Introduction to Uncertainty Quantification for Bayesian Inverse Problems

Babak Maboudi - day 1 - Jyväskylä summer school 2025

What are inverse problems?

- When we want to understand hidden causes from indirect measurements.
- It is best understood by examples!

Examples of Inverse Problems

X-ray Computed Tomography (CT) or CAT scan

Anna Bertha Ludwig's hand

X-ray by Wilhelm Röntgen

1895

First X-ray image in space

2025

Examples of Inverse Problems

X-ray Computed Tomography (CT) or CAT scan

X-ray radiography

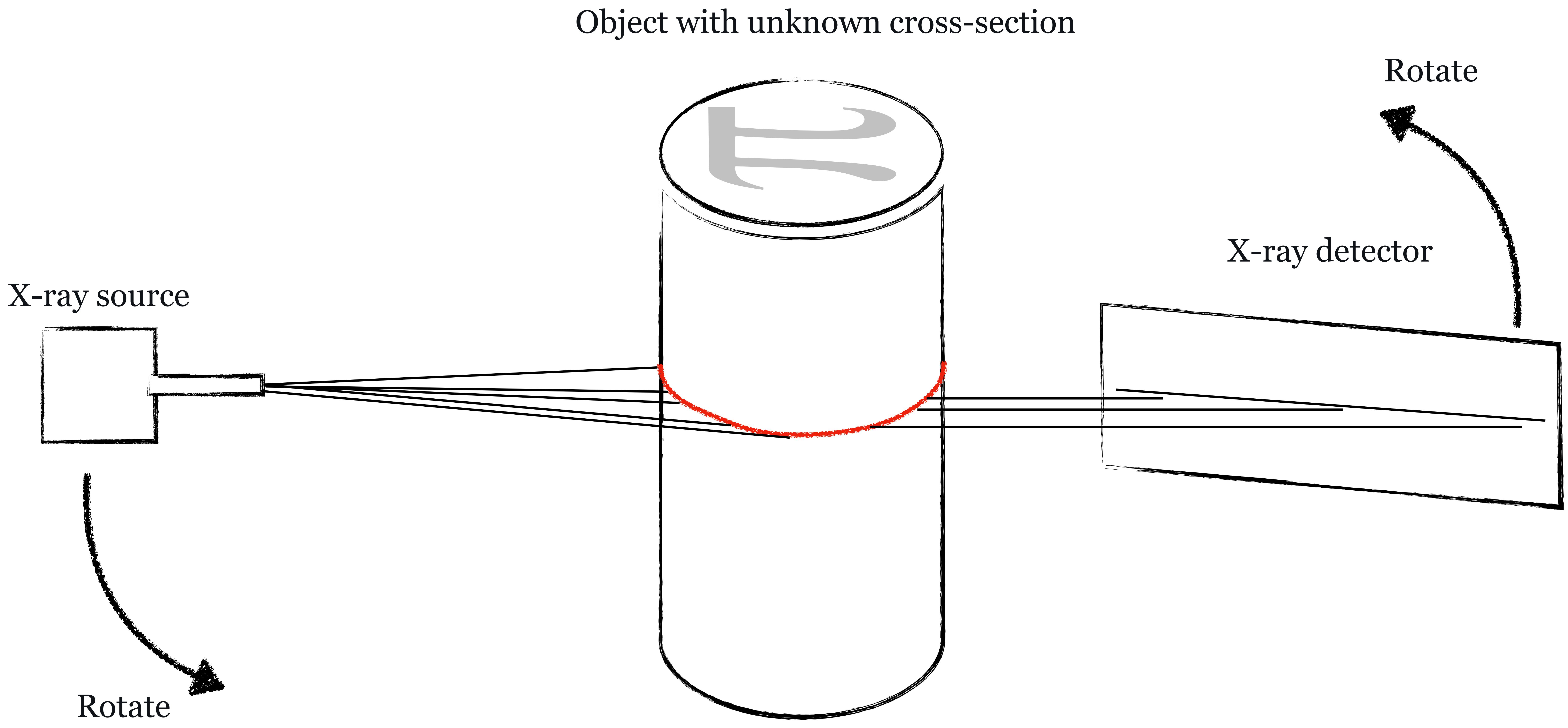
One X-ray image

X-ray computed tomography (CT)

