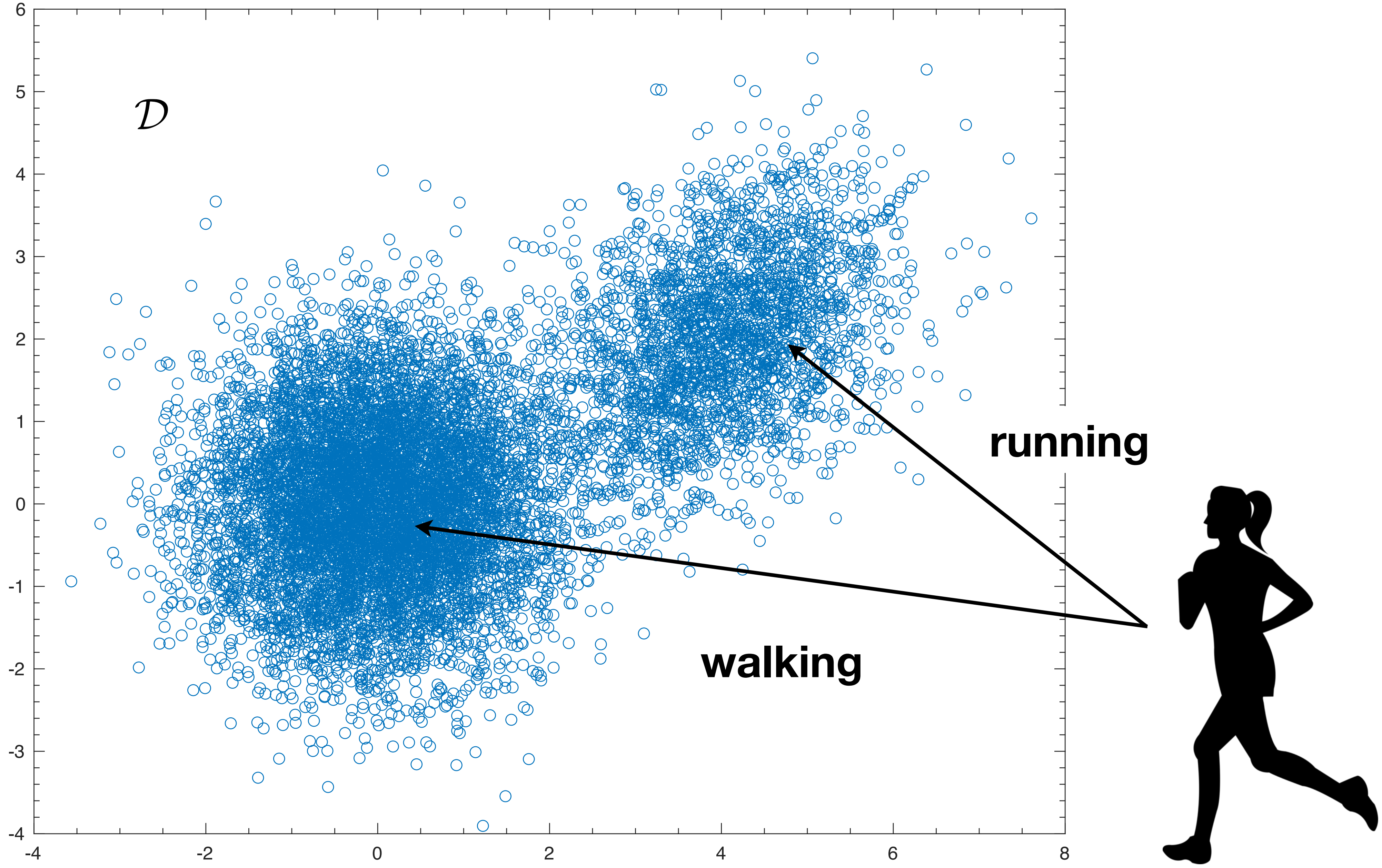


$$\left|\hat{\mathcal{J}}_{\mathcal{T}}(\hat{w}^\star) - \hat{\mathcal{J}}_{\mathcal{D}}(\hat{w}^\star)\right| = 0.0012$$

learning was successful

Gaussian mixture data

$$(X_1, X_2) \sim 0.75 \cdot \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) + 0.25 \cdot \mathcal{N}\left(\begin{bmatrix}4\\2\end{bmatrix}, \begin{bmatrix}1 & 0.4\\0.4 & 1\end{bmatrix}\right)$$



