Probabilistic operator learning:
Generative modeling and uncertainty quantification for foundation models of differential equations

Benjamin Zhang
Division of Applied Mathematics
Brown University

Sampling, Inference, and Data-driven Physical Modeling in Scientific Machine Learning
Institute for Pure and Applied Mathematics
15 July 2025

Acknowledgments

Siting Liu

UC RIVERSIDE

Markos Katsoulakis

UMass Amherst

Stan Osher

UCLA

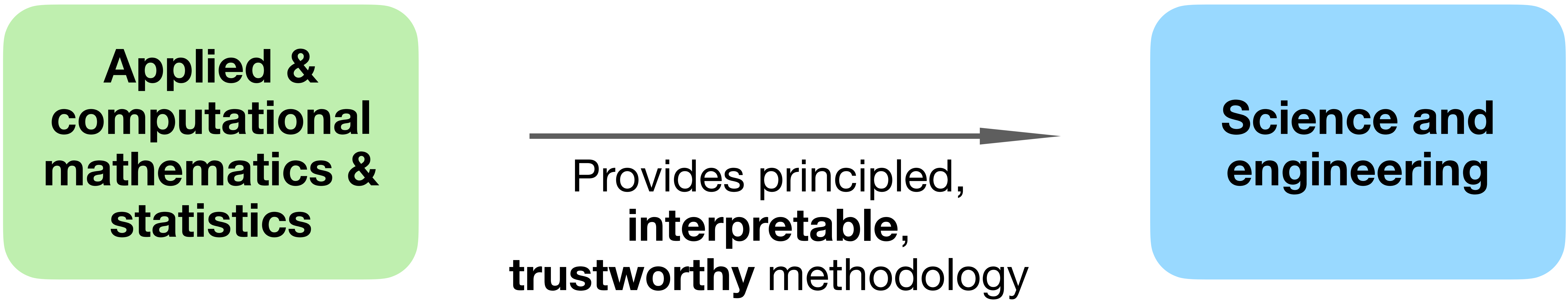
Funding agencies

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH · UNITED STATES AIR FORCE

NSF

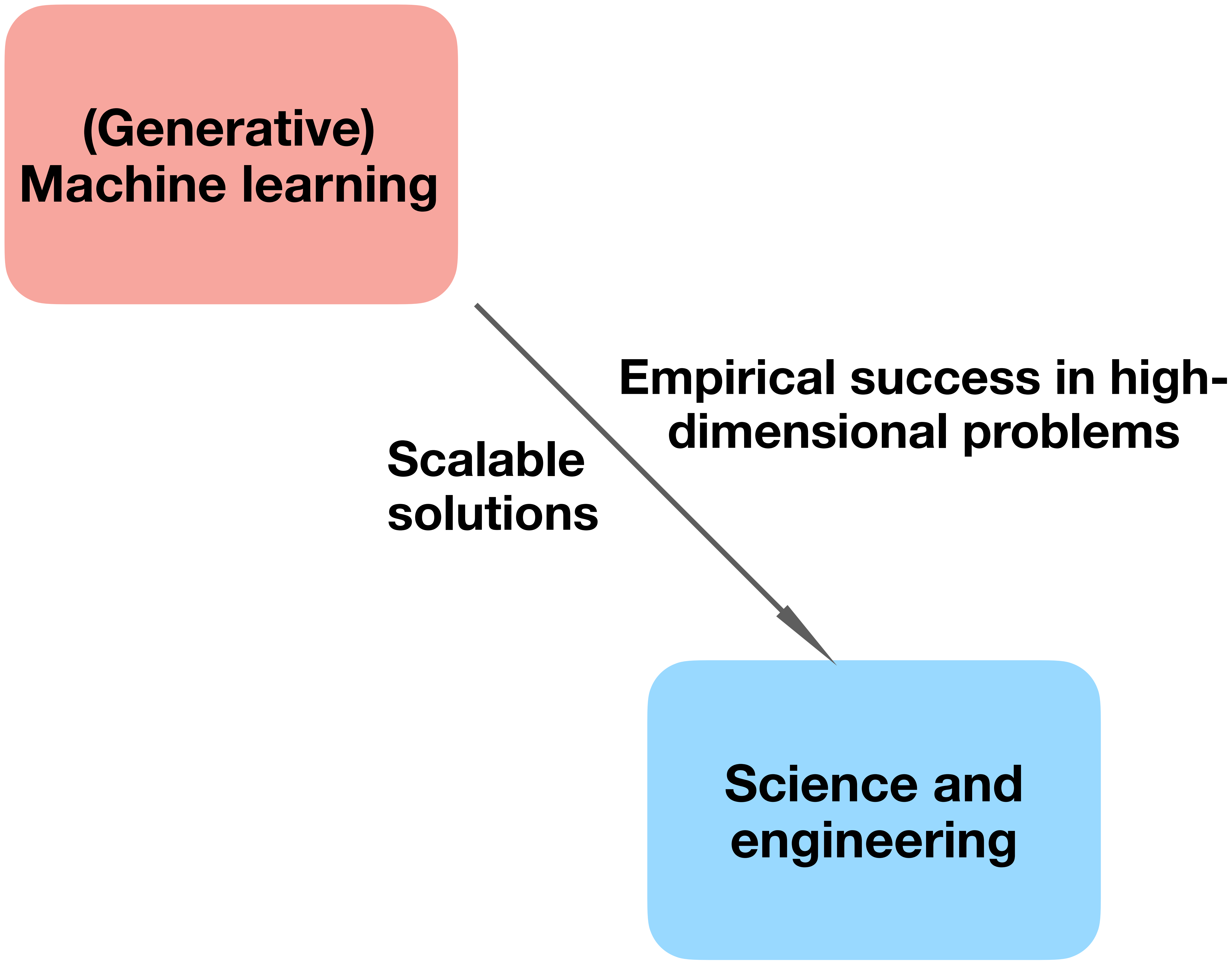
The establishment

Why do we trust computational methods?



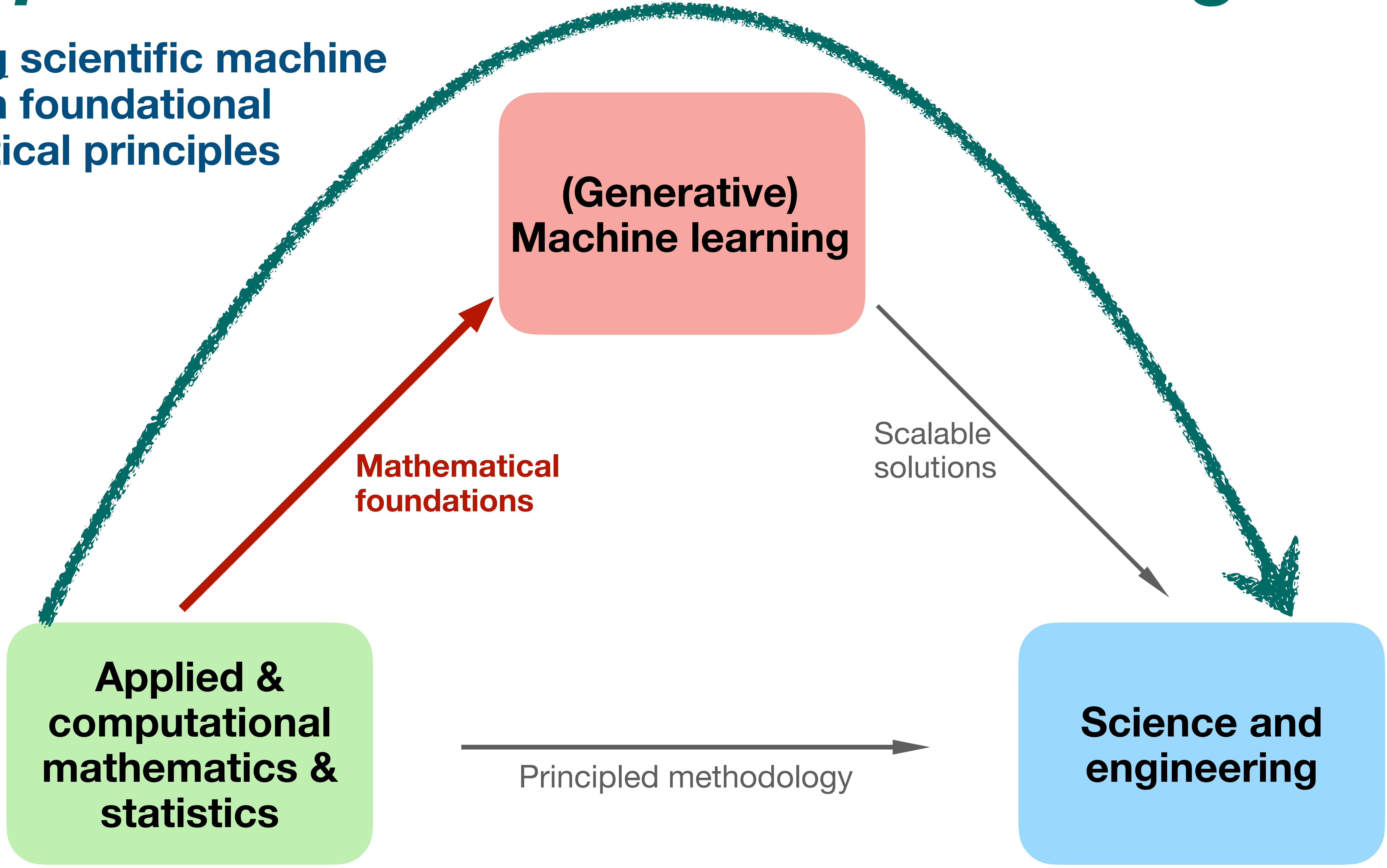
The insurgent

Why is machine learning attractive?