Sequence of X-ray images

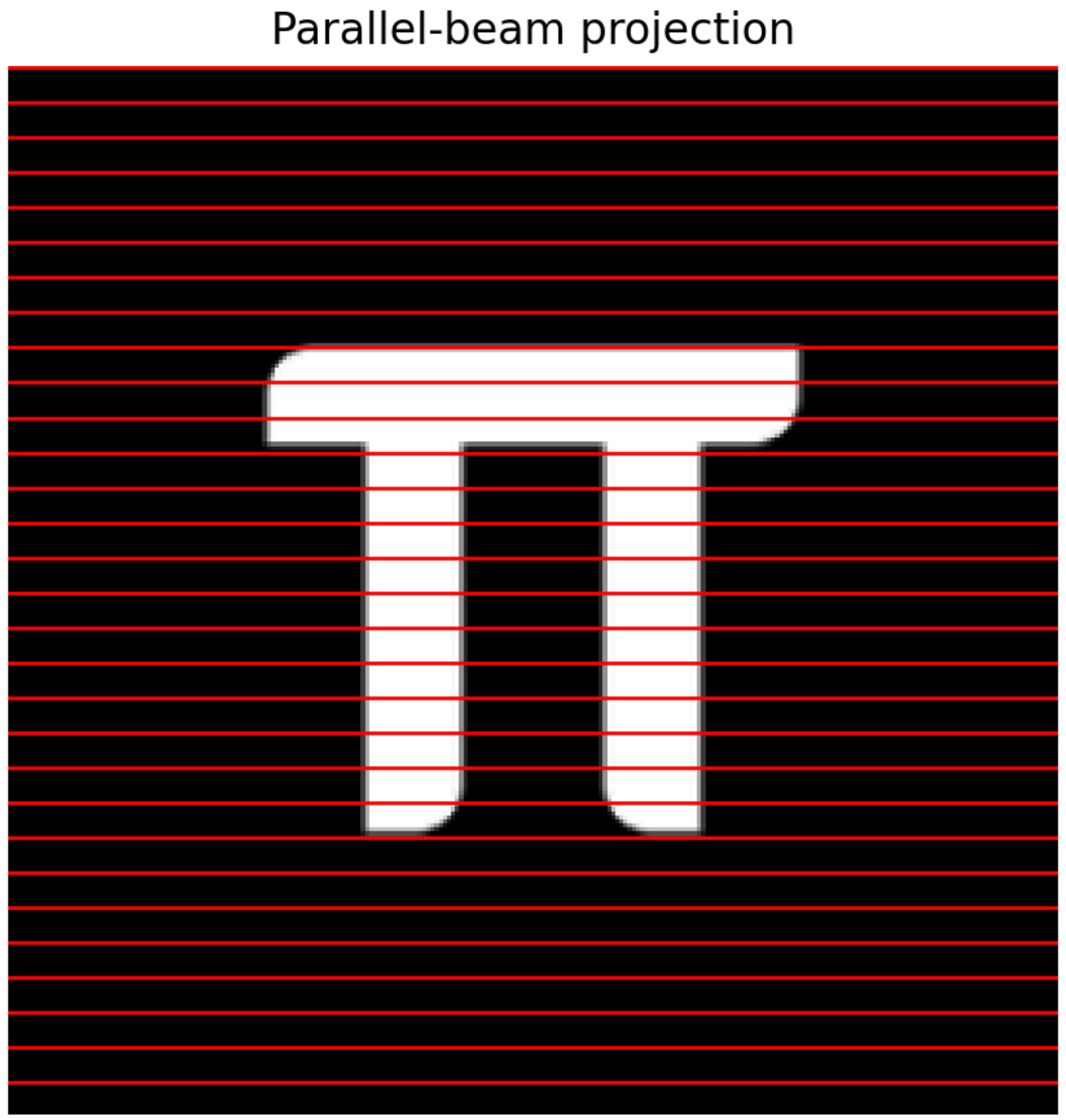
Examples of Inverse Problems

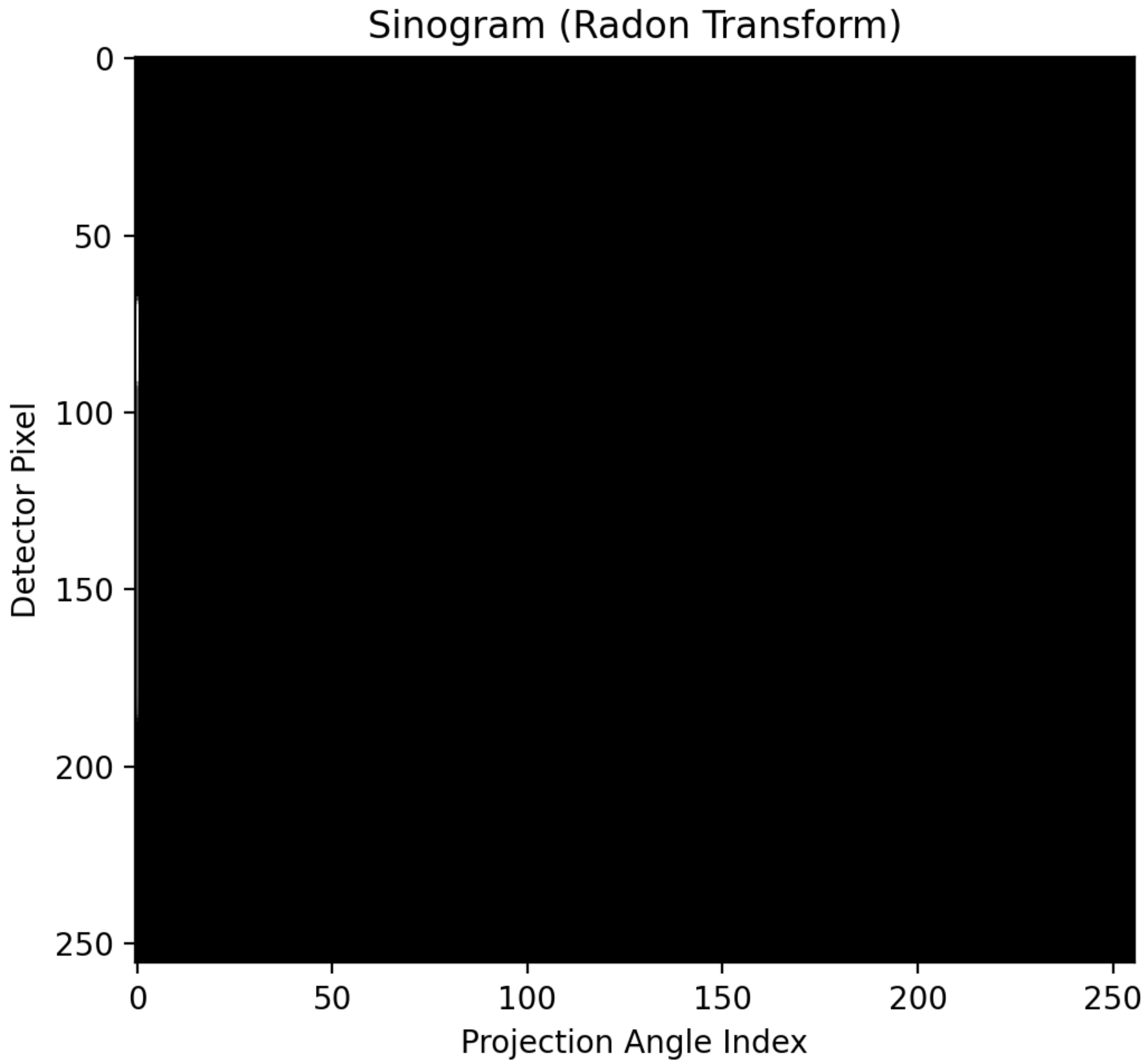
X-ray CT, a 2D example



Examples of Inverse Problems

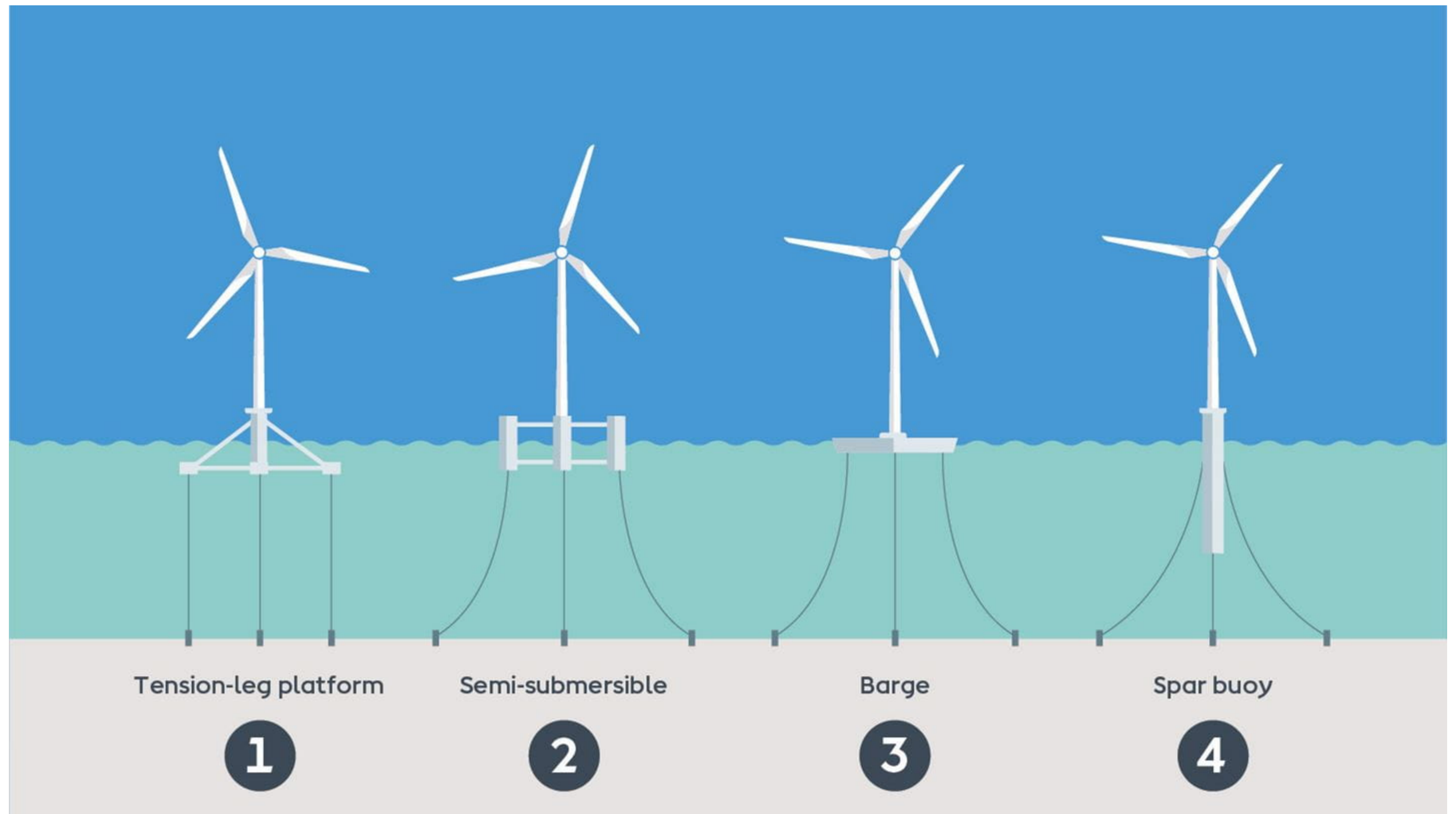
X-ray CT, a 2D example

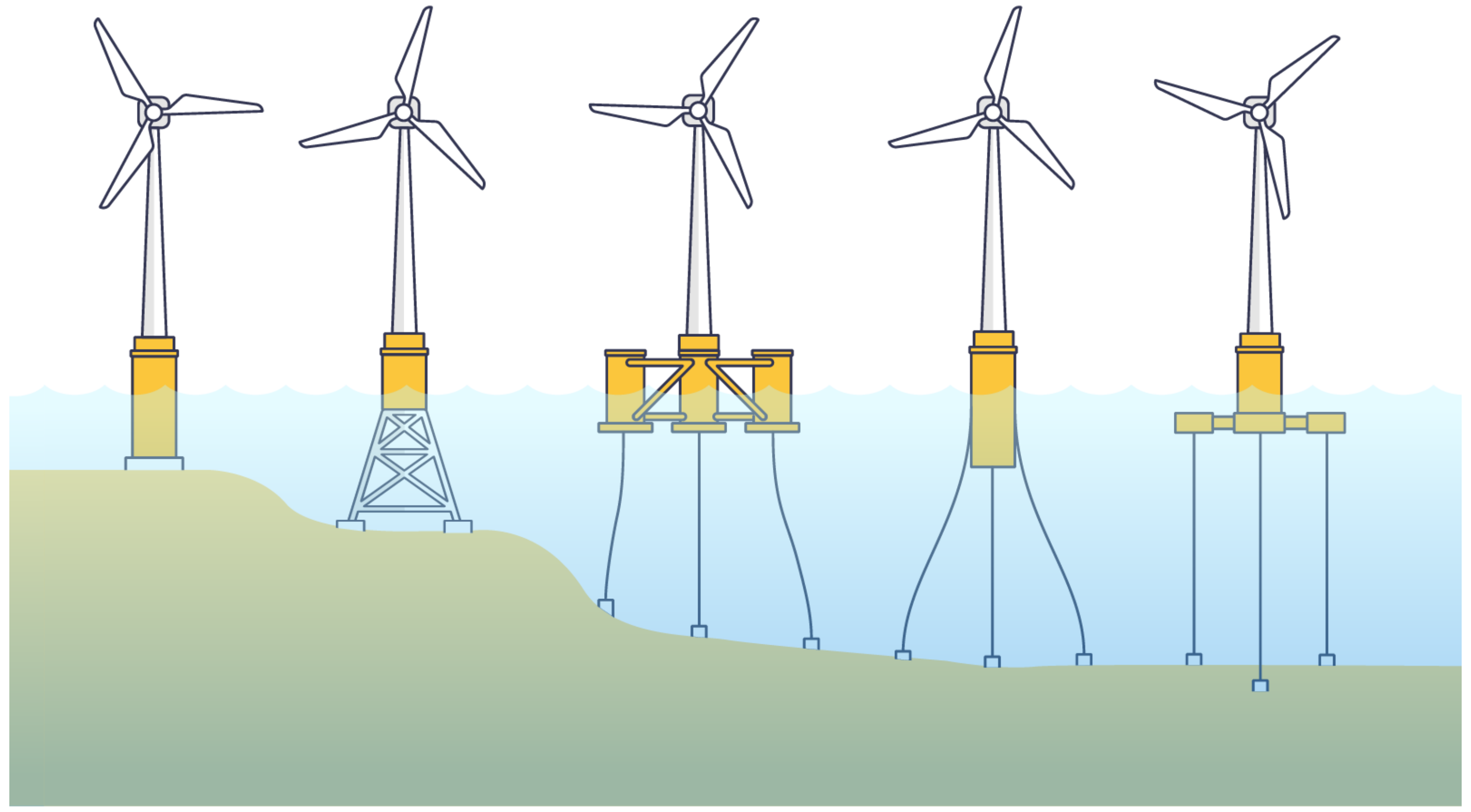




Examples of Inverse Problems

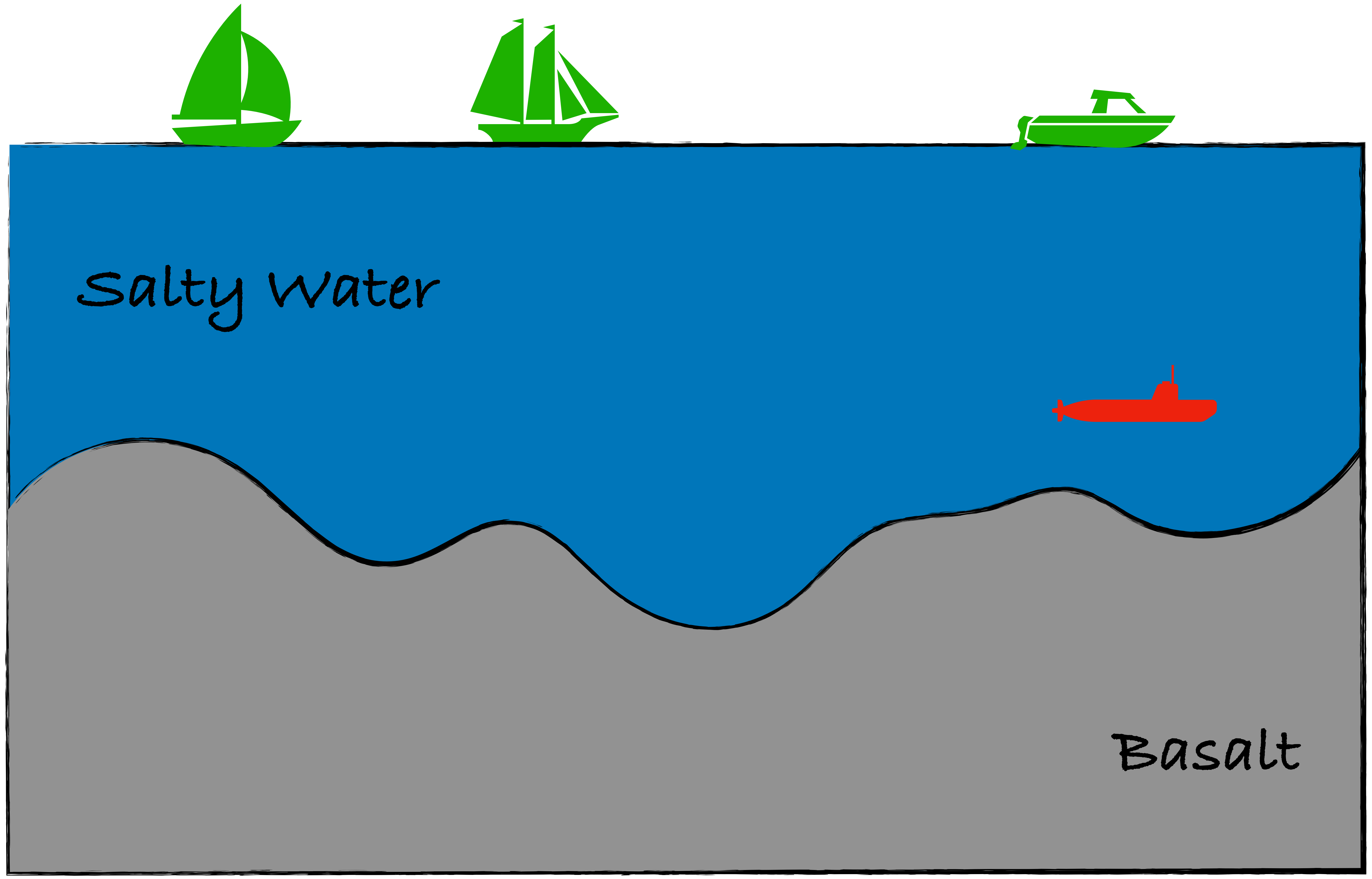
