AK: Attentive Kernel for Information Gathering

Weizhe Chen, Roni Khardon, and Lantao Liu

Bardia Mojra
January 25, 2023
Robotic Vision Lab
The University of Texas at Arlington

Introduction

- Mapping topological features in a salient environment with sparse samples and an online planner
- Robotic Information Gathering (RIG) planners
  - Rely on the uncertainty of the probabilistic topological model
- Common kernel for spatial modeling:
  - Gaussian process (GP)
  - Radial basis function (RBF)
- Informative planning:
  - RIG planners prioritize regions with the highest uncertainly

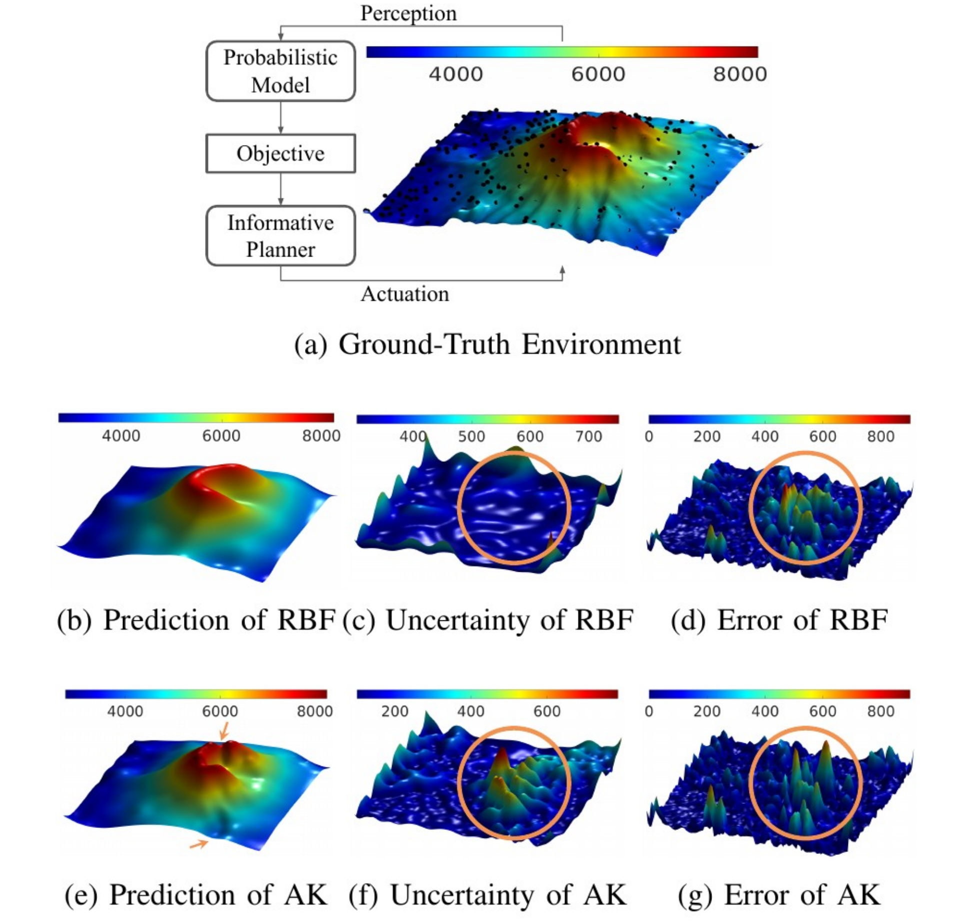


Fig. 1. **Comparison of GPR models with RBF kernel and the AK in terrain mapping**. The color indicates elevation, and black dots are training samples. The AK portrays the salient environmental features in more detail and assigns higher uncertainty to the high-error area.

Related Work

- Gaussian process regress (GPR) with stationary kernels:
  - Struggles to capture fine and random topological details
  - Inconsistent prediction error and uncertainty
- Non-stationary models:
  - Often too flexible to be trained
  - Categories:
    - Input-independent length-scale: correlation scales at different input locations
    - Input warping: maps input to a distorted space and assumed stationarity holds
    - The mixture of experts: a gating network that allows tuning of local hyper-parameters

Problem Statement

- Gaussian process regress (GPR) with stationary kernels:
  - Struggles to capture fine and never-seen-before topological features
  - Inconsistent prediction error and uncertainty
- Non-stationary models:
  - Often too flexible to be trained