Gaussian mixture data

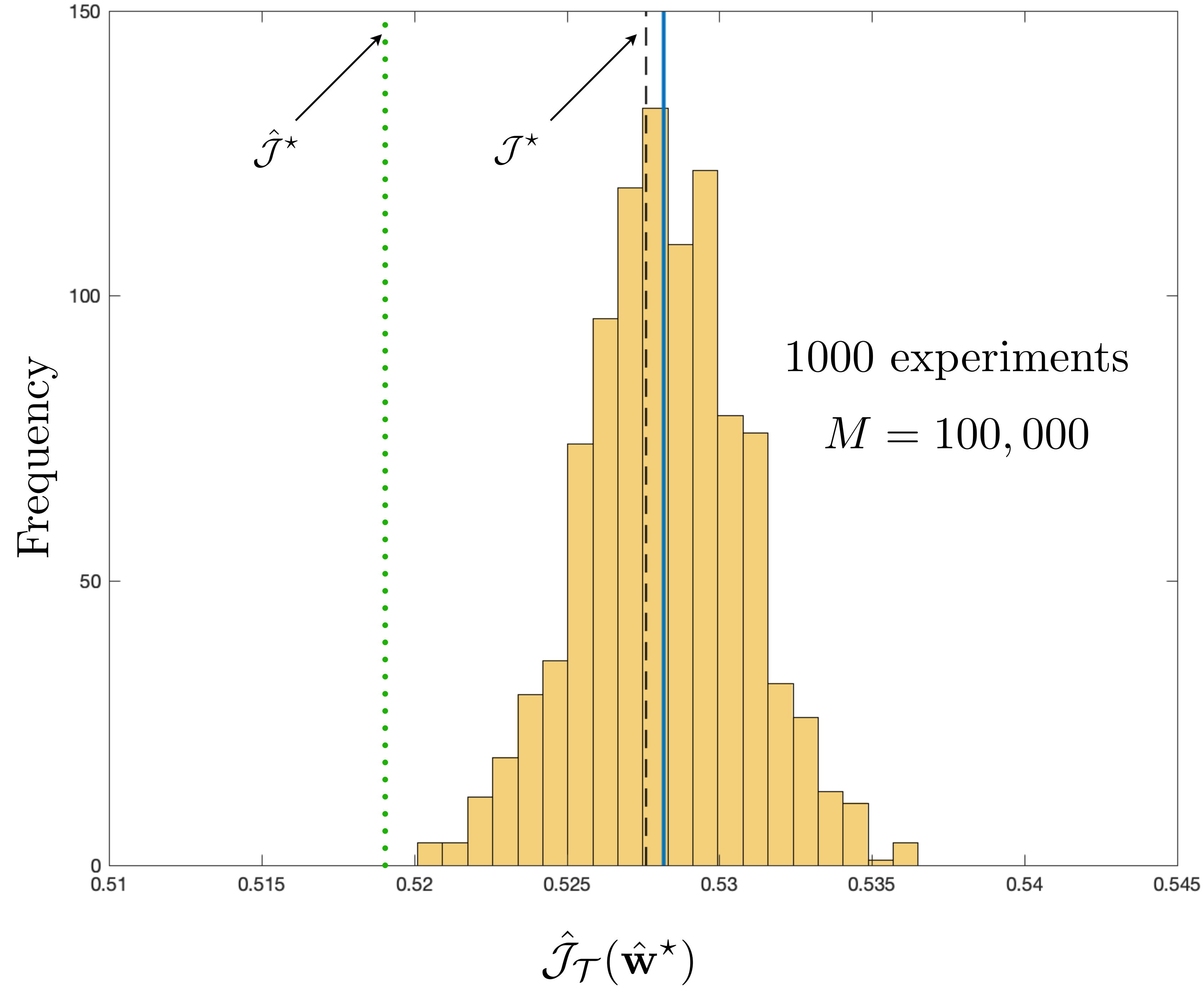
training

$N = 10,000$ samples

$$\hat{\mathbf{w}}^\star = \begin{bmatrix} -0.2218 \\ +0.4089 \\ +0.0644 \\ +0.2485 \end{bmatrix}$$