Ocean Floor Detection/Exploration





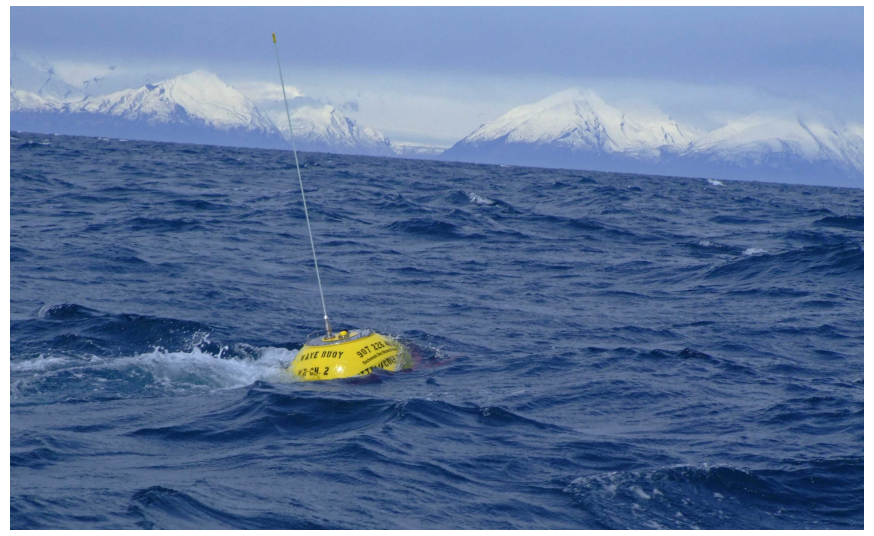
Examples of Inverse Problems

Ocean Floor Detection/Exploration

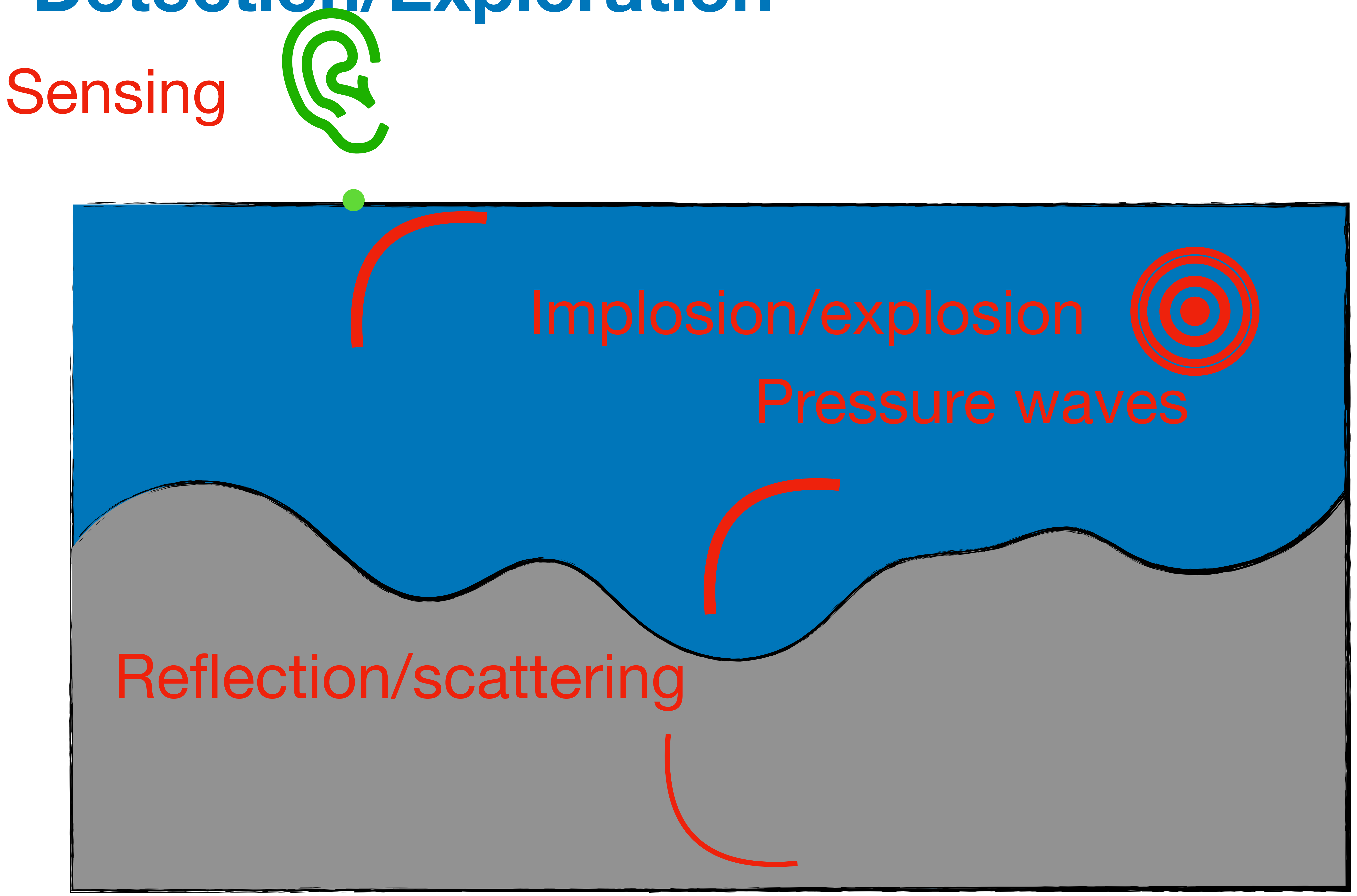


Examples of Inverse Problems

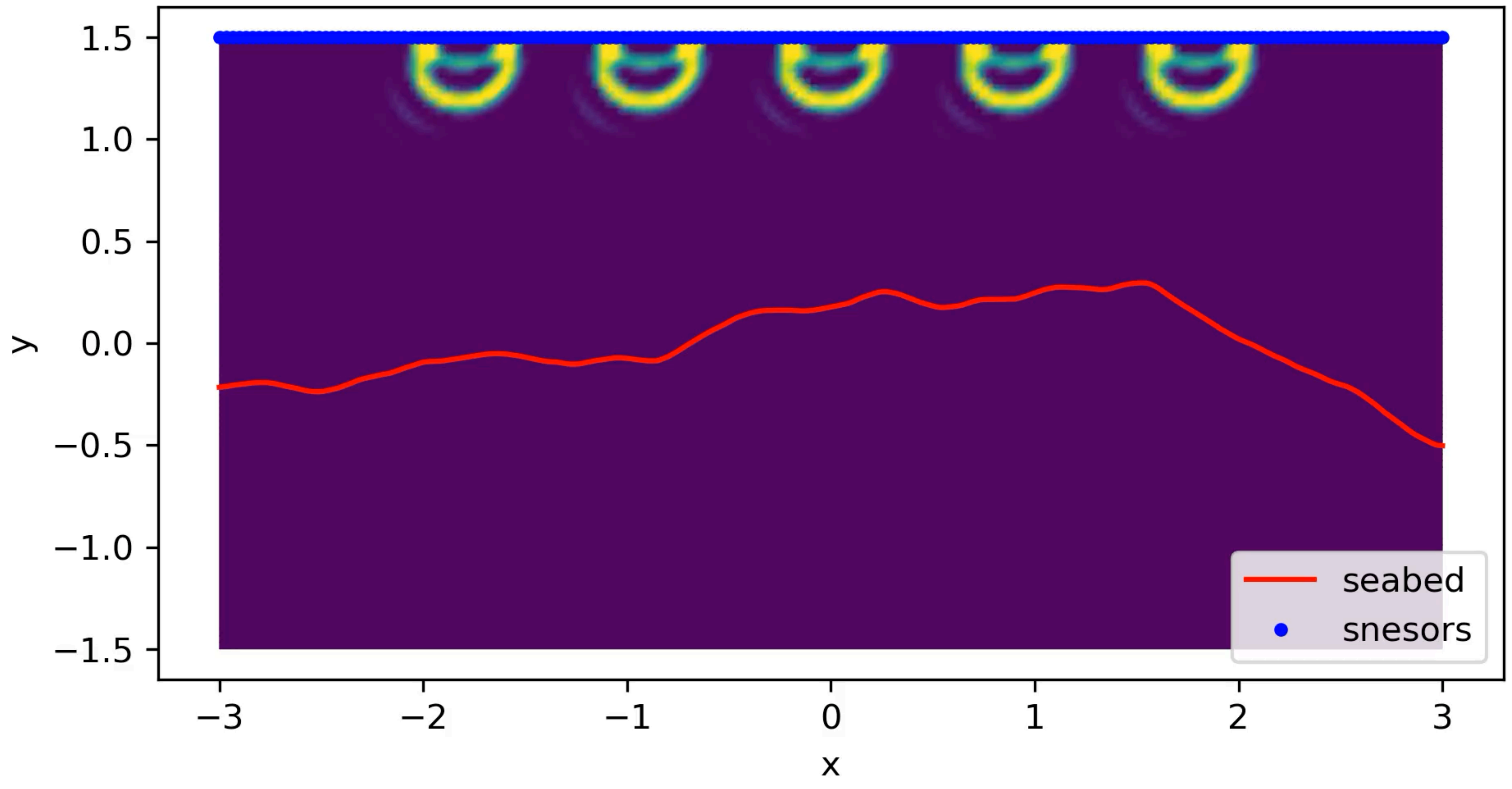
Ocean Floor Detection/Exploration



Wave Buoys

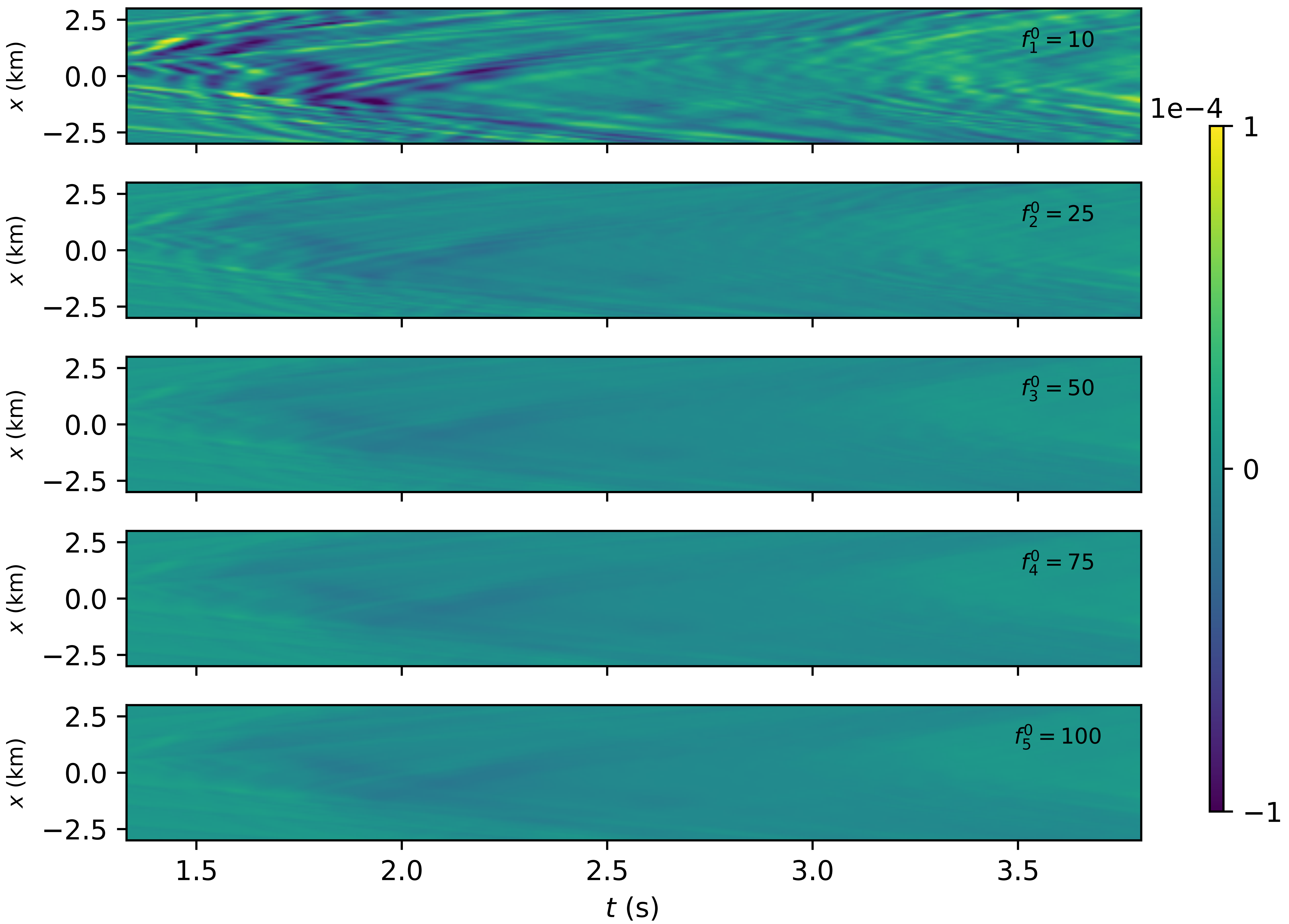