Problem Statement – RIG

- To map an initially unknown envrionment efficiently using sparse active sensing.
- The goal is to find sampling locations that minimize the expected error after updating the model.
- Eq (1) cannot be used as an objective function as f_env is unknown.
- Eq (2) aims to find locations that minimize an information-theoretic objective function, e.g., entropy.
  - Assumes well-calibrated uncertainty model!
- The goal is to develop a kernel to improve:
  - Uncertainty quantification
  - Prediction accuracy

$$\arg\min_{\mathbf{X}_t} \mathrm{E}_{\mathbf{x}^\star}\left[\mathtt{error}\left(\mathrm{f}_{\text{env}}(\mathbf{x}^\star), \mu_t(\mathbf{x}^\star), \nu_t(\mathbf{x}^\star)\right)\right]. \quad (1)$$

$$\arg\min_{\mathbf{X}_t} \mathrm{E}_{\mathbf{x}^\star}\left[\mathtt{info}\left(\nu_t(\mathbf{x}^\star)\right)\right]. \quad (2)$$

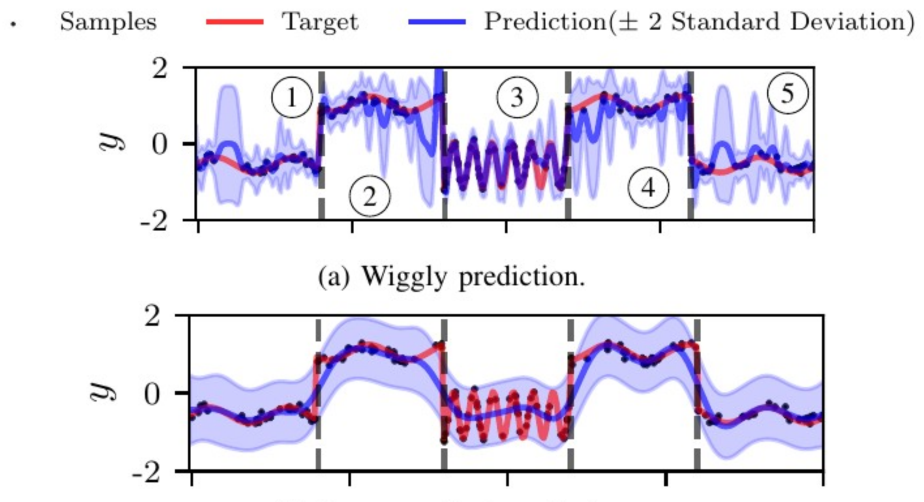
Method – Active Kernel

- Eq (3) defines the Active Kernel,
  - $$\mathrm{ak}(\mathbf{x}, \mathbf{x}') = \alpha \bar{\mathbf{z}}^{\intercal} \bar{\mathbf{z}}' \sum_{m=1}^{M} \bar{w}_m \mathrm{k}_m(\mathbf{x}, \mathbf{x}') \bar{w}'_m. \qquad (3)$$
  -
- It learns parametric functions that map each input x to w and z.
- Similarity attention scores for the set of base kernels
  - $\bar{w}_m \acute{\bar{w}}_m$
- Visibility attention score to mask the kernel value
  - $\bar{Z}^T \acute{Z}$
- Base kernels
  - $\{k_m(X, \acute{X})\}_{m=1}^{M}$

Method – Nonstationary Kernel

- RBF kernels, eq (4) with M evenly spaced lengthscales
- Hyperparameters: α , θ , ϕ , σ
- Input dependence:
  - Different lengthscales for different input locations
  - Break correlations among data points in different partitions
- In training, for every input location,
  - Selects a set of GPs with different predefined primitive lengthscales
  - Selects which training samples are used when making a prediction

$$\mathrm{k}_m(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|_2^2}{2\ell_m^2}\right), m = 1, \ldots, M. \quad (4)$$



(a) Wiggly prediction.

(b) Oversmoothed prediction.