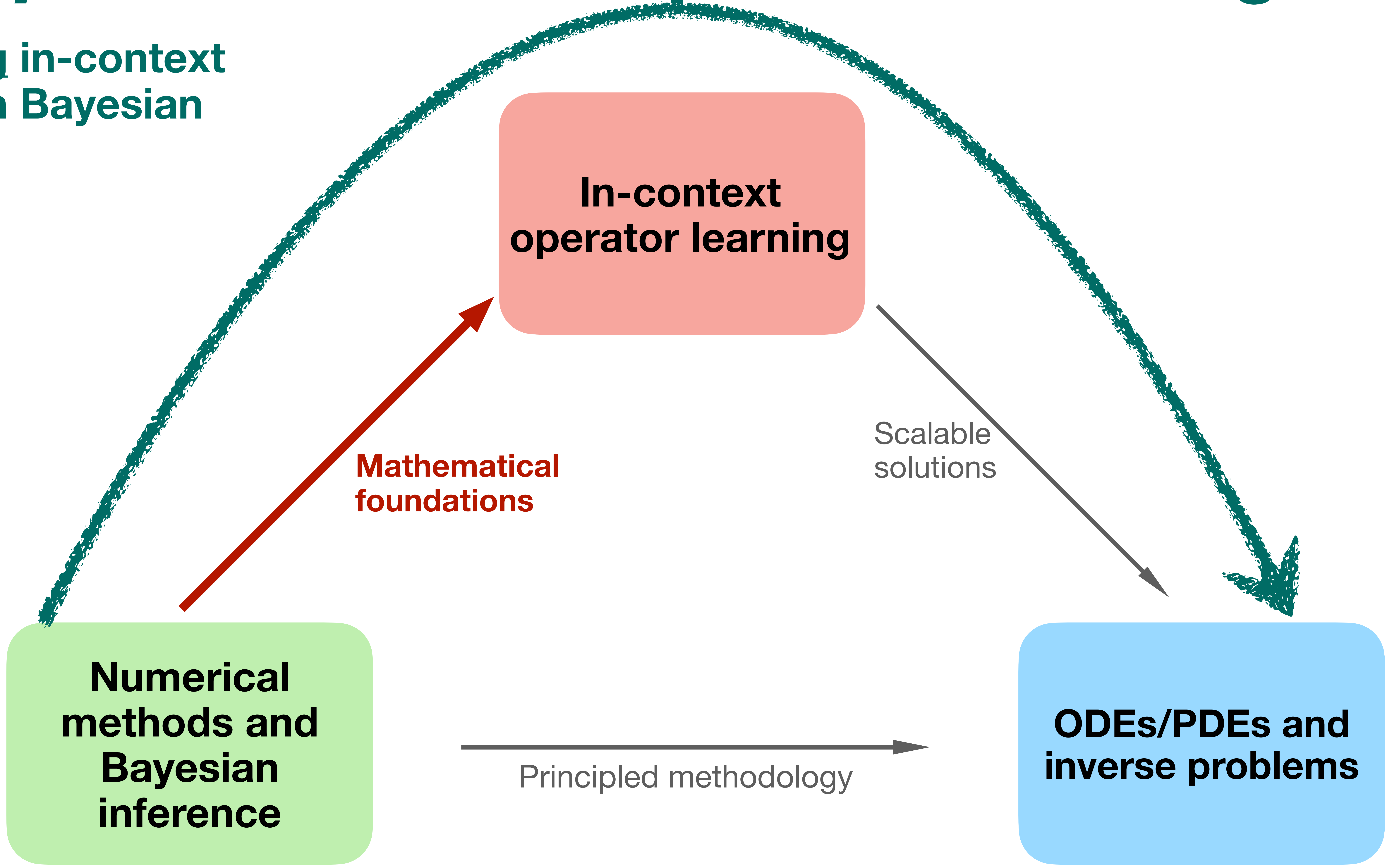
Principled scientific machine learning

<u>Grounding</u> scientific machine learning in foundational mathematical principles



Principled in-context operator learning

<u>Grounding</u> in-context learning in Bayesian inference



“Traditional” operator learning

Standard information flow for traditional operator learning

Operator Learning: algorithms and Analysis, Kovachki et al. 2024

Parameters: $\boldsymbol{\alpha}$ **Conditions: y** **Quantities-of-interest: z**

$\mathscr{A}(\alpha)z = 0,$ in D $\mathscr{B}(\alpha)z = y,$ on ∂D

Initial value problems:

$z'(t) = f(z(t); \alpha)$

$z(0) = y$

Boundary value problems:

$-\nabla \cdot (\alpha_1(x)\nabla z(x)) = \alpha_2(x)$ in D

$z(x) = y$ on ∂D

Inverse problems:

$-\nabla \cdot (\alpha_1(x)\nabla y(x)) = \alpha_2(x)$ in D

$y(x) = z$ on ∂D

Training data: $\{(\boldsymbol{\alpha}^{(i)}, y^{(i)}, z^{(i)})\}_{i=1}^N$

Model: $z = \mathscr{F}_\theta(y; \boldsymbol{\alpha})$

Training objective: $\min_\theta \sum_{i=1}^N \|z^{(i)} - \mathscr{F}_\theta(y^{(i)}; \boldsymbol{\alpha}^{(i)})\|_2^2$

$\mathscr{F}$ is a neural operator (Deep-O-net, FNO, PCA-net, etc.)

In-context operator learning

Information flow of ICON induces learning of ‘context’

Parameters: $\boldsymbol{\alpha}$ **Conditions:** y **Quantities-of-interest:** z

Group according to $\boldsymbol{\alpha}$: $\left\{ \left\{ (\boldsymbol{\alpha}^{(i)}, y^{(i,j)}, z^{(i,j)}) \right\}_{j=1}^{J} \right\}_{i=1}^{N}$

Training data omits $\boldsymbol{\alpha}$: $\left\{ \left\{ (y^{(i,j)}, z^{(i,j)}) \right\}_{j=1}^{J} \right\}_{i=1}^{N}$

$$\mathscr{A}(\alpha)z = 0, \text{ in } D$$

$$\mathscr{B}(\alpha)z = y, \text{ in } \partial D$$

Example Condition-QoI pairs

In-context operator network: $z = \mathscr{T}_\theta \left(y; \left\{ (y^{(j)}, z^{(j)}) \right\}_{j=1}^{J-1} \right)$

Query-prediction pair

Informs ‘context’

Training objective: $\min_\theta \sum_{i=1}^{N} \left\| z^{(J)} - \mathscr{T}_\theta \left(y^{(J)}; \left\{ (y^{(j)}, z^{(j)}) \right\}_{j=1}^{J-1} \right) \right\|_2^2$

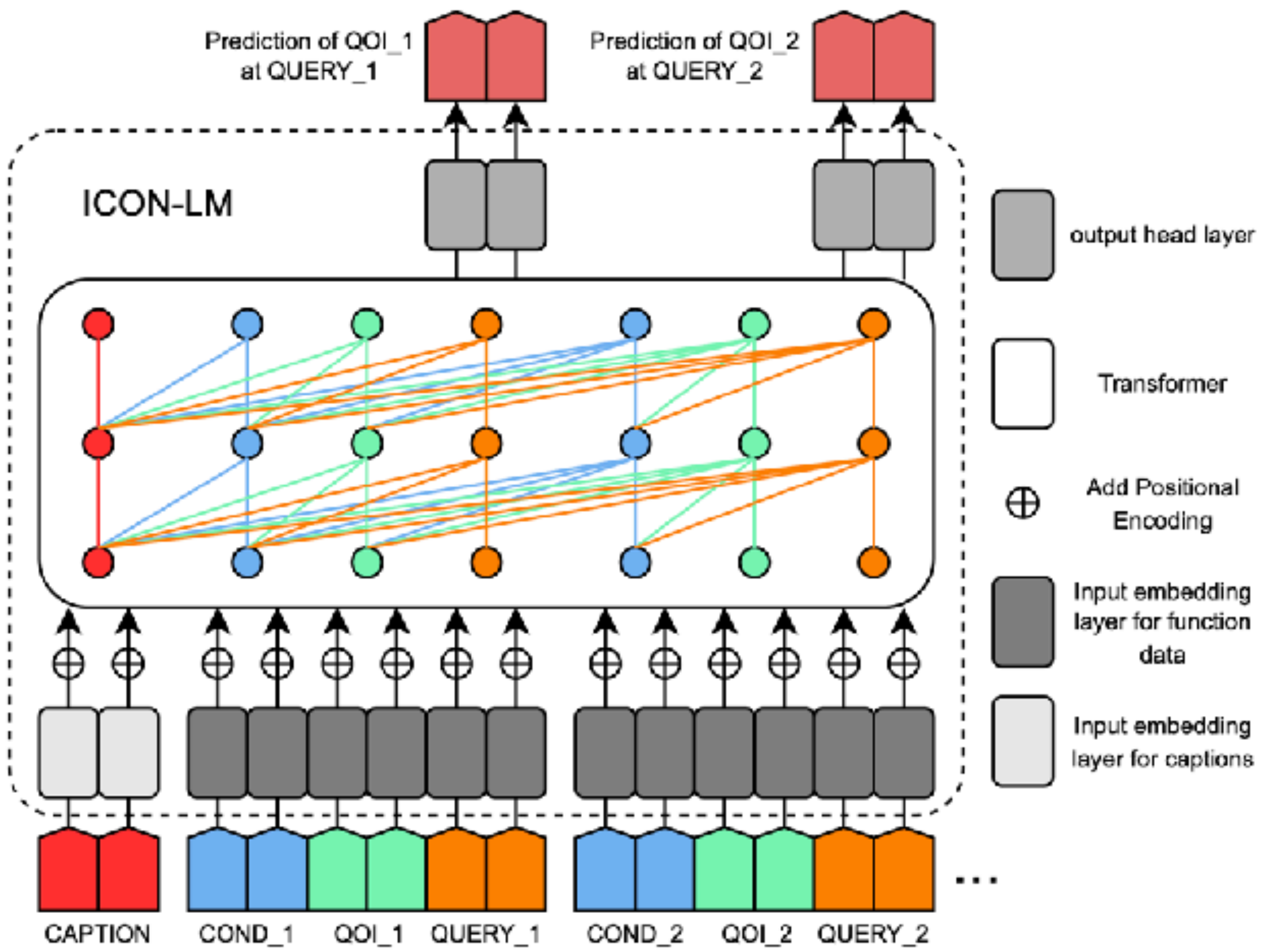
Note training does not involve $\boldsymbol{\alpha}$