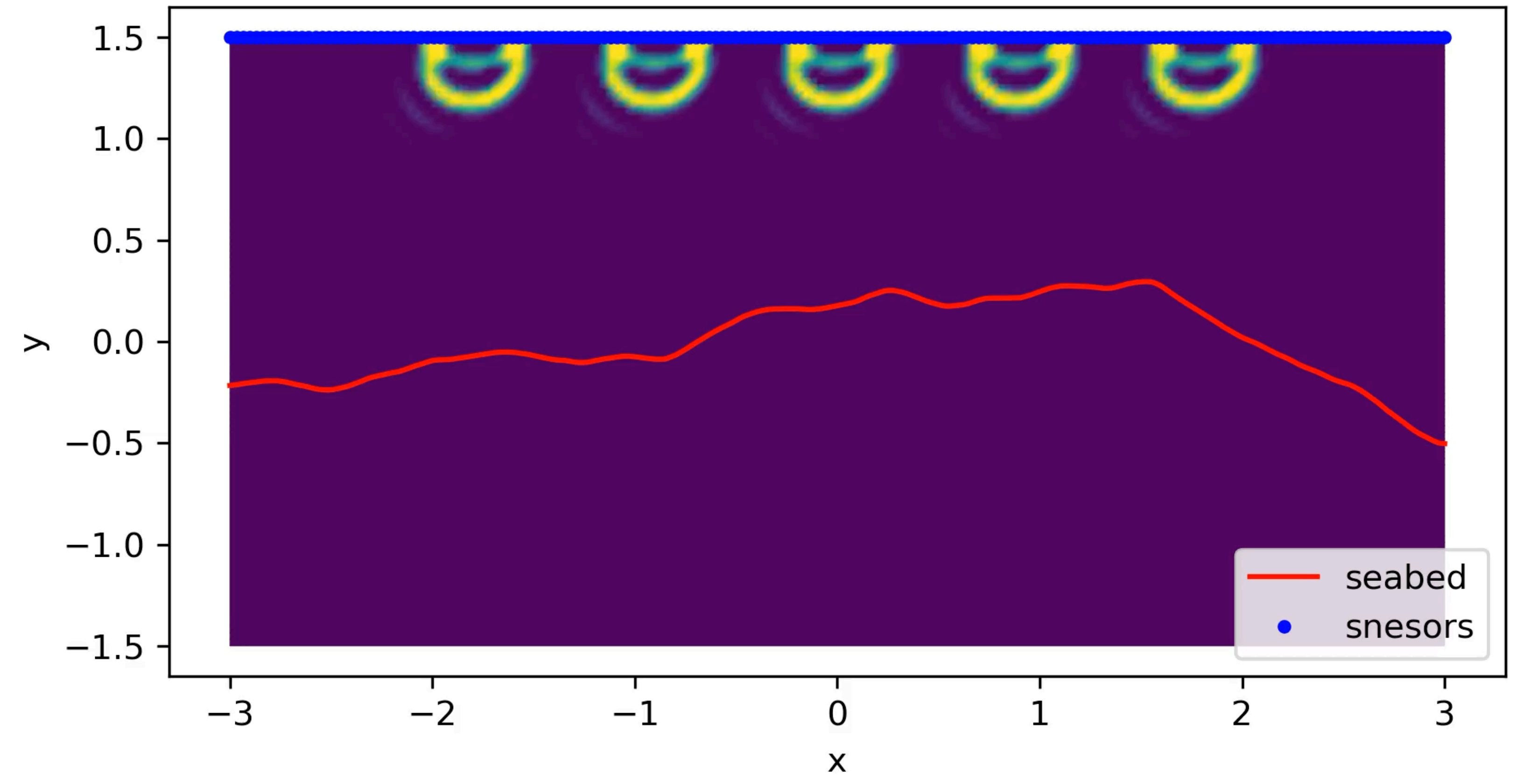
Examples of Inverse Problems
Ocean Floor Detection/Exploration

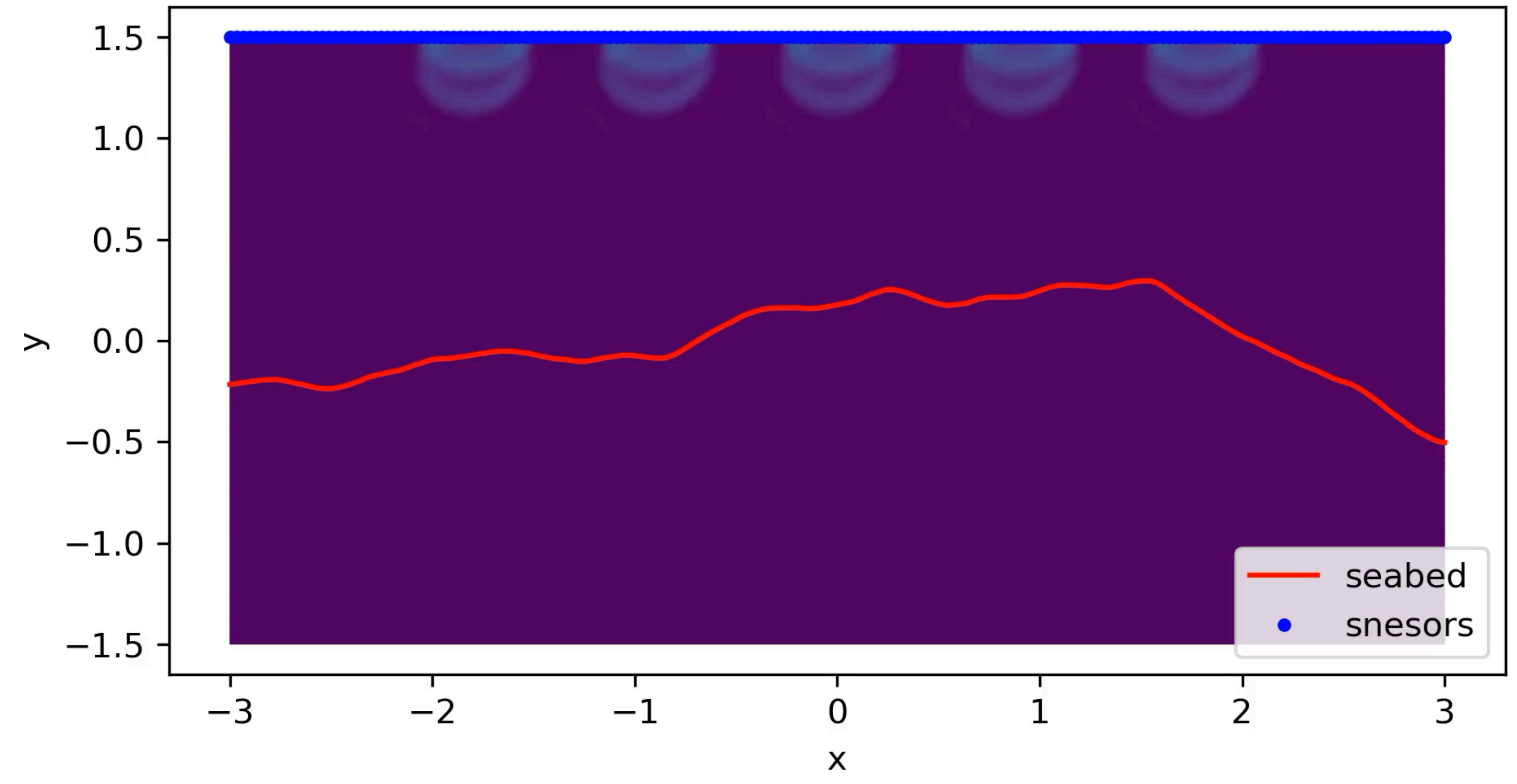


Examples of Inverse Problems

Ocean Floor Detection/Exploration

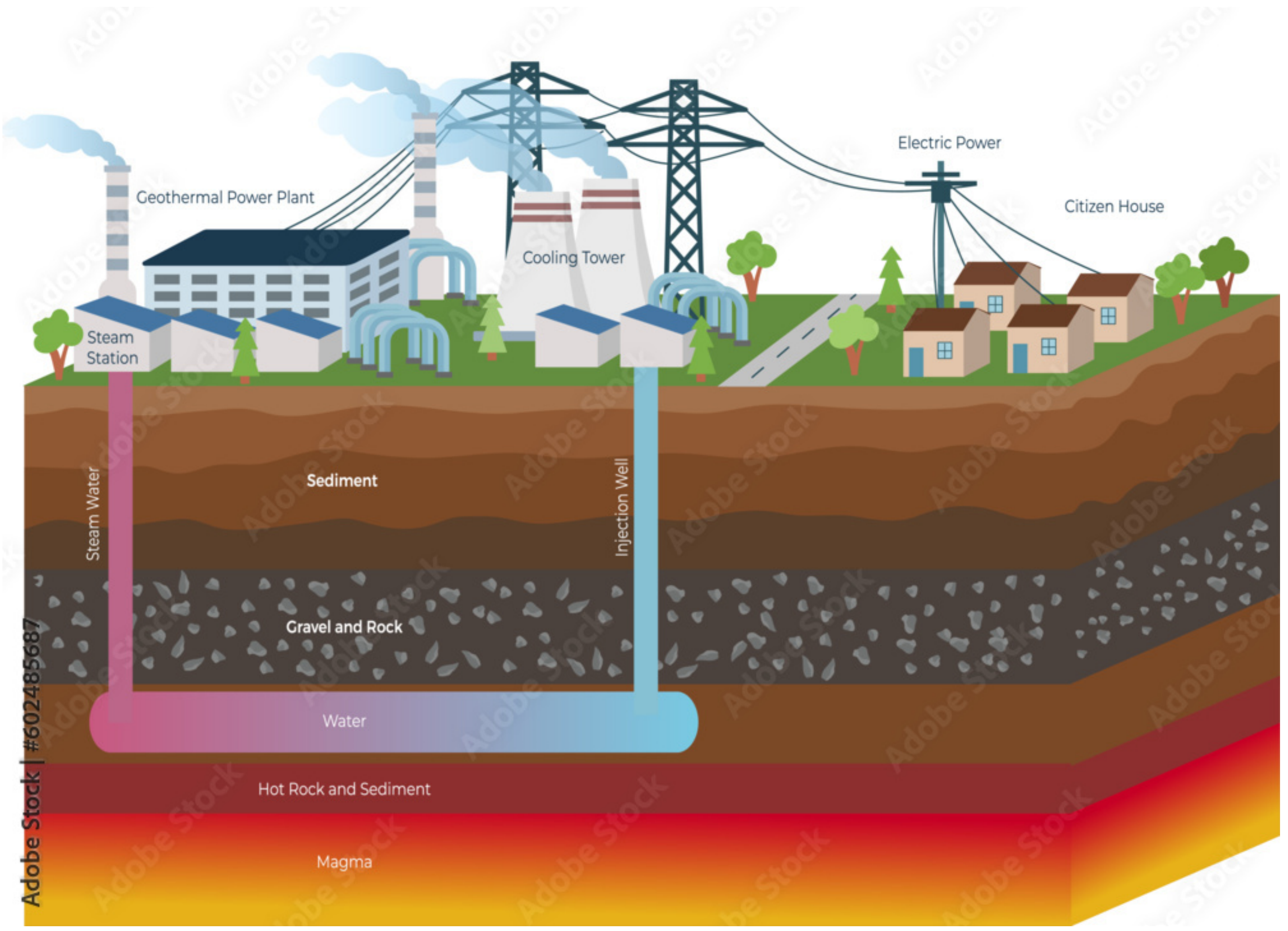






Examples of Inverse Problems

Geo-thermal Power stations (Motivation for Course Project)