Fig. 2. **Learning a nonstationary function using GPR with RBF kernel.**

Lengthscale & Instance Selection

- Softmax was used to make W differentiable w.r.t. θ
- Allows for gradual change of nonstationary kernels
- To accommodate abrupt changes in the input and loose correlations, they select similar data points by using a membership function.

$$\mathtt{f}(\mathbf{x}) = \sum_{m}^{M} \mathtt{w}_m(\mathbf{x})\mathtt{g}_m(\mathbf{x}), \text{ where} \tag{5}$$

$$\mathtt{g}_m(\mathbf{x}) \sim \mathcal{GP}(0, \mathtt{k}(\mathbf{x}, \mathbf{x}'|\ell_m)). \tag{6}$$

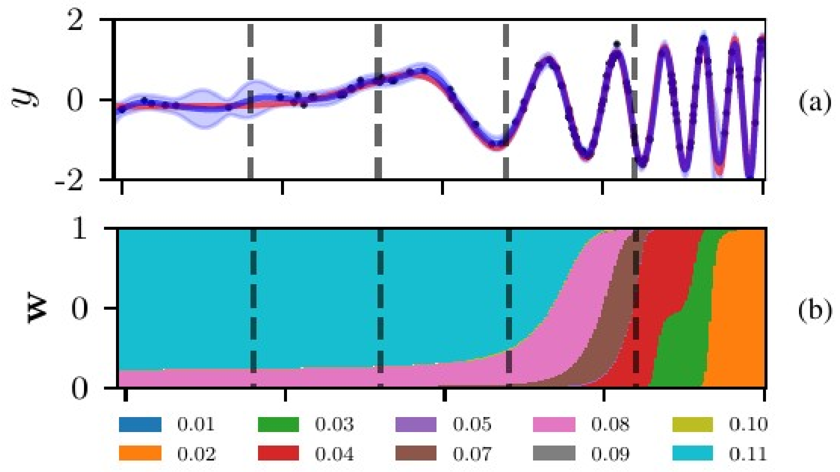


Fig. 3. **Learning $f(x) = x\sin(40x^4)$ with soft lengthscale selection.**

AK Model

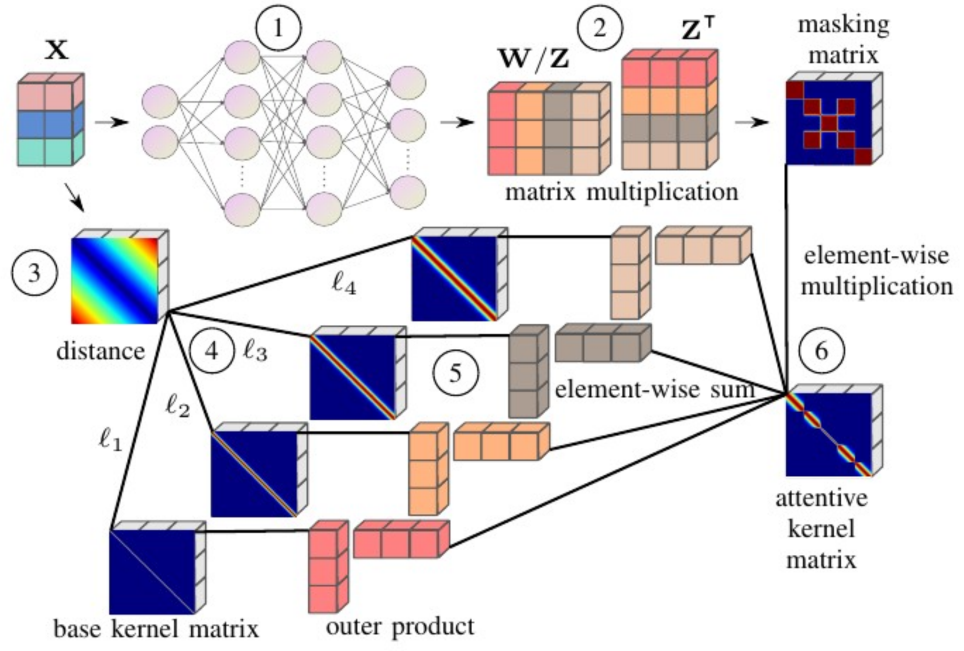


Fig. 6. **Computational diagram of the AK.**

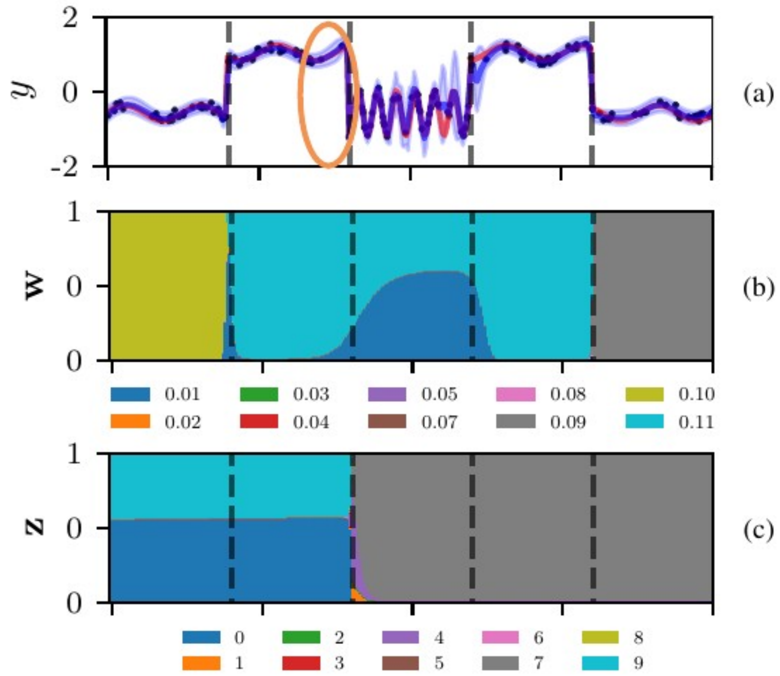


Fig. 5. **Learning the same function as in Fig. 2 using AKGPR.**

AK Model

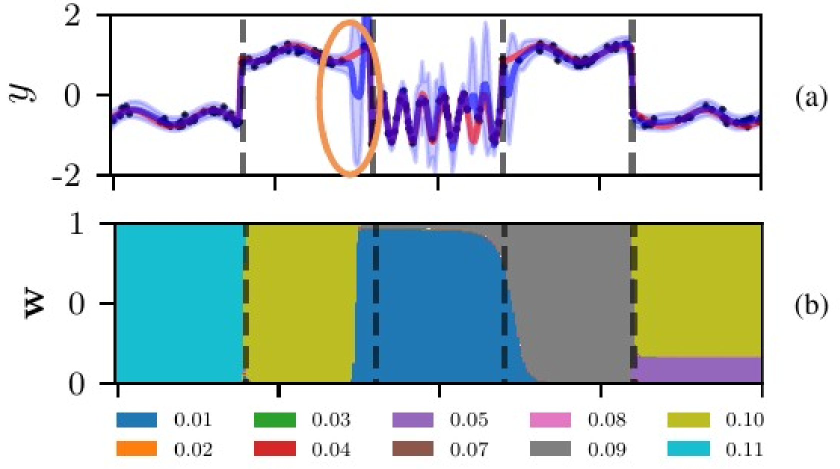