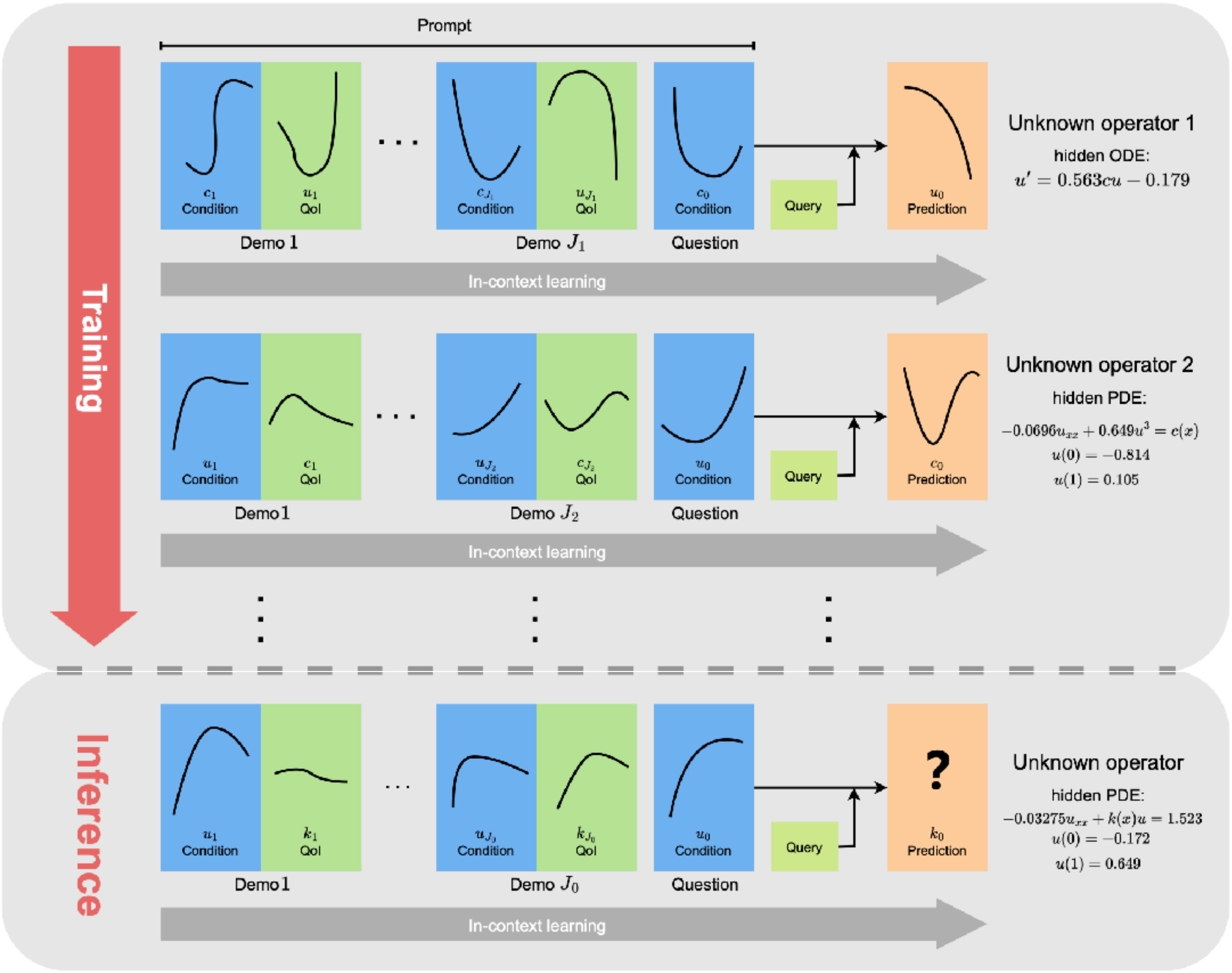
In-context operator networks

In-Context Operator Learning with Data Prompts for Differential Equation Problems, *PNAS,* L. Yang, S. Liu, T. Meng and S. Osher

Fine-Tune Language Models as Multi-Modal Differential Equation Solvers, Neural Networks, L. Yang, S. Liu, S. Osher

Differential equations	Parameters	Conditions	QoIs
$\frac{d}{dt}u(t) = a_1 c(t) + a_2$	a_1, a_2	$u(0), c(t), t \in [0,1]$	$u(t), t \in [0,1]$
for $t \in [0,1]$		$u(t), t \in [0,1]$	$c(t), t \in [0,1]$
$\frac{d}{dt}u(t) = a_1 c(t)u(t) + a_2$	a_1, a_2	$u(0), c(t), t \in [0,1]$	$u(t), t \in [0,1]$
for $t \in [0,1]$		$u(t), t \in [0,1]$	$c(t), t \in [0,1]$
$\frac{d}{dt}u(t) = a_1 u(t) + a_2 c(t) + a_3$	a_1, a_2, a_3	$u(0), c(t), t \in [0,1]$	$u(t), t \in [0,1]$
for $t \in [0,1]$		$u(t), t \in [0,1]$	$c(t), t \in [0,1]$
$u(t) = A\sin(\frac{2\pi}{T}t + \eta)e^{-kt}$	k	$u(t), t \in [0, 0.5)$	$u(t), t \in [0.5, 1]$
for $t \in [0,1]$		$u(t), t \in [0.5, 1]$	$u(t), t \in [0, 0.5)$
$\frac{d^2}{dx^2}u(x) = c(x)$	$u(0), u(1)$	$c(x), x \in [0,1]$	$u(x), x \in [0,1]$
for $x \in [0,1]$		$u(x), x \in [0,1]$	$c(x), x \in [0,1]$
$-\lambda a \frac{d^2}{dx^2}u(x) + k(x)u(x) = c$	$u(0), u(1), a, c$	$k(x), x \in [0,1]$	$u(x), x \in [0,1]$
for $x \in [0,1]$, $\lambda = 0.05$		$u(x), x \in [0,1]$	$k(x), x \in [0,1]$
$-\lambda a \frac{d^2}{dx^2}u(x) + ku(x)^3 = c(x)$	$u(0), u(1), k, a$	$c(x), x \in [0,1]$	$u(x), x \in [0,1]$
for $x \in [0,1]$, $\lambda = 0.1$		$u(x), x \in [0,1]$	$c(x), x \in [0,1]$
$\inf_{\rho,m} \iint c\frac{m^2}{2\rho}dxdt + \int g(x)\rho(1,x)dx$	$g(x), x \in [0,1]$	$\rho(t=0,x), x \in [0,1]$	$\rho(t=1,x), x \in [0,1]$
s.t. $\partial_t \rho(t,x) + \nabla_x \cdot m(t,x) = \mu \Delta_x \rho(t,x)$		$\rho(t=0,x), x \in [0,1]$	$\rho(t,x), t \in [0.5,1], x \in [0,1]$
for $t \in [0,1], x \in [0,1]$,		$\rho(t,x), t \in [0,0.5), x \in [0,1]$	$\rho(t,x), t \in [0.5,1], x \in [0,1]$
$c = 20, \mu = 0.02$,	$\rho(t=0,x)$	$g(x), x \in [0,1]$	$\rho(t=1,x), x \in [0,1]$
Periodic spatial boundary condition	$x \in [0,1]$		$\rho(t,x), t \in [0.5,1], x \in [0,1]$





What exactly is in-context operator learning?

What is the fundamental math problem ICON solves?

- What are the tools we use to approach this problem?
  - A probabilistic formulation for operator learning: **random differential equations**
- What exactly *is* in-context operator learning in the probabilistic formulation?
  - In-context operator learning implicitly performs **Bayesian inference**
- What does the Bayesian interpretation provide us?
  - A motivated **generative** formulation of ICON provides **uncertainty quantification**

A probabilistic formulation for operator learning

Random differential equations describes the data generating process

Key idea: parameters, conditions, and QoIs are random variables

Fix a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$

Model the family of ODEs/PDEs as RDEs.

Parameters: $\boldsymbol{\alpha}(\omega)$ **Conditions:** $y(\omega)$ **Quantities-of-interest:** $z(\omega)$

$$\begin{cases} \frac{d}{dt} z(t,\omega) = \alpha(t,\omega) z(t,\omega) \\ z(0,\omega) = z_0(\omega) = y(\omega) \end{cases}$$

RDEs induce probability measure which describes **data generating process** (data generating measure)

$$\mathcal{Q}(d\alpha, dy, dz)$$

In practice, requires discretization $0 = t_0 < t_1 < \cdots < t_N = T$

$$\{(\boldsymbol{\alpha}^{(i)}, y^{(i)}, z^{(i)})\}_{i=1}^{N} \sim \mathcal{Q}^N(d\alpha, dy, dz_1, \ldots, dz_N) = \mathcal{Q}_y(dy)\mathcal{Q}_\alpha(d\alpha)\prod_{n=0}^{N-1}\mathcal{K}(dz_{n+1} \,|\, z_n, \alpha)$$

$$\approx \mathcal{Q}(d\alpha, dy, dz)$$

Recall: Regression and conditional expectation

Least squares regression approximates conditional expectation

Let X and Y be random variables with joint distribution p_{XY}, and $h(X)$ be a predictor for Y. The predictor that minimizes the mean-squared error is the conditional expectation of Y given X.

$$\mathbb{E}[Y\,|\,X] = \arg\min_h \mathbb{E}_{p_{XY}} \|Y - h(X)\|_2^2$$

Joint: $p_{XY}(x, y)$

Marginal: $p_X(x) = \int p_{XY}(x, y)dy$

Conditional: $p_{Y|X}(y\,|\,x) = \dfrac{p_{XY}(x, y)}{p_X(x)}$

$$\mathbb{E}[Y\,|\,X = x] = \int y p_{Y|X}(y\,|\,x)dy$$

What does this mean for "traditional" operator learning?