Examples of Inverse Problems

- A typical formulation of an inverse problem

$$\mathbf{y} = F(\mathbf{x}) + \varepsilon$$

  - Here $\mathbf{x}$ is a mathematical representation of the unknown. It can be a parameter in $\mathbb{R}^+$, a vector of parameters in $\mathbb{R}^d$, a function in a function space, …
  - F is called the forward operator and is the (physical or approximate) process that creates noise-free measurements from a known $\mathbf{x}$.
  - ε is the noise in the measurement instruments.
  - $\mathbf{y}$ is the raw measurement.
- Discuss in groups what are $\mathbf{x}, \mathbf{y}, \varepsilon$ and F in the examples we discussed previously.

Uncertainty in Inverse Problems

Exercise 1

- In each of these examples investigate what are sources of uncertainty?
  - X-ray CT
  - Ocean floor detection with waves
  - Geo-thermal power station (both for exploration and monitoring)
- What are the consequences of uncertainties in each case?

Uncertainty in Inverse Problems
Exercise 2

- Choose an inverse problem of your choice. Investigate what are the sources of uncertainties.
- What are the consequences of uncertainty in your example?

Teaching Objectives of This Course

By the end of this course:

- You can formulate an inverse problem in a statistical setting.
- You can incorporate prior knowledge into the statistical problem in terms of a prior distribution.
- You can use the Bayes' theorem to formulate the solution to an inverse problem as the posterior distribution.
- You can write an algorithm and a Python code that can explore the posterior distribution with a sampling method.
- You can interpret samples of the posterior distribution as level of uncertainty (quantified uncertainty).
- You can perform uncertainty quantification for both linear and non-linear inverse problems.
- You will deliver outputs through teamwork.

What to expect in this course?

- You must work in groups of 3.
- You need to write codes in Python. You need numpy, matplotlib and scipy.
- We will have an active learning teaching method (appose to in-active students).
- Every session will have “homework” which we will do during the lectures.
- Ideally all homework will be finished in class.
- You will hand-in selected homework as your course report in groups of 3. You will be evaluated based on the final report.
- Advice: Start your Latex report as we go through the week.

Course Overview

- Day 1: Introduction to Inverse Problems, Uncertainty Quantification and revision on basic probability theory.
- Day 2: Markov chain and acceptance/rejection sampling.
- Day 3: The famous random-walk Metroplis-Hastings algorithm.
- Day 4: Choices of priors for Bayesian inverse problems.
- Day 5: Continuous and differentiable priors for Bayesian inverse problems.

Bayesian vs. Frequentist Debate

Source for the coin analogy:

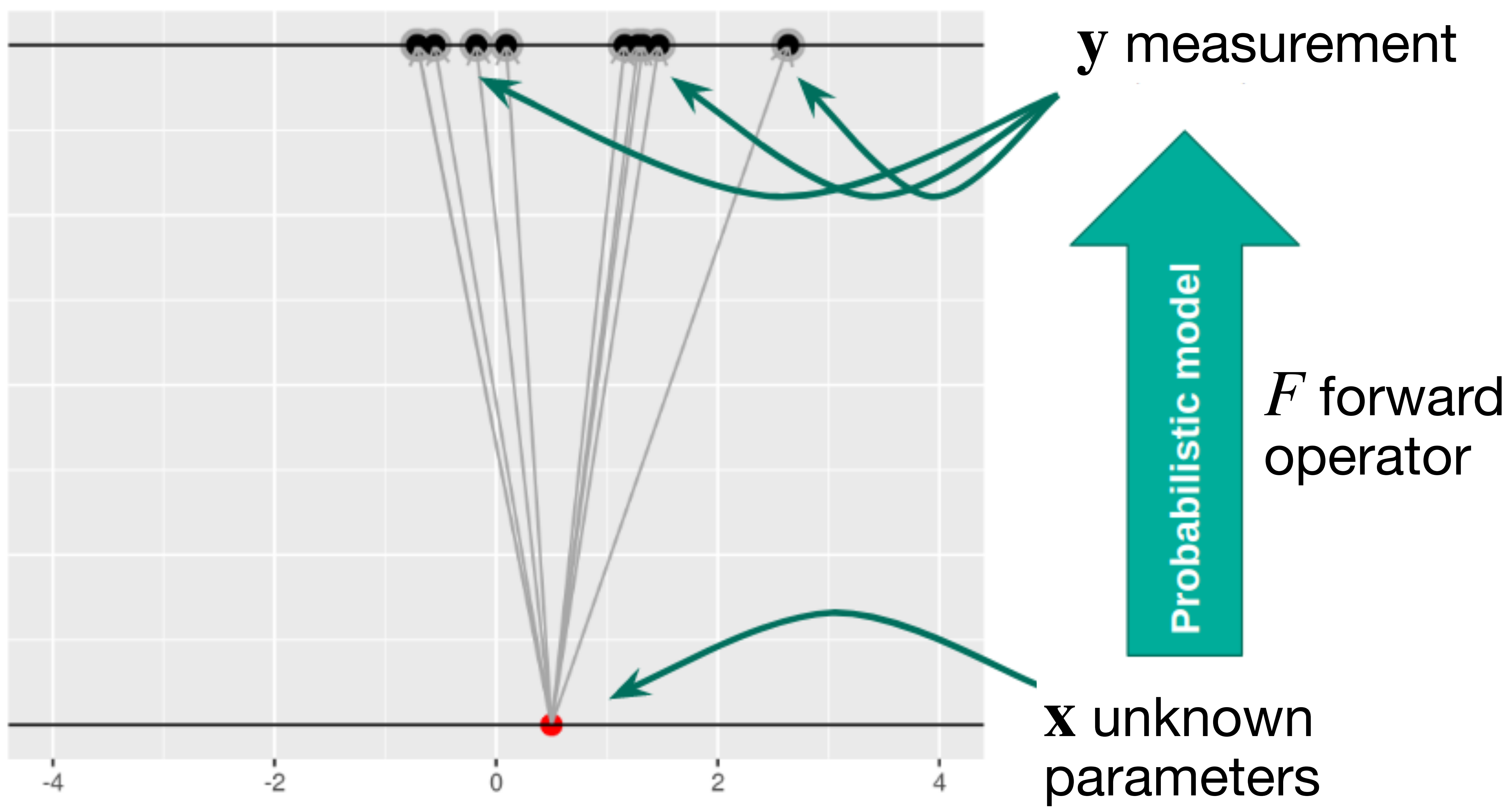
Cassie Kozyrkov
Are you Bayesian or Frequentist?

https://www.youtube.com/watch?v=GEFxFVESQXc&t=2s

Inverse Problems

A generic view

- Recall
  - $\mathbf{x}$ is the unknown
  - $\mathbf{y}$ is the measurement
  - F is the forward operator