Fig. 4. **Learning the same function as in Fig. 2 using lengthscale selection.**

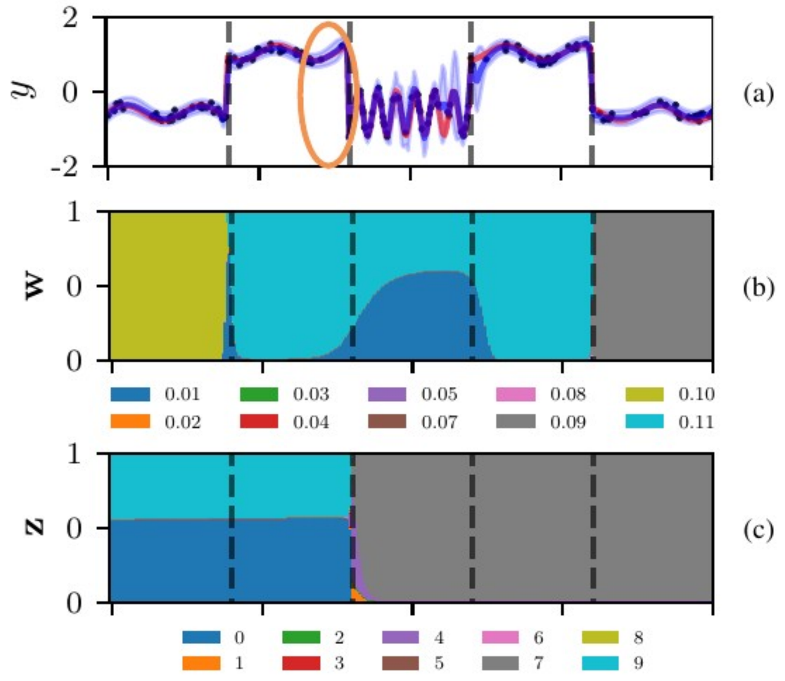


Fig. 5. **Learning the same function as in Fig. 2 using AKGPR.**

AK Model

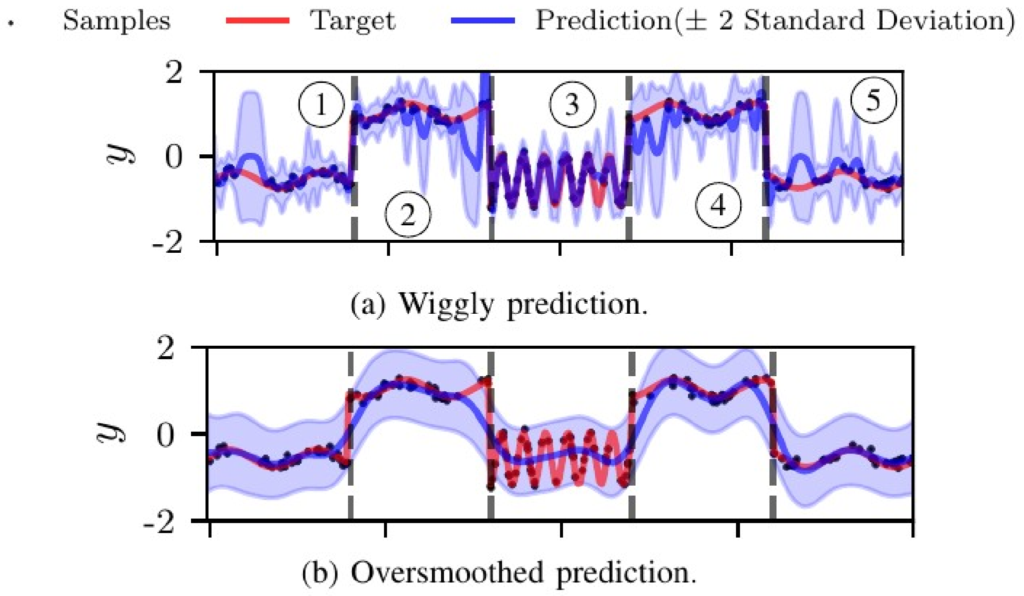


Fig. 2. **Learning a nonstationary function using GPR with RBF kernel.**

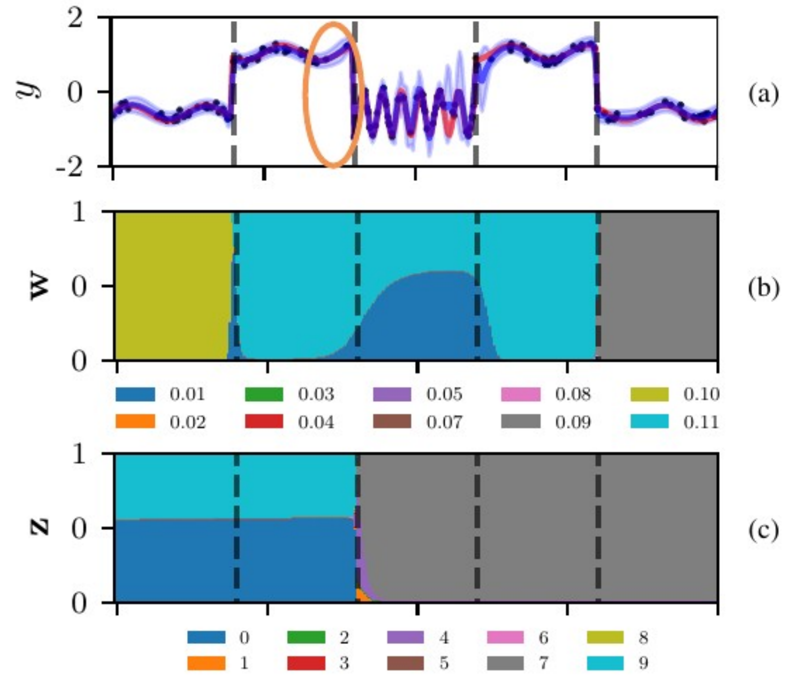


Fig. 5. **Learning the same function as in Fig. 2 using AKGPR.**

Experiments

- Refer to the paper.

- Thank you!