Training data: $\{(\boldsymbol{\alpha}^{(i)}, y^{(i)}, z^{(i)})\}_{i=1}^N \sim \mathcal{Q}(\boldsymbol{\alpha}, y, z)$

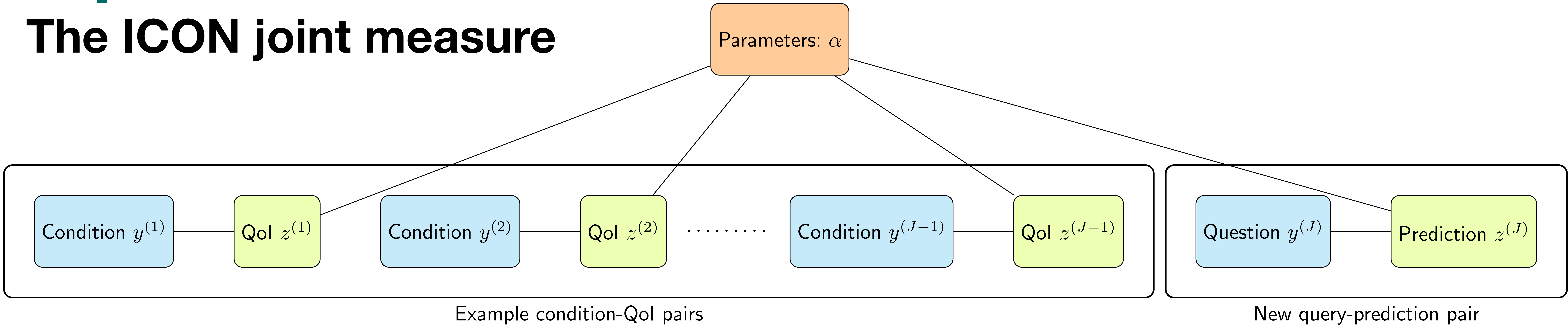
Objective: $\min_\theta \sum_{i=1}^N \|z^{(i)} - \mathcal{F}_\theta(y^{(i)}; \boldsymbol{\alpha}^{(i)})\|_2^2$

The optimal learned operator approximates conditional expectation

$$\mathbb{E}[z\,|\,y, \boldsymbol{\alpha}] \approx \mathcal{F}_{\theta^\star}(y^{(i)}; \boldsymbol{\alpha}^{(i)})$$

A probabilistic formulation for ICON

The ICON joint measure



Group according to $\boldsymbol{\alpha}$: $\left\{ \left\{ (\boldsymbol{\alpha}^{(i)}, y^{(i,j)}, z^{(i,j)}) \right\}_{j=1}^{J} \right\}_{i=1}^{N} \sim \mathcal{Q}(\boldsymbol{\alpha}, y, z) = \mathcal{Q}(z \,|\, y, \boldsymbol{\alpha}) \mathcal{Q}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \mathcal{Q}_y(y)$

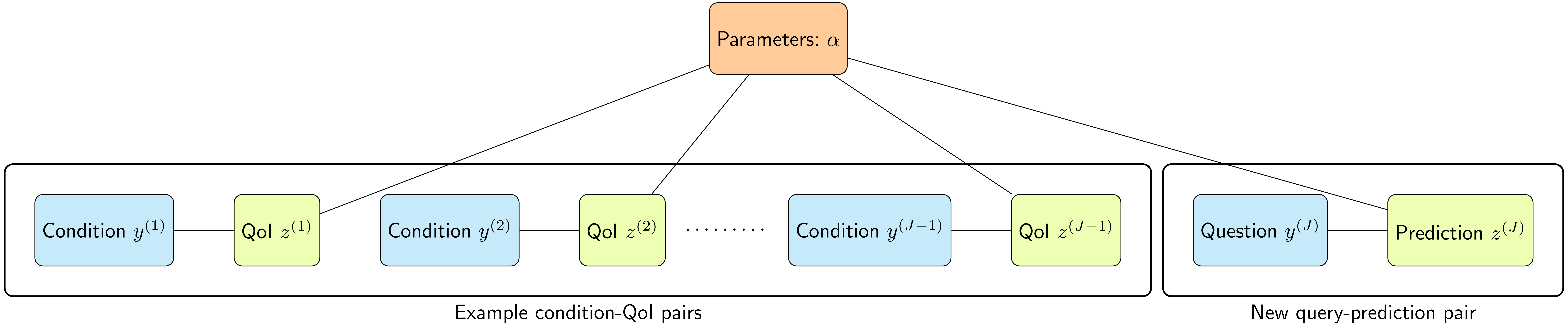
<u>Training data omits</u> $\boldsymbol{\alpha}$: $\left\{ \left\{ (y^{(i,j)}, z^{(i,j)}) \right\}_{j=1}^{J} \right\}_{i=1}^{N} \sim \mathcal{Q}\left(\{(y^{(j)}, z^{(j)})\}_{j=1}^{J} \right)$ **Subtly non-i.i.d!**

The joint measure of conditions-QoIs pairs

$$\mathcal{Q}\left(\{(y^{(j)}, z^{(j)})\}_{j=1}^{J} \right) = \int \mathcal{Q}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \prod_{j=1}^{J} \mathcal{Q}(z^{(j)} \,|\, y^{(j)}, \alpha) \mathcal{Q}_y(y^{(j)}) d\boldsymbol{\alpha}$$

What conditional expectation does ICON approximate?

ICON estimates expectation of the posterior predictive



Training data: $\left\{ \{(y^{(i,j)}, z^{(i,j)})\}_{i=1}^{N} \right\}_{j=1}^{J} \sim \mathcal{Q}\left(\{(y^{(j)}, z^{(j)})\}_{j=1}^{J} \right)$ Training objective: $\min_{\theta} \sum_{i=1}^{N} \left\| z^{(J)} - \mathcal{T}_{\theta}\left(y^{(J)}; \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1} \right) \right\|_2^2$

<u>The conditional measure</u> of prediction conditioned on query and example condition-QoI pairs

$$\mathcal{Q}\left(z^{(J)} \middle| y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1} \right)$$

Posterior predictive distribution

The in-context operator network approximates conditional expectation

$$\mathbb{E}\left[z^{(J)} \middle| y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1} \right] \approx \mathcal{T}_{\theta^\star}\left(y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1} \right)$$

Expectation of posterior predictive

What is the posterior predictive distribution?

What is the classical Bayesian approach to the ICON problem?

Model: $z'(t) = f(z(t); \alpha)$
$z(0) = y$

Task: Given $\{(y^{(j)}, z^{(j)})\}_{j=1}^{J}$ (from same α) and $y^{(J)}$, predict $z^{(J)}$

Step 1: Bayesian parameter estimation (compute the posterior distribution)

Bayes's Rule

$$\mathcal{Q}\left(\boldsymbol{\alpha} \,\middle|\, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}\right) \propto \mathcal{Q}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \prod_{i=1}^{J-1} \mathcal{Q}\left(y^{(j)}, z^{(j)} \,\middle|\, \boldsymbol{\alpha}\right)$$

Posterior — Prior — Likelihood

Step 2: Compute the posterior predictive (Predictions under an inferred α)

$$\mathcal{Q}\left(z^{(J)} \,\middle|\, y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}\right) = \int \mathcal{Q}\left(z^{(J)} \,\middle|\, y^{(J)}, \boldsymbol{\alpha}\right) \mathcal{Q}\left(\boldsymbol{\alpha} \,\middle|\, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}\right) d\boldsymbol{\alpha}$$

Posterior predictive — Prediction for a given $y^{(J)}$ and $\boldsymbol{\alpha}$ — Posterior on $\boldsymbol{\alpha}$

ICON implicitly performs Bayesian inference

Task: Given $\{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}$ and new question $y^{(J)}$, predict $z^{(J)}$

Classical Bayesian approach

Step 1: Compute posterior distribution

$$Q\left(\alpha \middle| \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) \propto Q_0(\alpha) \prod_{j=1}^{J-1} Q\left(y^{(j)}, z^{(j)} \middle| \alpha\right)$$

Step 2: Compute posterior predictive distribution

$$Q\left(z^{(J)} \middle| y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) = \int Q\left(z^{(J)} \middle| y^{(J)}, \alpha\right) Q\left(\alpha \middle| \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) \mathrm{d}\alpha$$

In-context operator networks

Least squares regression

$$\mathcal{T}^\star\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) = \arg\min_{\mathcal{T}} \mathbb{E}\left[\left\|z^{(J)} - \mathcal{T}\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right)\right\|_2^2\right]$$

Expectation taken over Joint measure of condition-QoI pairs

Conditional expectation predicts $z^{(J)}$

$$\mathbb{E}\left[z^{(J)} \middle| y^{(J)}, \{y^{(j)}, z^{(j)}\}_{j=1}^{J-1}\right] = \mathcal{T}^\star\left(y^{(J)}, \{y^{(j)}, z^{(j)}\}_{j=1}^{J-1}\right)$$

ICON directly approximates mean of posterior predictive distribution

An explanation of in-context learning as implicit Bayesian inference, Xie et. al 2021