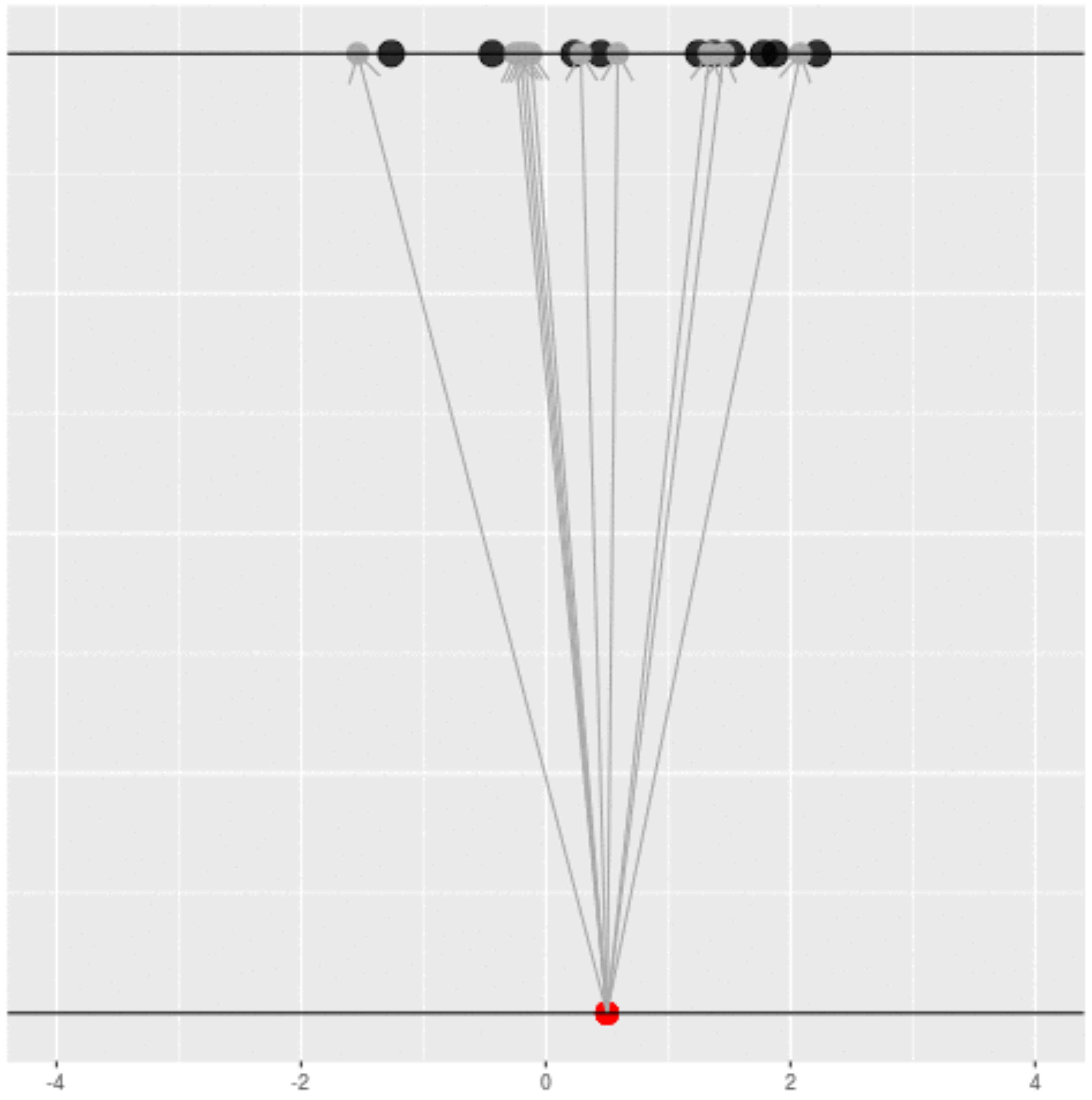


Inverse Problems

A frequentist approach

- There is a true parameter $\mathbf{x}$

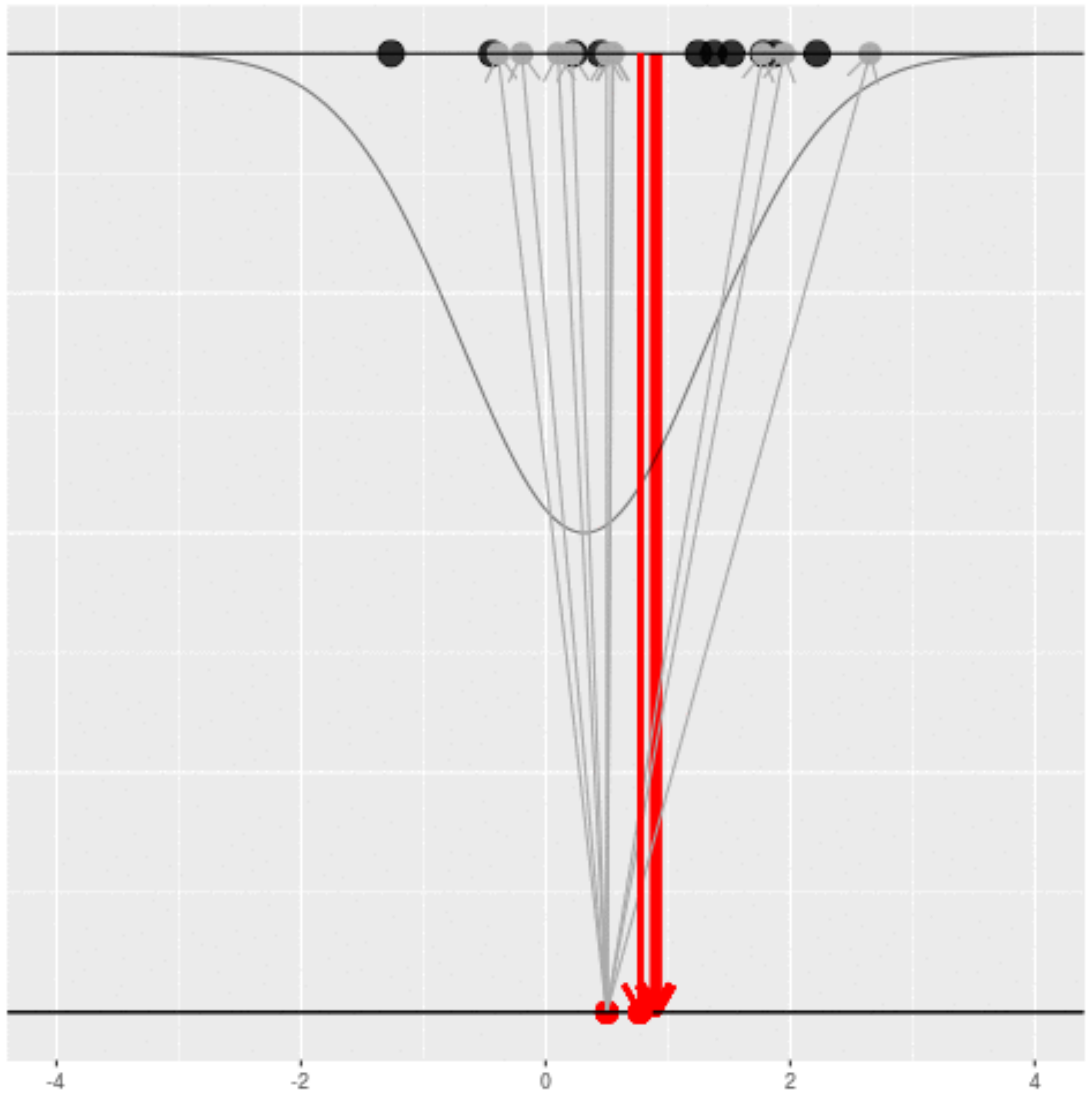
- Due to noise, every measurement is different, i.e., same parameter can result in different measurement data.



Inverse Problems

A frequentist uncertainty quantification

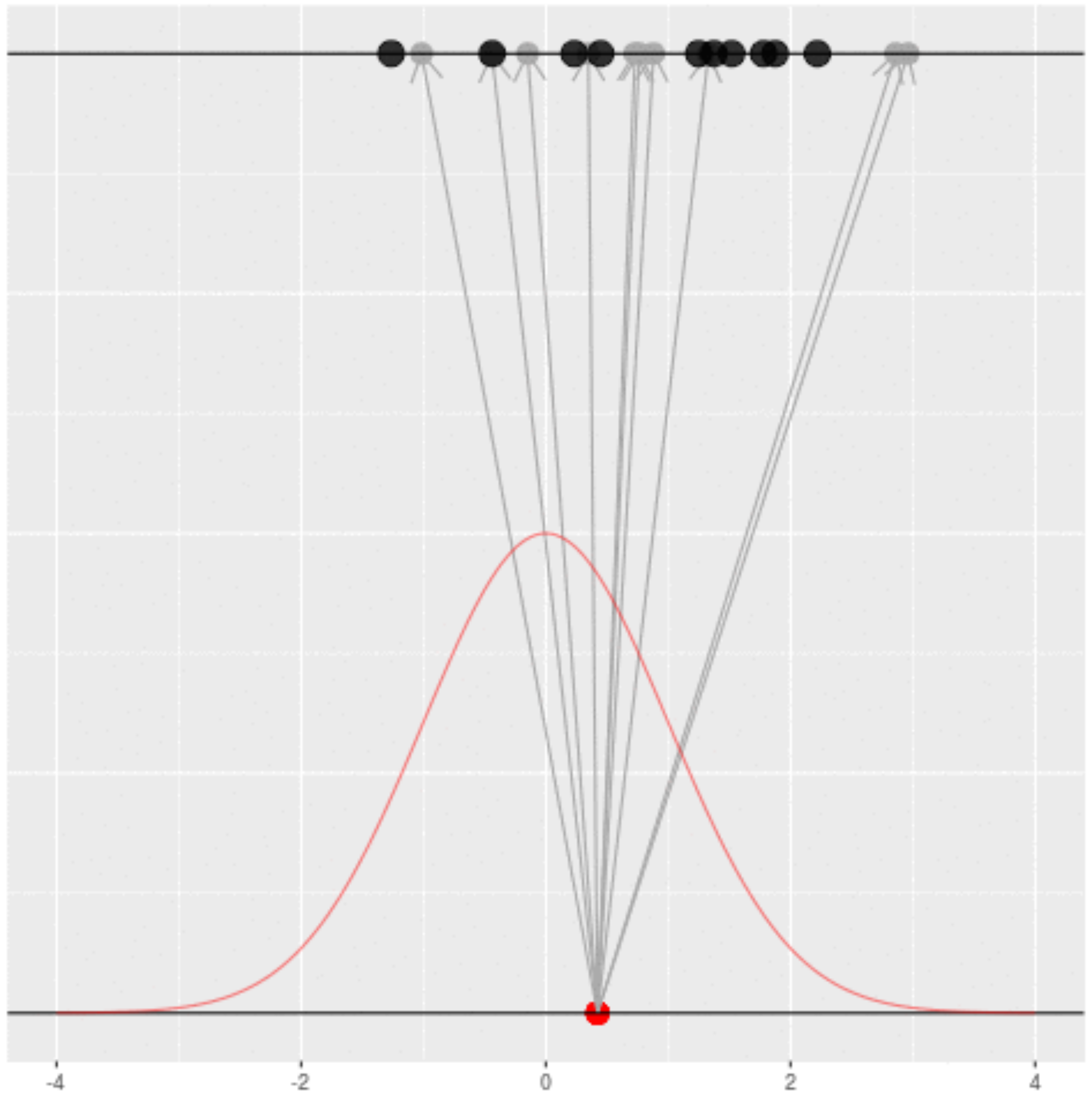
- Pick an estimate strategy.
- The range these estimates create, if we repeated the experiment, quantifies the uncertainty.



Inverse Problems

A Bayesian approach

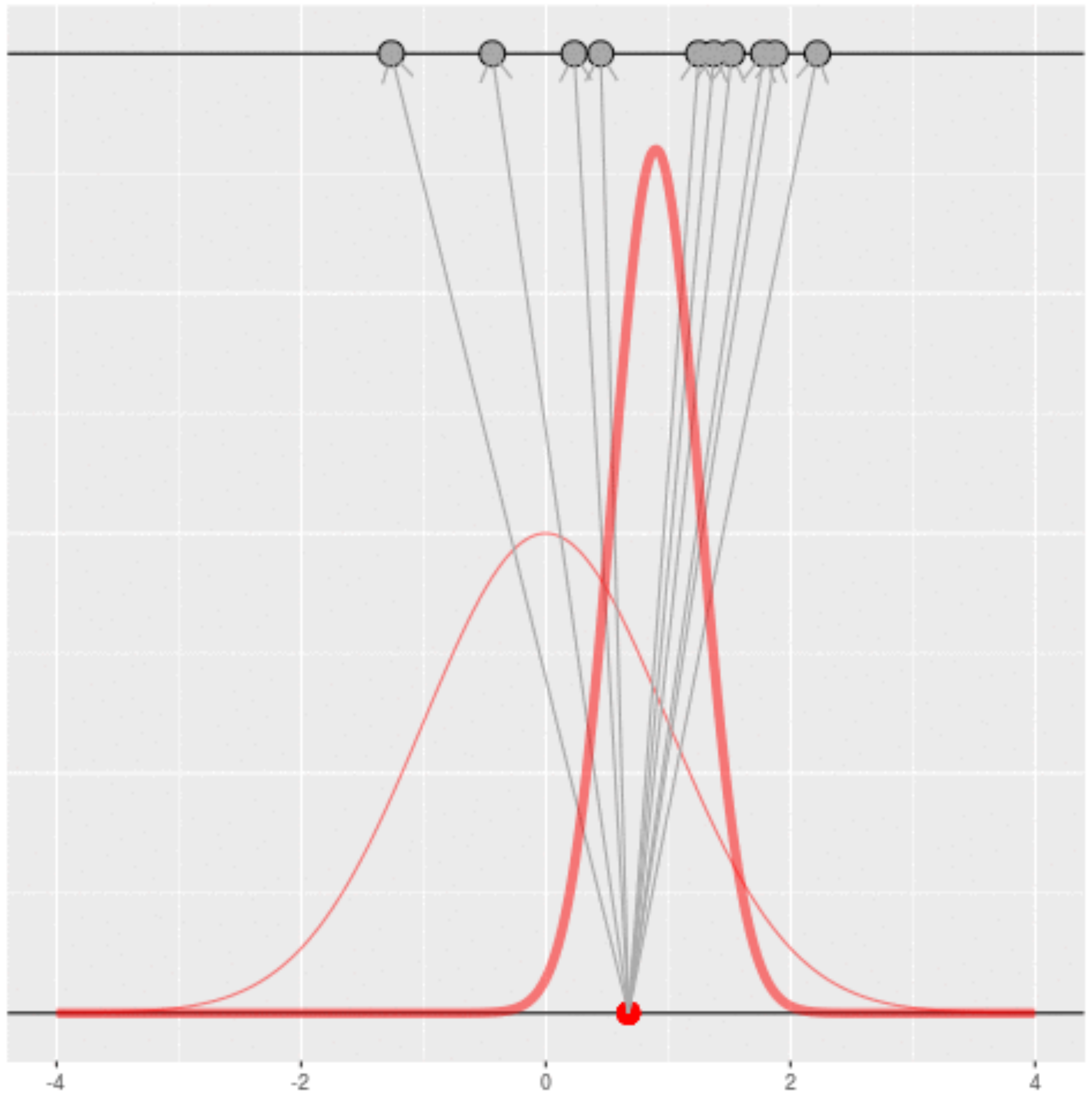
- There is a no true parameter $\mathbf{x}$
- We have an opinion, or a prior belief on what the value of $\mathbf{x}$ is.
- Different parameters can create different measurement data.