$$\hat{\mathcal{J}}_{\mathcal{D}}(\hat{\mathbf{w}}^\star) = 0.5186$$

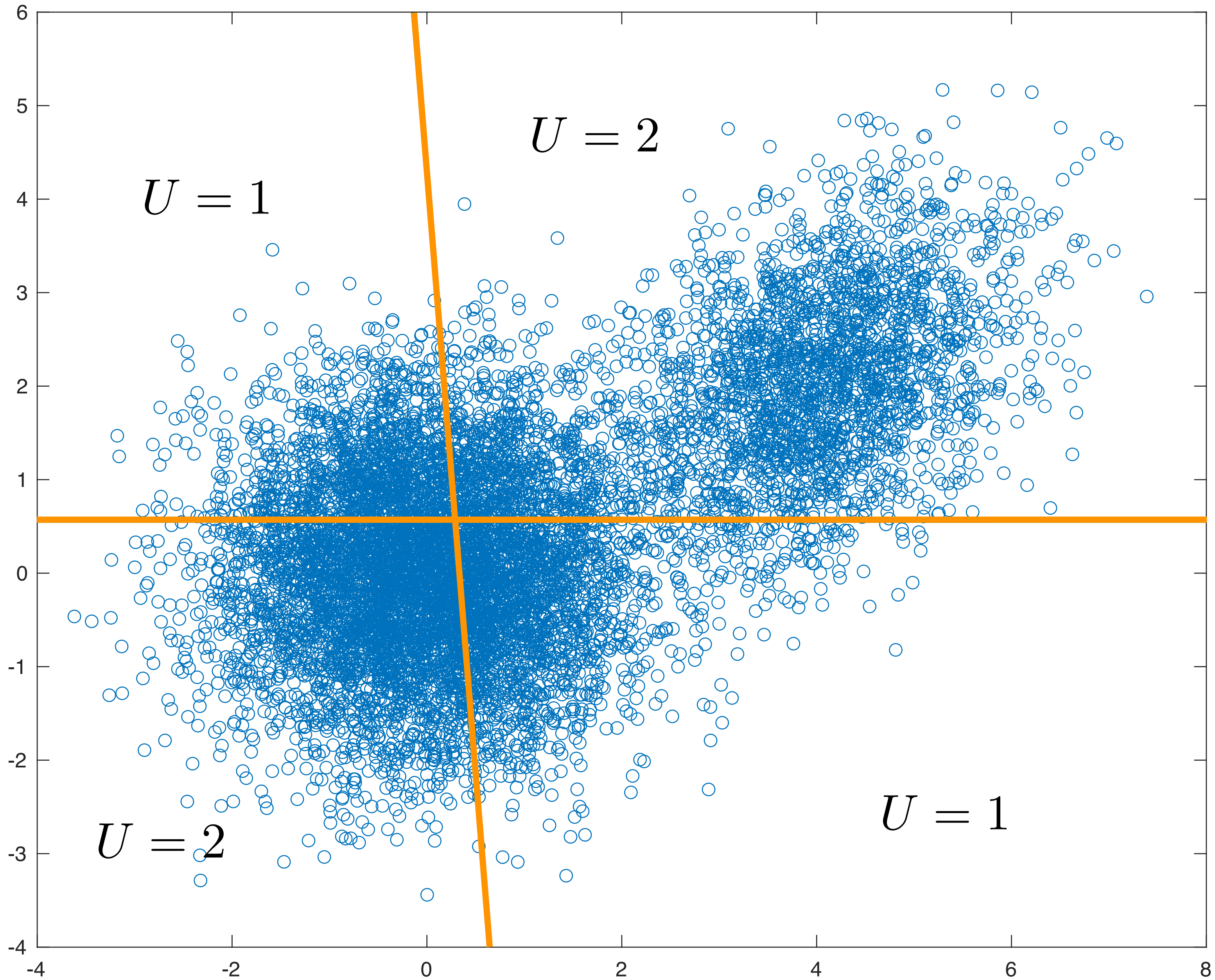
out-of-sample validation



$$\left|\mathbf{E}\left[\hat{\mathcal{J}}_{\mathcal{T}}(\hat{\mathbf{w}}^\star)\right] - \hat{\mathcal{J}}_{\mathcal{D}}(\hat{\mathbf{w}}^\star)\right| = 0.0096$$

$$\hat{\mathbf{w}}^\star \approx \mathbf{w}^\star$$

Data-driven scheduler



Summary

Fundamentals of data-driven scheduling for sensors

with unknown probabilistic models

Combined ideas from:

1. **Quantization theory (nearest neighbor condition)**
2. **Nonconvex optimization (DoC, CCP)**

Our algorithm:

is guaranteed to converge

works with data samples drawn from any joint PDF

Future work

1. Approximate the optimal estimators using Neural Networks

Can we find NNs such that the DoC decomposition holds?

2. Online learning of optimal data-driven schedulers

data samples arrive one at a time

3. Distributionally robust data-driven optimal scheduling

$$\min_{\mathbf{w}} \left\{ \max_{f_X \in \Pi(\mathcal{D})} \mathcal{J}(\mathbf{w}) \right\}$$

Shalev-Shwartz - *Found. Trends in ML* 2012
Delage & Ye - *Op. Res.* 2010

Relevant papers

(1) **Data-driven sensor scheduling**, M. Vasconcelos and U. Mitra, in preparation. **Arxiv version soon**.

(2) **Observation-driven scheduling for remote estimation of two Gaussian random variables**, M. Vasconcelos and U. Mitra, IEEE TCNS 2019. **Available on Arxiv**.

(3) **Optimal scheduling strategy for networked estimation with energy harvesting**, M. Vasconcelos, M. Gagrani, A. Nayyar and U. Mitra, Submitted to IEEE TCNS 2019. **Available on Arxiv**.

Thank you!

`mvasconc@usc.edu`