Ex: Reaction-diffusion equation

$$\begin{cases} -0.05y_1(\omega)u''(x,\omega) + \alpha(x,\omega)u(x,\omega) = y_2(\omega) \\ u(0,\omega) = y_3(\omega), u(1,\omega) = y_4(\omega) \end{cases}$$

$$z(x,\omega) = u(x,\omega) + \epsilon$$

$\epsilon \sim \mathcal{N}(0,\sigma_o^2)$, for each x

Parameters $\alpha(x,\omega)$

Conditions $\{y_i(\omega)\}_{i=1}^4$

QoI $z(x,\omega)$

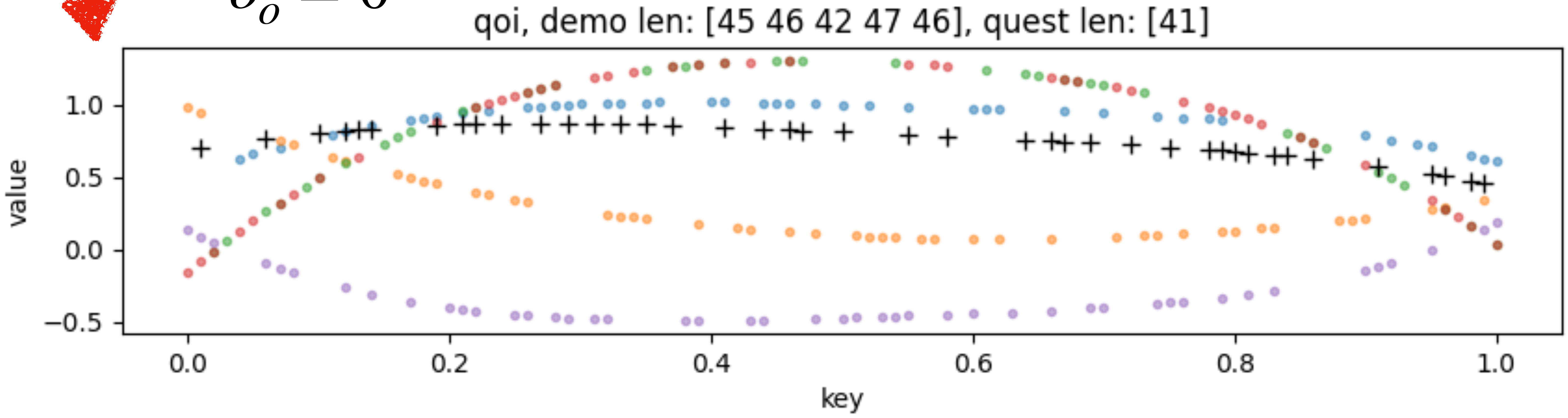
Noisy observations of the solution $u(x,\omega)$

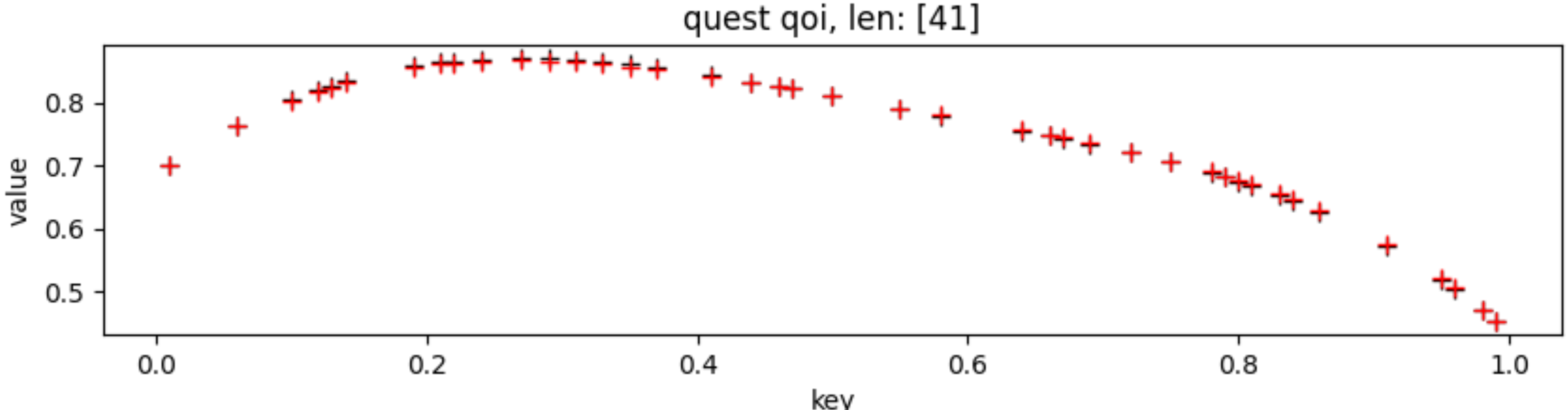
Data generating process

- $\alpha(x,\omega)$ rectified Gaussian process with squared exponential covariance
- $y_1 \sim \mathcal{U}[0.5,1.5], y_2 \sim \mathcal{U}[-2,2]$
  $y_3 \sim \mathcal{U}[-1,1], y_4 \sim \mathcal{U}[-1,1]$

QoI $z(x,\omega)$

$\sigma_o = 0$





- $J - 1 = 5$ Example condition-QoI pairs
- Colored lines: example solutions
- Truth (+) and prediction (+)
- **Learning from partial observations** (example pairs have different lengths)

Ex: Reaction-diffusion equation

$$\begin{cases} -0.05 y_1(\omega) u''(x,\omega) + \alpha(x,\omega) u(x,\omega) = y_2(\omega) \\ u(0,\omega) = y_3(\omega), u(1,\omega) = y_4(\omega) \end{cases}$$

$$z(x,\omega) = u(x,\omega) + \epsilon$$

$\epsilon \sim \mathcal{N}(0, \sigma_o^2)$, for each x

Parameters $\alpha(x,\omega)$

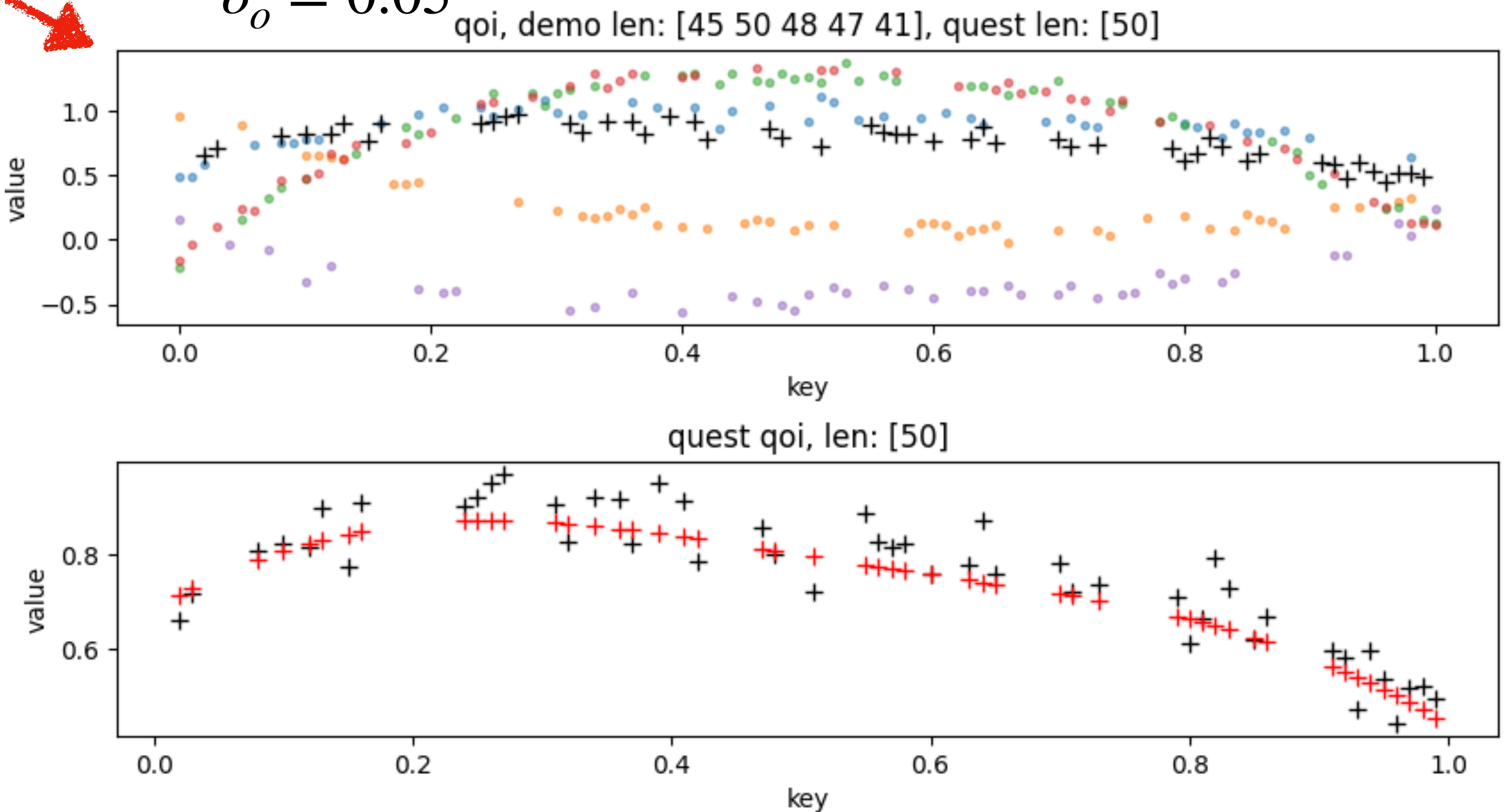
Conditions $\{y_i(\omega)\}_{i=1}^4$

QoI $z(x,\omega)$

QoI $z(x,\omega)$

Noisy observations of the solution $u(x,\omega)$

$\sigma_o = 0.05$



- $J - 1 = 5$ Example condition-QoI pairs
- Colored lines: example pairs
- Truth (+) and prediction (+)
- **ICON is de-noising**

$$\mathbb{E}\left[z^{(6)} \middle| y^{(6)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^5 \right]$$