Inverse Problems

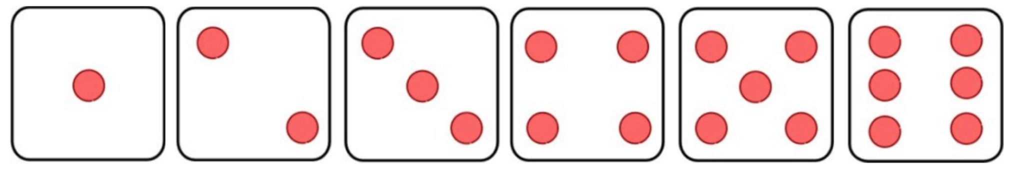
A Bayesian uncertainty quantification

- We delete the parameters that does not match with data.
- What remains is called the posterior.
- The uncertainty is then interpreted as the shape of the posterior.



Recap

Random variables

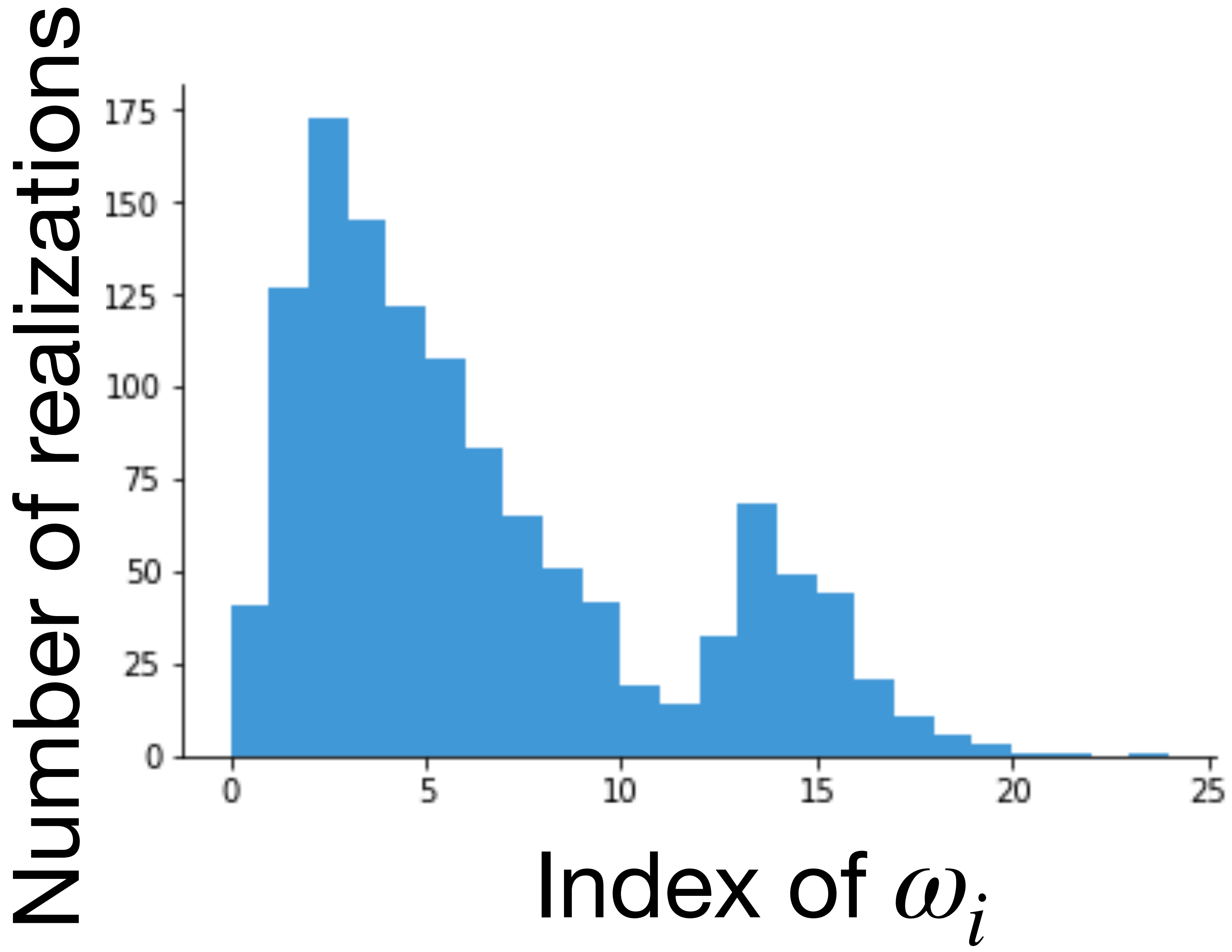
- We model random events that takes values in a set Ω as Ω-valued random variables and write them with capital letters, e.g., $X, Y, \ldots$
  - Let $A = \{$$\}$, then a dice roll is an A-valued random variable.
  - Let $B = \{$

$\}$, then a flip of a (Finnish) 1 Euro coin is a B-valued random variables.
  - Let $C = [0,3]$ meters, then measuring hight of people is a C-valued random variable.
  - Let D be the set of continuous and finite paths in 3D space. Then the path of a mosquito in the air is a D-valued random variable.

Recap

Histogram

- Let X be an Ω-valued random variables. Then under repeated realizations of X we record the outcomes in a sequence $\omega_1, \omega_2, \omega_1, \omega_4 \ldots$ A histogram is then a frequency plot.



Recap

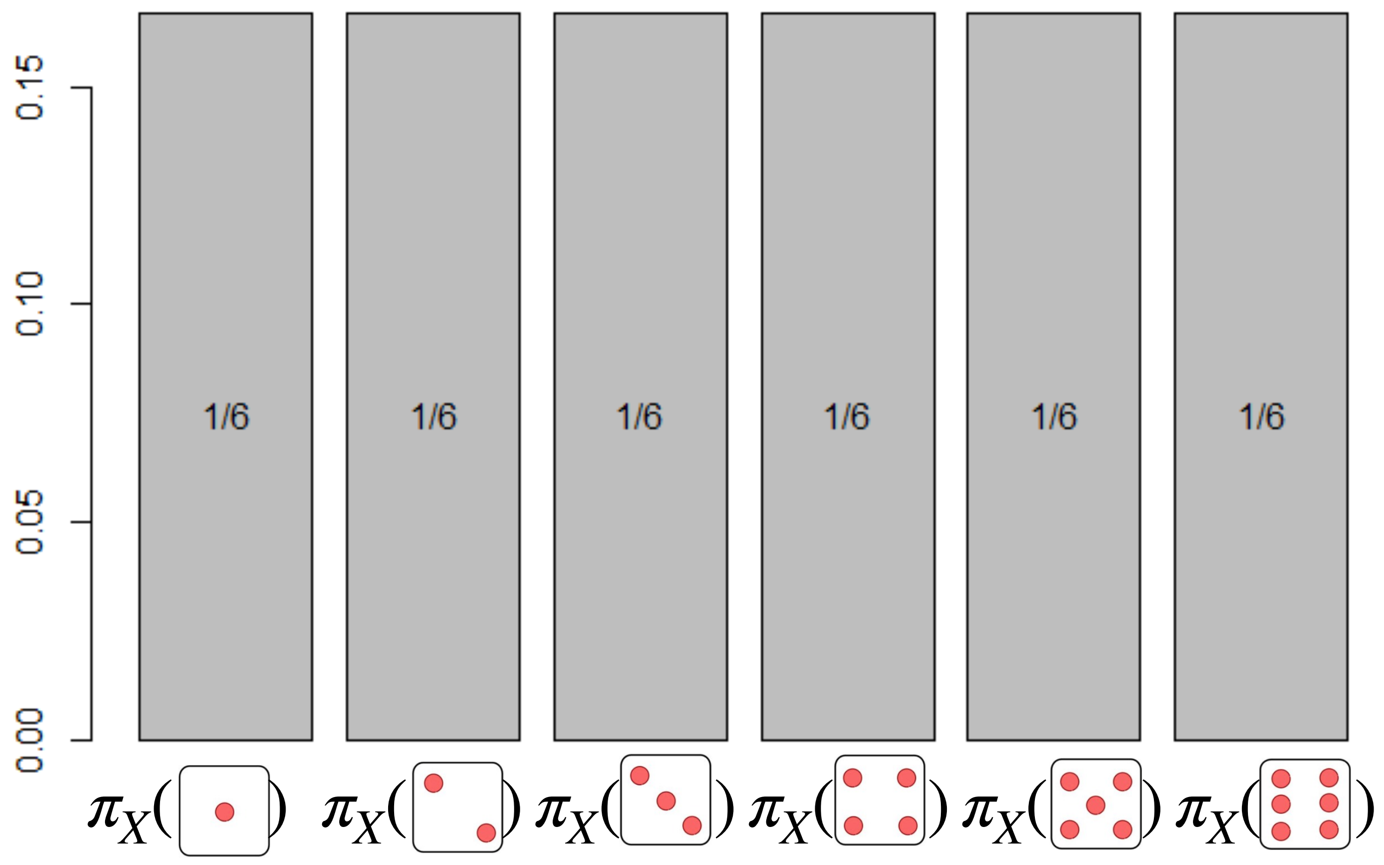
Density function

- The law, under which, a random variable behaves is called a distribution of a random variable.
    - For X being a fair dice roll, the distribution assigns equal probability of seeing each number.
- We can (sometimes) express a distribution using a density function. Which contains the “probability” of each event.
    - We represent a density as $\pi_X(\mathbf{x})$: This means the probability of the Ω-valued random variable X for the outcome $\mathbf{x} \in \Omega$.

Recap

Examples of Density function

- Let X be a random variable of a fair dice roll. Then we can represent its density function as



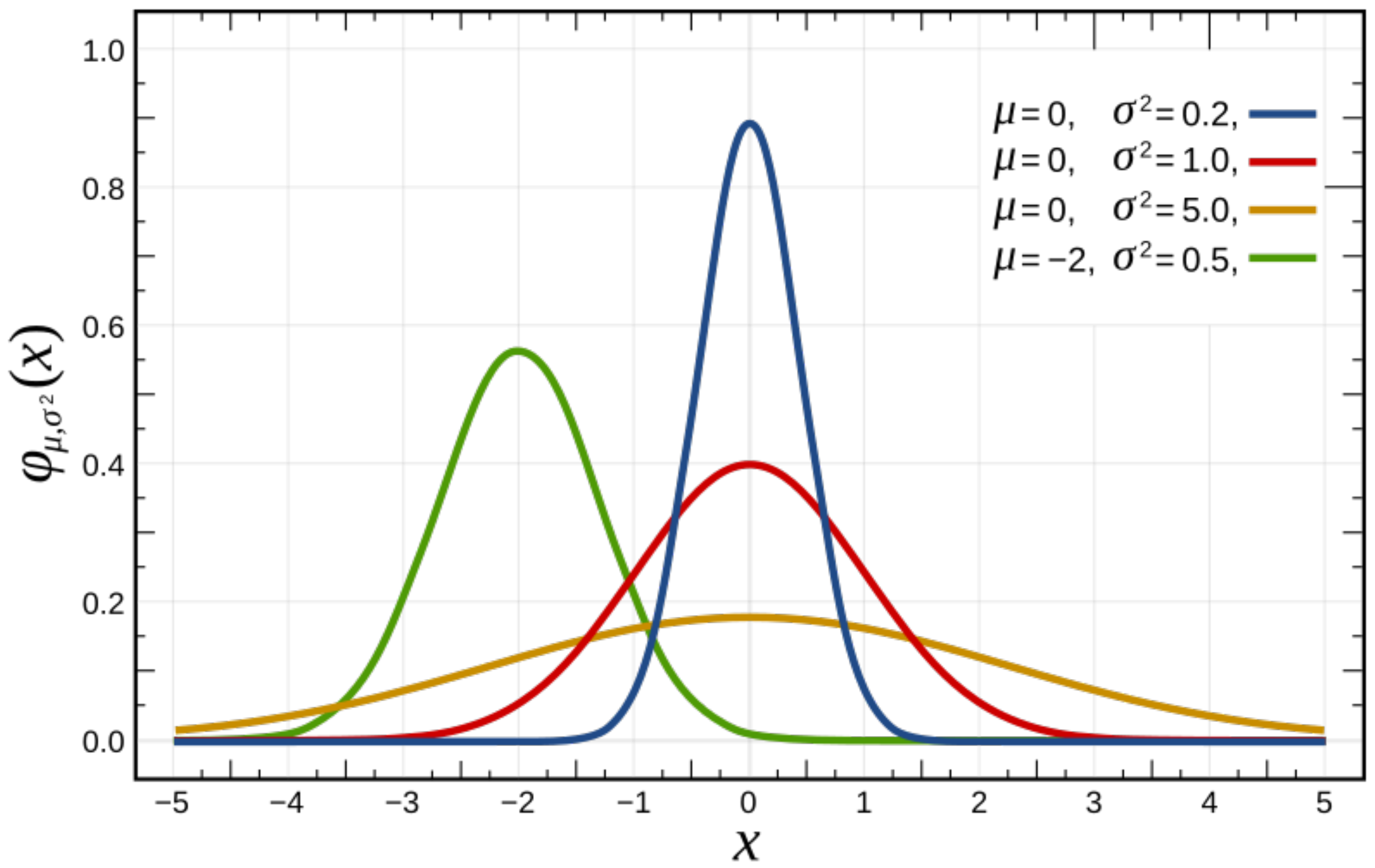
Recap

Examples of Density function

- Let X be an $\mathbb{R}$-valued random variable. Then the wrong way to think about its density function is to think of a function that assigns probability to each point $\mathbf{x} \in \mathbb{R}$. Although this analogy is wrong, we can use it for the purpose of this course.

- Normal distribution is a classic example. In this case:

$$\pi_X(\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2})$$



Recap

Examples of Density function - multivariate Gaussian

- Let X be an $\mathbb{R}^n$-valued Gaussian random variable. Then there is a vector $\mathbf{m} \in \mathbb{R}^n$ (the mean) and a symmetric and positive-definite matrix $\mathscr{C}$ (the covariance matrix) such that the density function of X is

$$\pi_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathscr{C}|}} \exp(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathscr{C}^{-1}(\mathbf{x} - \mathbf{m}))$$

  and we write $X \sim \mathscr{N}(\mathbf{m}, \mathscr{C})$.

- In this course we only consider zero-mean random variables.

Recap

Exercise

- What is the difference between a density function of random variable X and a histogram of a random variable X?

Statistical Formulation of Inverse Problems

- Recall formulation of an inverse problem.

$$\mathbf{y} = F(\mathbf{x}) + \varepsilon$$

- We define random variables to replace the components of the inverse problem.
    - define X to be the random variable of the unknown $\mathbf{x}$.
    - Define Y to be the random variable of the measurement $\mathbf{y}$.
    - Similarly E is the random variable of ε.
    - Note that randomness in F comes from $\mathbf{x}$, and F itself is not necessarily random.

Statistical Formulation of Inverse Problems

- The statistical modeling of the inverse problem is:

$$Y = F(X) + E$$

- What we need to define now is $\pi_X(\mathbf{x}), \pi_E(\mathbf{e})$:
  - $\pi_X(\mathbf{x})$ is called the prior density. It shows our belief (the probability) of any value $\mathbf{x}$ the.
  - $\pi_E(\mathbf{e})$ is the distribution of noise. Every sensor comes with a description of how precise measurements are.

- The solution to the inverse problem is then the conditional random variable $X \mid Y$, a.k.a. *the posterior*. We can also describe the posterior in terms of its density function.

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x})$$

Statistical Formulation of Inverse Problems

Bayes' rule

- We now use the Bayes' rule to simplify the posterior density function:

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x}) = \frac{\pi_{Y|X=\mathbf{x}}(\mathbf{y})\pi_X(\mathbf{x})}{\pi_Y(\mathbf{y})}$$

  - $\mathbf{y}$ is measurement data and we have it!
  - $\pi_{X|Y=\mathbf{y}}(\mathbf{x})$ is the posterior density function. This is what we want to compute.
  - $\pi_X(\mathbf{x})$ is the prior density function. This is something that we have.
  - $\pi_{Y|X}(\mathbf{y})$ is called the likelihood density function and is easy to evaluate (next slide).
  - $\pi_Y(\mathbf{y})$ is the probability of data. This is very hard to evaluate but generally unimportant. We find a way to deal with this term in the next days.

Statistical Formulation of Inverse Problems

Likelihood distribution $\pi_{Y|X}(\mathbf{y})$

- Recall a statistical inverse problem
$$Y = F(X) + E$$
- We want to compute $\pi_{Y|X=\mathbf{x}}(\mathbf{y})$. This means that we have $\mathbf{x}$:
$$Y = F(\mathbf{x}) + E$$
- Suppose that E is a Gaussian, i.e., $E \sim \mathcal{N}(0,\sigma^2)$, or $\pi_E(\mathbf{e}) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp(-\mathbf{e}^2/(2\sigma^2))$.
- $F(\mathbf{x})$ is not random! Take it to the other side of the equation and write:
$$Y - F(\mathbf{x}) = E$$
- Can you guess what is the density function of the left-hand-side? $F(\mathbf{x})$ can be a mean of a distribution.

Statistical Inversion

Linear problem with Gaussian noise

- Recall a statistical inverse problem
$$Y = F(X) + E,$$
Let
  - $X \sim \mathcal{N}(0,1)$, with
  $$\pi_X(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mathbf{x}^2}{2}\right)$$
  - $E \sim \mathcal{N}(0,\sigma^2)$, with
  $$\pi_E(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$
  - $\pi_{Y|X=\mathbf{x}} \sim \mathcal{N}(F(\mathbf{x}), \sigma^2)$, with
  $$\pi_{Y|X=\mathbf{x}}(\mathbf{y}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mathbf{y} - F(\mathbf{x}))^2}{2\sigma^2}\right)$$
  - Since $\mathbf{y}$ is already measured, then $\pi_Y(\mathbf{y})$ is a constant.

Statistical Inversion

Linear problem with Gaussian noise

- Now we can put everything into the Bayes' rule:

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x}) = \frac{\pi_{Y|X=\mathbf{x}}(\mathbf{y})\pi_X(\mathbf{x})}{\pi_Y(\mathbf{y})}$$

- This gives us the relation:

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x}) = \frac{1}{c}\exp\left(-\frac{(\mathbf{y}-F(\mathbf{x}))^2}{2\sigma^2}\right)\exp\left(-\frac{\mathbf{x}^2}{2}\right)$$

- The constant c is referred to as the normalization constant.

Statistical Inversion

Linear problem with Gaussian noise - multivariate case $\mathbf{x} \in \mathbb{R}^n$

- Now we can put everything into the Bayes' rule:

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x}) = \frac{\pi_{Y|X=\mathbf{x}}(\mathbf{y})\pi_X(\mathbf{x})}{\pi_Y(\mathbf{y})}$$

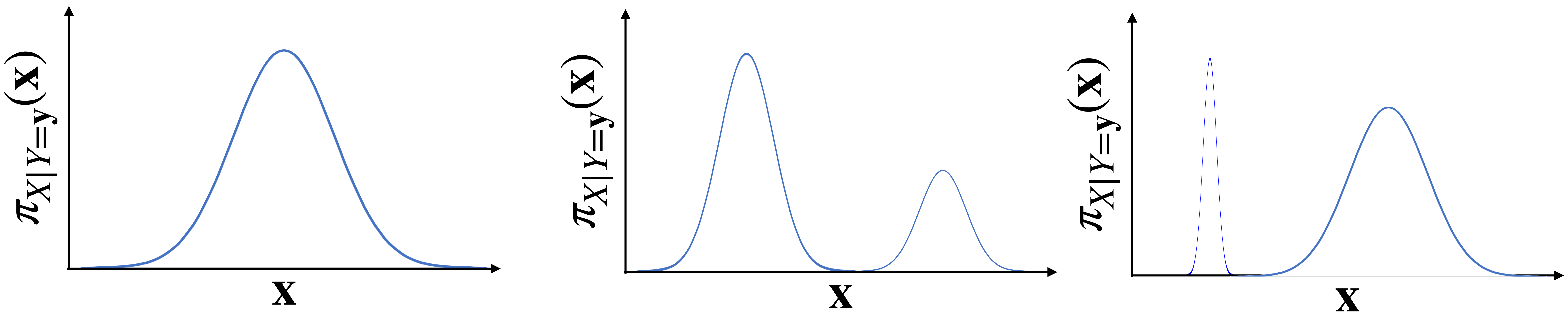
- This gives us the relation:

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x}) = \frac{1}{c}\exp\left(-\frac{\|\mathbf{y} - F(\mathbf{x})\|_2^2}{2\sigma^2}\right)\exp\left(-\frac{\mathbf{x}^T\mathscr{C}^{-1}\mathbf{x}}{2}\right)$$

- The constant c is referred to as the normalization constant.

Point Estimations of the Posterior

- Now that we have a distribution $\pi_{X|Y=\mathbf{y}}$, what should we report as the solution to the inverse problem.
- Some examples of possible posterior distributions:



Point Estimations of the Posterior

- The maximum a posterior (MAP) estimation

$$\mathbf{x}_{\text{MAP}} := \arg\max_{\mathbf{x}} \quad \pi_{X|Y=\mathbf{y}}(\mathbf{x})$$

- The posterior mean (or sometimes conditional mean)

$$\mathbf{x}_{\text{mean}} := \mathbb{E}(X \,|\, Y = \mathbf{y}) = \int_{\Omega} \mathbf{x}\pi_{X|Y=\mathbf{y}}(\mathbf{x})\; d\mathbf{x}$$

- However, there is no right or wrong point estimators!

Connection to the Tikhonov regularization

Exercise (HW1)

- For the inverse problem:

$$\mathbf{y} = F(\mathbf{x}) + \mathbf{e}$$

- A classic solution to inverse problems are Tikhonov regularized optimization:

$$\mathbf{x}_{\text{Tyk}} := \arg\min_{\mathbf{x}} \ \|F(\mathbf{x}) - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_2^2$$

- Show that for the posterior

$$\pi_{X|Y=\mathbf{y}}(\mathbf{x}) = \frac{1}{c} \exp\left(-\frac{\|\mathbf{y} - F(\mathbf{x})\|_2^2}{2\sigma^2}\right) \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2}\right),$$

  i.e., linear inverse problem with Gaussian noise and standard-normal prior ,i.e., $X \sim \mathcal{N}(0, I_n)$, we have

$$\mathbf{x}_{\text{Tik}} = \mathbf{x}_{\text{MAP}}$$

- What is the regularization parameter λ in this case?

Connection to the Tikhonov regularization

Exercise (HW1)

- Hints: Start with the MAP minimization problem.
- In the arg-min problem, you can take $\log$ of and/or multiply the argument with a positive constant without effecting its results.
- Same is true when dropping constants.
- When multiplying an arg-min problem with a negative value, the problem becomes arg-max, and vice-versa.