Ex: Reaction-diffusion equation

$$\begin{cases} -0.05y_1(\omega)u''(x,\omega) + \alpha(\omega)u(x,\omega) = y_2(\omega) \\ u(0,\omega) = y_3(\omega), u(1,\omega) = y_4(\omega) \end{cases}$$

$$z(x,\omega) = u(x,\omega) + \epsilon$$

$\epsilon \sim \mathcal{N}(0,\sigma_o^2)$, for each x

Parameters $\alpha(\omega)$

Conditions $\{y_i(\omega)\}_{i=1}^4$

QoI $z(x,\omega)$

- **ICON learns from partial observations**
- **ICON is de-noising**

How? Implicitly performing Bayesian inference

Posterior $\mathcal{Q}\left(\boldsymbol{\alpha} \middle| \{(y^{(j)}, z^{(j)}\}_{j=1}^5\right)$

Posterior predictive $\mathcal{Q}\left(z^{(6)} \middle| y^{(6)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^5\right)$

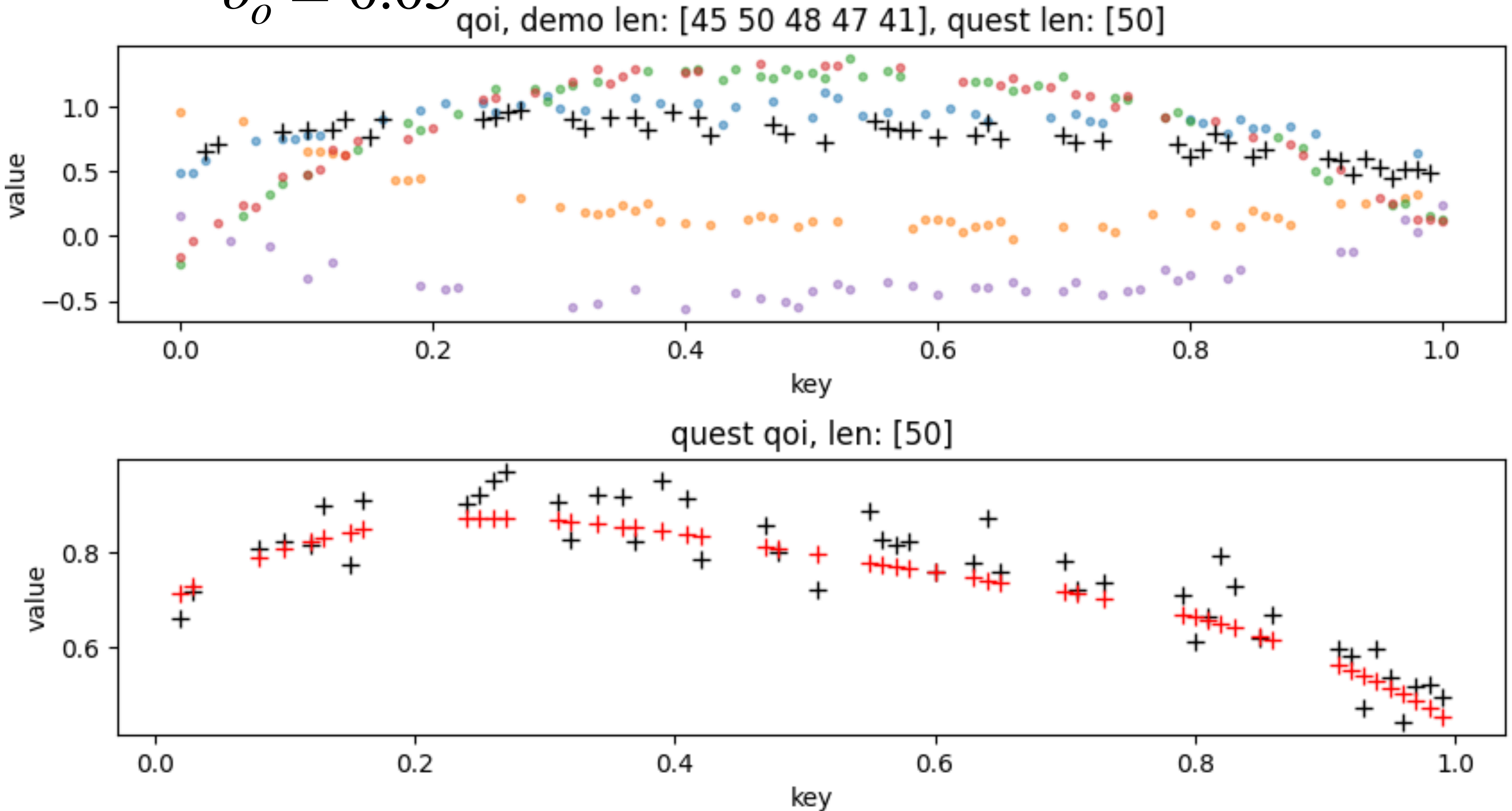
$$\mathbb{E}\left[z^{(6)} \middle| y^{(6)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^5\right]$$

$$= \mathbb{E}\left[u^{(6)} + \epsilon \middle| y^{(6)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^5\right]$$

$$= u^{(6)} \middle| y^{(6)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^5$$

Noisy observations of the solution $u(x,\omega)$

$\sigma_o = 0.05$



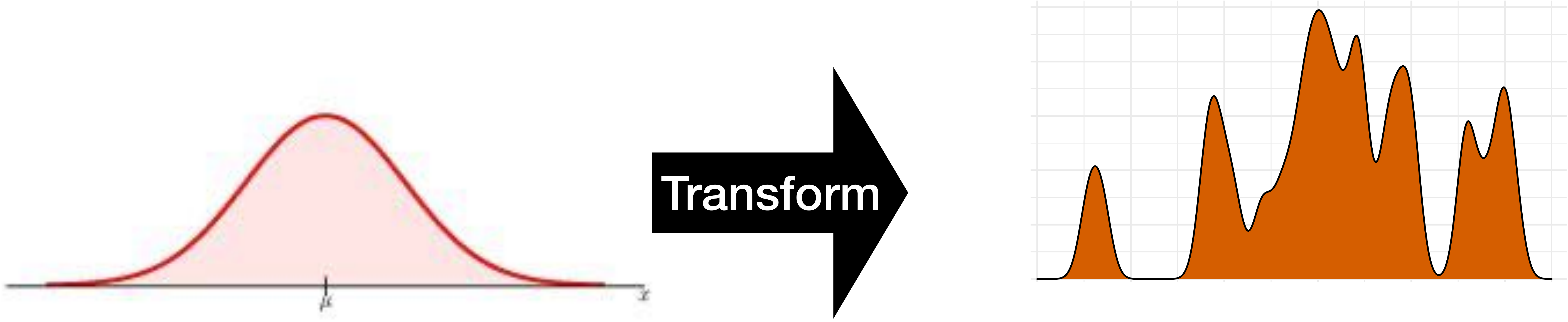
Can we sample from the posterior predictive?

A soupçon of generative modeling

Goal: Given data $\{Z_i\}_{i=1}^N \sim \mathbb{Q}$ from (unknown) **target distribution** $\mathbb{Q}$ **,** produce *new* samples from $\mathbb{Q}$

- Pick a **source distribution** p_X, easy to simulate (e.g. Gaussian).
- **Generative map:** Learn a **transport map** $\mathcal{G}$ such that:

$$\mathcal{G}_{\#}p_X \approx \mathbb{Q}, \textbf{ s.t. } X \sim p_X, \mathcal{G}(X) \sim \mathbb{Q}_\theta \approx \mathbb{Q}$$



source ρ (simple) **target** π (complex)

Can we sample from the posterior predictive?

A generative ICON produces samples from the posterior predictive

Generative in-context operator network samples from the posterior predictive

Let p_X be a reference distribution. **Generative ICON** is a mapping such that

$$\mathcal{G}^\star\left(\cdot\,, y^{(J)}, \left\{(y^{(j)}, z^{(j)})\right\}_{j=1}^{J-1}\right)_\sharp p_X(x) = \mathcal{Q}\left(\cdot\,\middle|\, y^{(J)}, \left\{(y^{(j)}, z^{(j)})\right\}_{j=1}^{J-1}\right)$$

$$x \sim p_X \implies \mathcal{G}^\star\left(x, y^{(J)}, \left\{(y^{(j)}, z^{(j)})\right\}_{j=1}^{J-1}\right) = z^{(J)} \sim \mathcal{Q}\left(\cdot\,\middle|\, y^{(J)}, \left\{(y^{(j)}, z^{(j)})\right\}_{j=1}^{J-1}\right)$$

Original ICON is the average Generative ICON!

$$\mathcal{T}^\star\left(y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}\right) = \mathbb{E}\left[z^{(J)}\,\middle|\, y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}\right]$$

$$= \mathbb{E}_X\left[\mathcal{G}^\star\left(x, y^{(J)}, \left\{(y^{(j)}, z^{(j)})\right\}_{j=1}^{J-1}\right)\right] = \int \mathcal{G}^\star\left(x, y^{(J)}, \left\{(y^{(j)}, z^{(j)})\right\}_{j=1}^{J-1}\right) p_X(x)dx$$

The posterior predictive provides uncertainty quantification

Why we should sample from the posterior predictive

Original ICON

$$\mathcal{T}^\star\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) = \mathbb{E}\left[z^{(J)} \middle| y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right]$$

- ICON **approximates mean** of posterior predictive distribution
- $\mathcal{T}^\star(\cdot,\cdot)$ is an operator learner

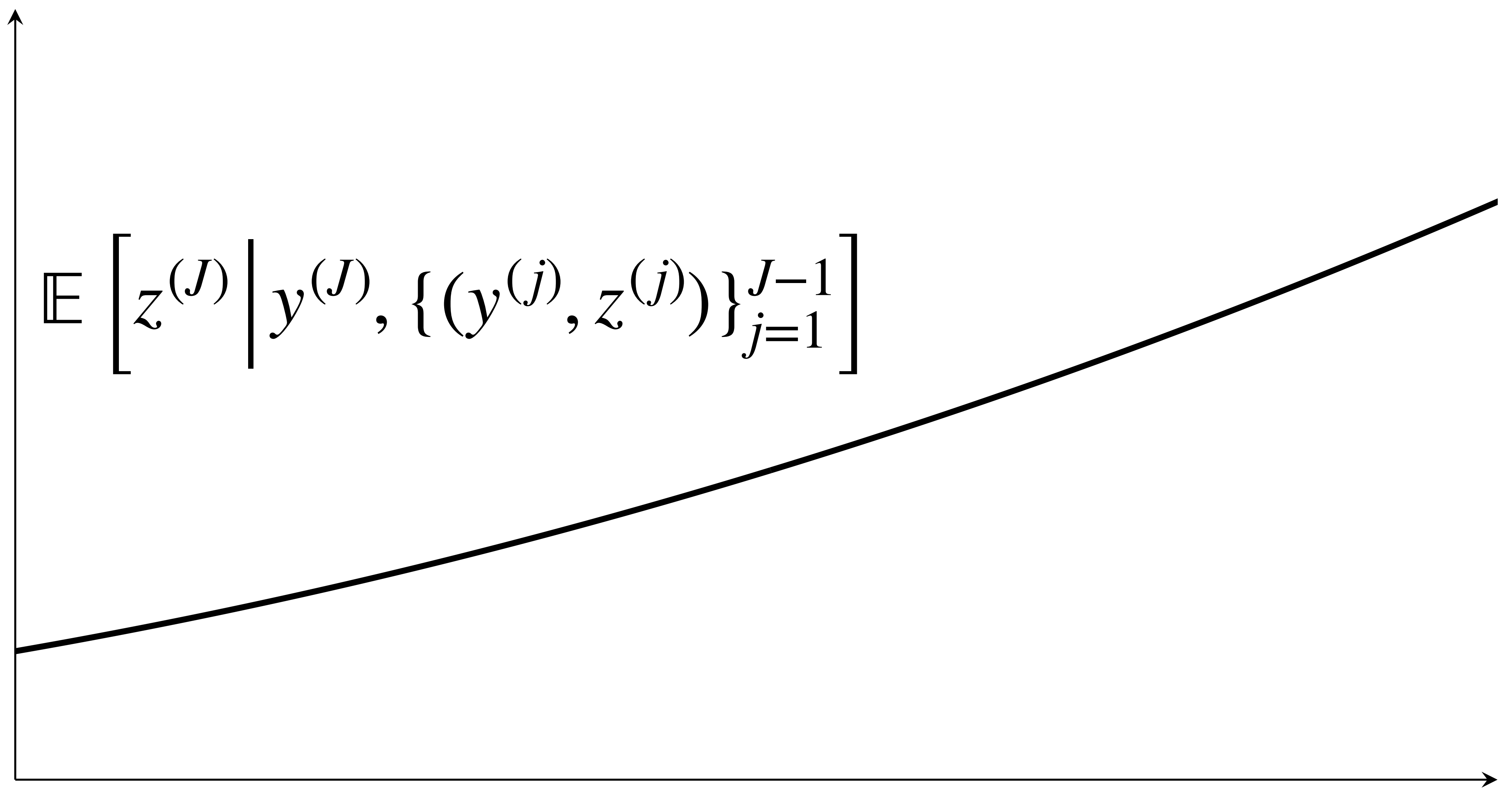
Theorem: Average generative ICON equals Original ICON

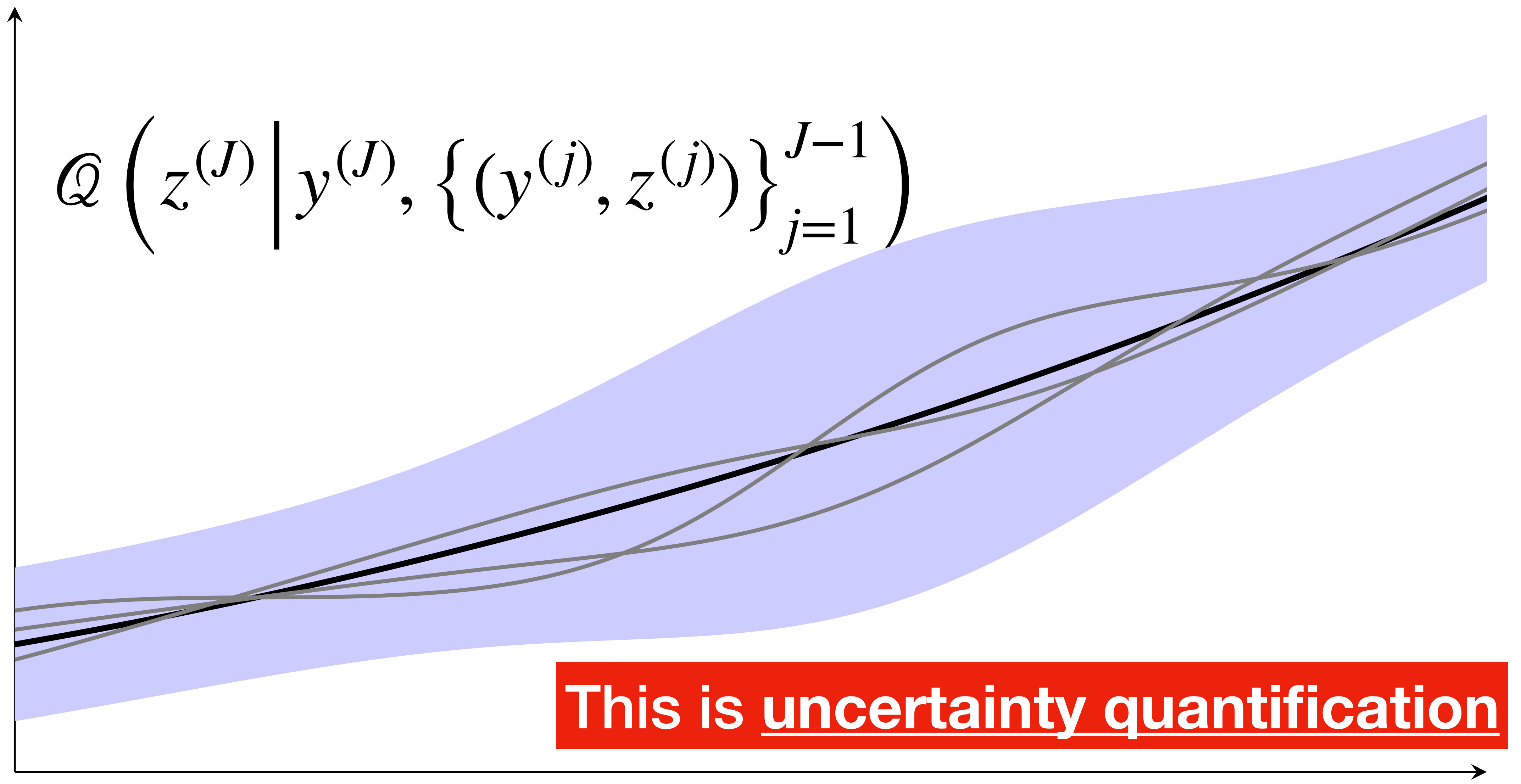
$$\mathcal{T}^\star\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) = \int \mathcal{G}^\star\left(x, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) p_X(x)\, dx$$

Generative ICON

$$\mathcal{G}^\star\left(\cdot, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right)_\sharp p_X = Q\left(\cdot \middle| y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right)$$

- Generative ICON **generates** from the posterior predictive distribution
- $\mathcal{G}^\star(x,\cdot,\cdot)$ is a *sample* operator learner





Generative in-context operator networks

Conditional sampling from the posterior predictive distribution

Example: training via conditional Generative Adversarial Networks

$$\min_{\theta} \mathscr{D}\left(\mathcal{Q}\left(z^{(J)} \,|\, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right), \mathcal{Q}_\theta\left(z^{(J)} \,|\, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right)\right) = \min_{\theta} \mathbb{E}_{\mathcal{Q}_\theta}\left[f\left(\frac{d\mathcal{Q}}{d\mathcal{Q}_\theta}\right)\right]$$

$$= \min_{\theta} \max_{\phi} \mathbb{E}_{Q}[D_\phi(y^{(J)})] - \mathbb{E}_{Q_\theta}[(f^\star \circ D_\phi)(y^{(J)})]$$

Training data: $\left\{ \left\{(y^{(i,j)}, z^{(i,j)})\right\}_{i=1}^{N} \right\}_{j=1}^{J} \sim \mathcal{Q}\left(\{(y^{(j)}, z^{(j)})\}_{j=1}^{J}\right)$

$$\min_{\theta} \max_{\phi} \left\{ \frac{1}{M} \sum_{m=1}^{M} D_\phi\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}, z^{(J)}\right) - \frac{1}{N} \sum_{n=1}^{N} f^\star\left(D_\phi\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}, \mathscr{G}_\theta(x^{(n)}, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1})\right)\right) \right\}$$

Example: Generative ICON for Poisson

$$\begin{cases} -u''(x,\omega) = y(x,\omega) \\ u(0,\omega) = \alpha_1(\omega), u(1,\omega) = \alpha_2(\omega) \end{cases}$$

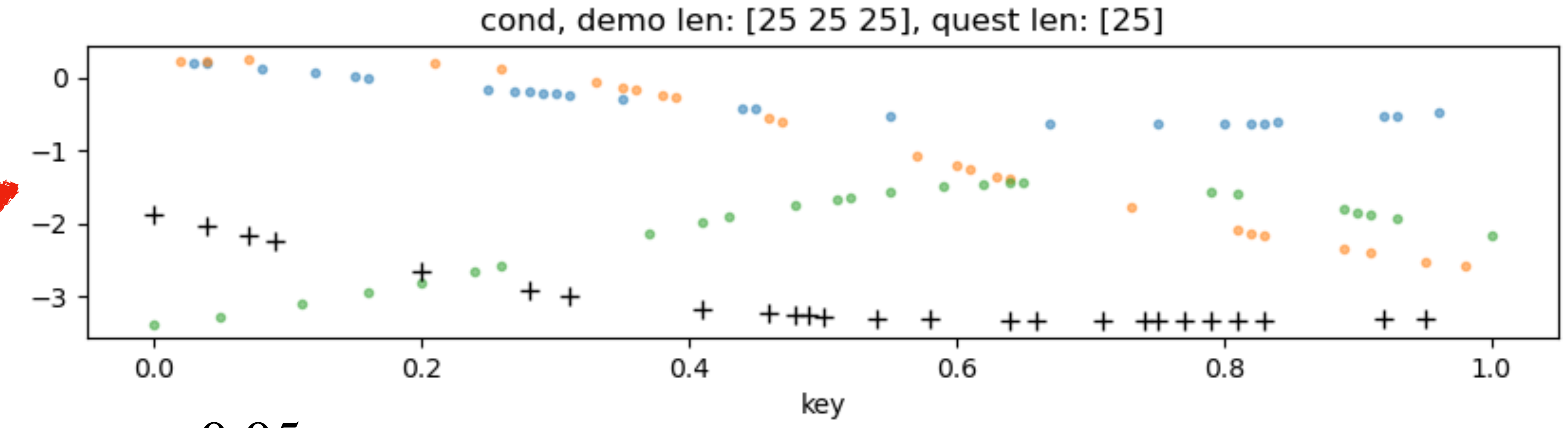
$z(x,\omega) = u(x,\omega) + \epsilon$

$\epsilon \sim \mathcal{N}(0,\sigma_o^2)$, for each x

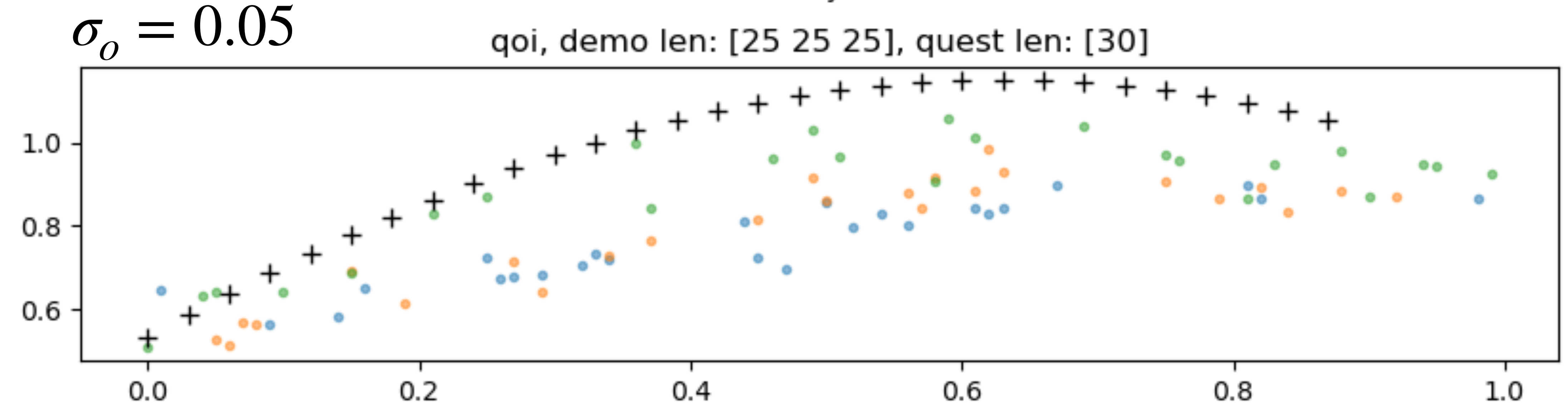
Parameters $\alpha_1(\omega)\ \alpha_2(\omega)$

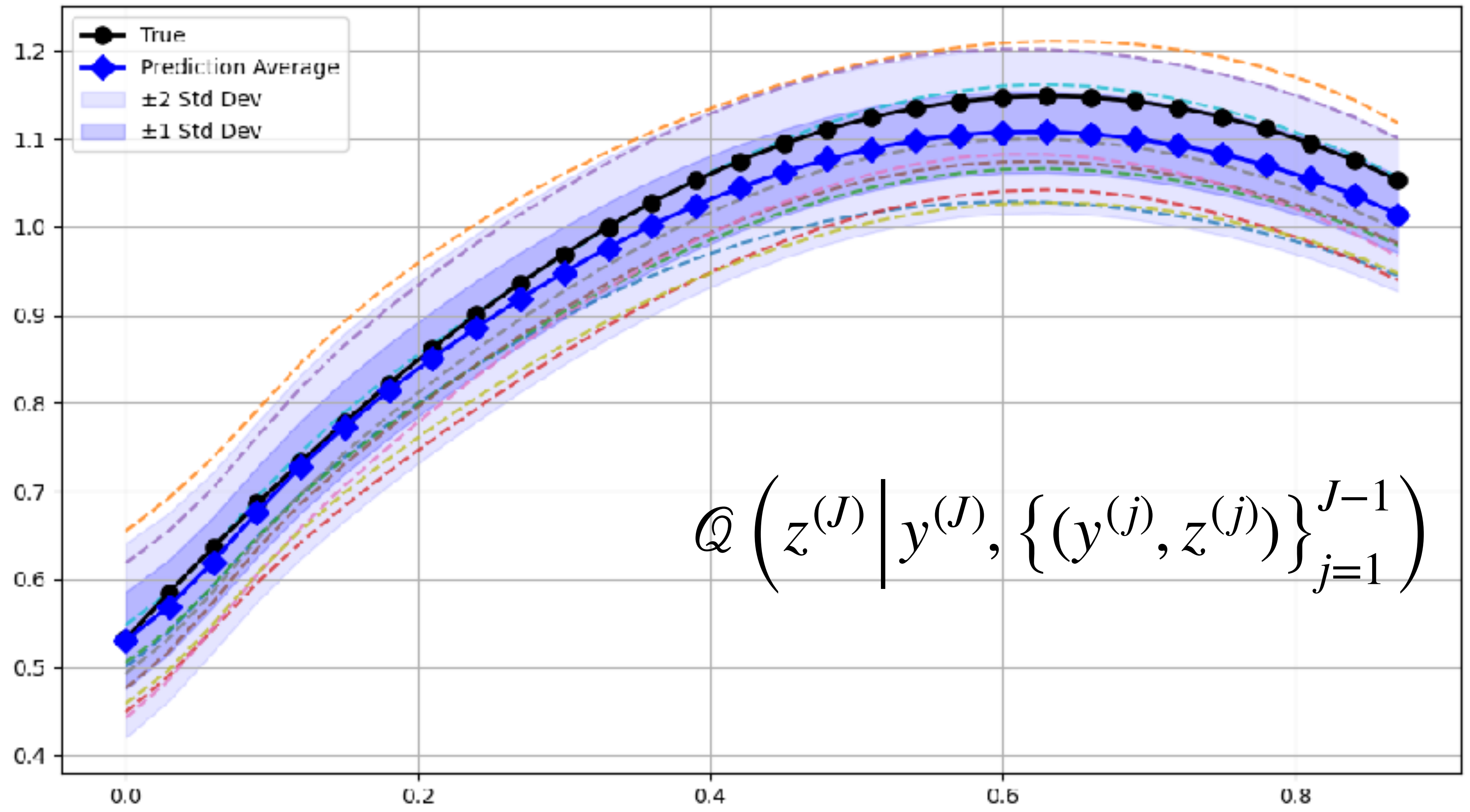
Conditions $y(x,\omega)$

QoI $z(x,\omega)$



$\sigma_o = 0.05$





Data generating process

- $y(x,\omega)$ Gaussian process with squared exponential covariance
- $\alpha_1 \sim \mathcal{U}[0,1]$, $\alpha_2 \sim \mathcal{U}[0,1]$
- $J - 1 = 3$ Example condition-QoI pairs
- Colored lines: example pairs
- Truth (+)
- Uncertainty bands computed from generated samples from $\mathcal{Q}\left(z^{(J)} \middle| y^{(J)}, \{(y^{(j)}, z^{(j)})\}_{j=1}^{J-1}\right)$

Conclusion and outlook

In-context operator learning implicitly performs Bayesian inference for differential equations

- **Random differential equations** provides a framework for probabilistic operator learning
- ICON directly approximates mean of the **Bayesian posterior predictive distribution**
- **Generative ICON** samples from the posterior predictive: a form of **uncertainty quantification**

Original ICON

$$\mathcal{T}^\star\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) = \mathbb{E}\left[z^{(J)} \middle| y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right]$$

- ICON **approximates mean** of posterior predictive distribution
- $\mathcal{T}^\star(\cdot,\cdot)$ is an operator learner

Theorem: Average generative ICON equals Original ICON

$$\mathcal{T}^\star\left(y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) = \int \mathcal{G}^\star\left(x, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right) p_X(x)\, dx$$

Generative ICON

$$\mathcal{G}^\star\left(\cdot, y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right)_\sharp p_X = Q\left(\cdot \middle| y^{(J)}, \left\{y^{(j)}, z^{(j)}\right\}_{j=1}^{J-1}\right)$$

- Generative ICON **generates** from the posterior predictive distribution
- $\mathcal{G}^\star(x,\cdot,\cdot)$ is a *sample* operator learner

Outlook

- Motivation to find Bayesian structure in other multi-task operator learners?
  - **PROSE [Liu et al.], BCAT [Liu et. al], POSEIDON [Herde et al.] and more…**
- Using other types of generative models beyond GANs

Thank you!