Contour location via entropy reduction leveraging multiple information sources

A. Marques, R. Lam, Massachusetts Institute of Technology
K. Willcox, University of Texas at Austin

Motivation

We address the problem of locating contours of functions that are expensive to evaluate. This problem arises in several areas of science and engineering:

classification | **constrained optimization** | **stability** | **reliability**

We consider the case when multiple information sources are available, in the form of relatively inexpensive, biased, and noisy approximations to the original function. **Our goal is to use information from all available sources to produce the best estimate of a contour under a fixed budget.**

Main contributions

- **Contour entropy:** a measure of uncertainty about the zero contour estimated by a statistical surrogate model
- **Acquisition function:** quantifies the expected reduction of contour entropy per unit cost
- **CLoVER** (**C**ontour **Lo**cation **V**ia **E**ntropy **R**eduction): an active-learning algorithm that that combines data from multiple information sources to locate contours of expensive functions at low cost

Problem statement

- Expensive function (information source 0): $g_0 : \mathcal{D} \subset \mathbb{R}^d \mapsto \mathbb{R}$
- Additional information sources: g_ℓ, $\ell \in \{1, \dots, M\}$ (biased, noisy)
- Known cost and variance: c_ℓ, λ_ℓ $\ell \in \{0, \dots, M\}$
- **Goal:** estimate $\mathcal{Z} = \{z \in \mathcal{D} \mid g_0(z) = 0\}$ at low cost using evaluations of all available information sources g_ℓ

Statistical multi-information source surrogate

Multi-information source Gaussian process (GP) surrogate introduced by Poloczek et al (NeurIPS 2017). This surrogate fits data from all available information sources and encodes the correlations between them.

$f(\ell, \boldsymbol{x})$ denotes the GP approximation to $g_\ell(\boldsymbol{x})$

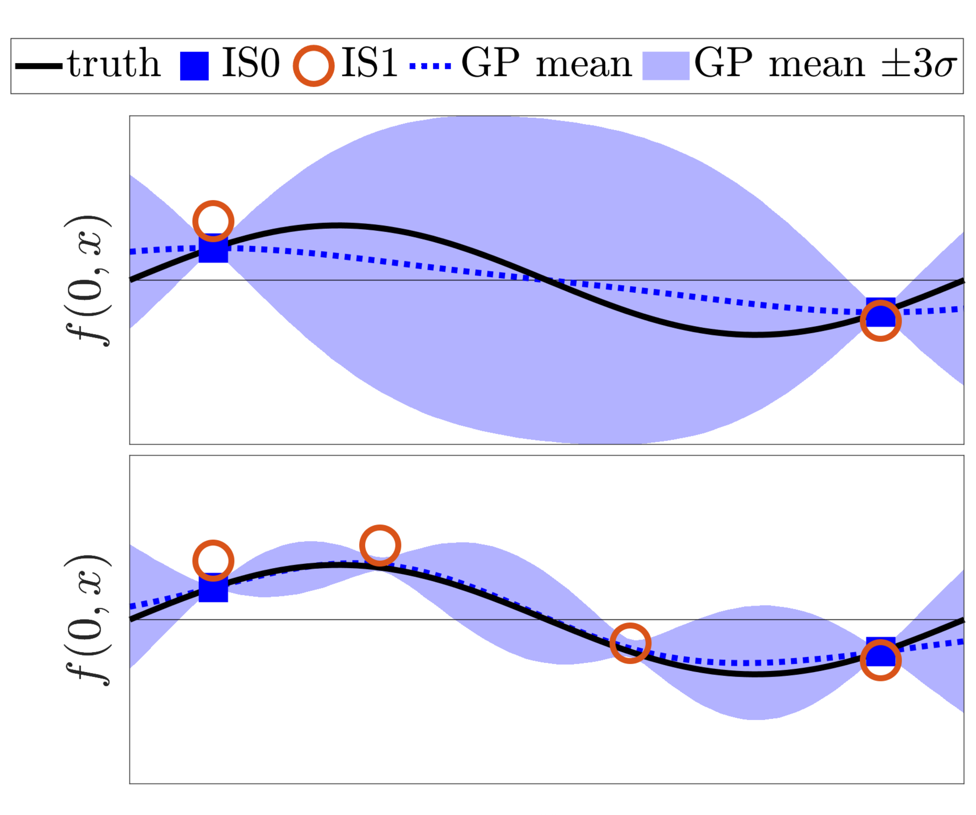


Figure 1: Multi-information source GP surrogate. Mean and variance of GP representation of expensive function g_0.

Top: two samples from both IS0 and IS1. Surrogate fits data and learns correlation between information sources.

Bottom: two additional samples of IS1 only. Surrogate improves the estimate of g_0 based only on IS1 data.

Contour entropy

For any $\boldsymbol{x} \in \mathcal{D}$, and $\epsilon > 0$, we define the following set of events based on the GP surrogate:

$f(0,\boldsymbol{x}) < -\epsilon$ (event L) | $|f(0,\boldsymbol{x})| < \epsilon$ (event C) | $f(0,\boldsymbol{x}) > \epsilon$ (event U)

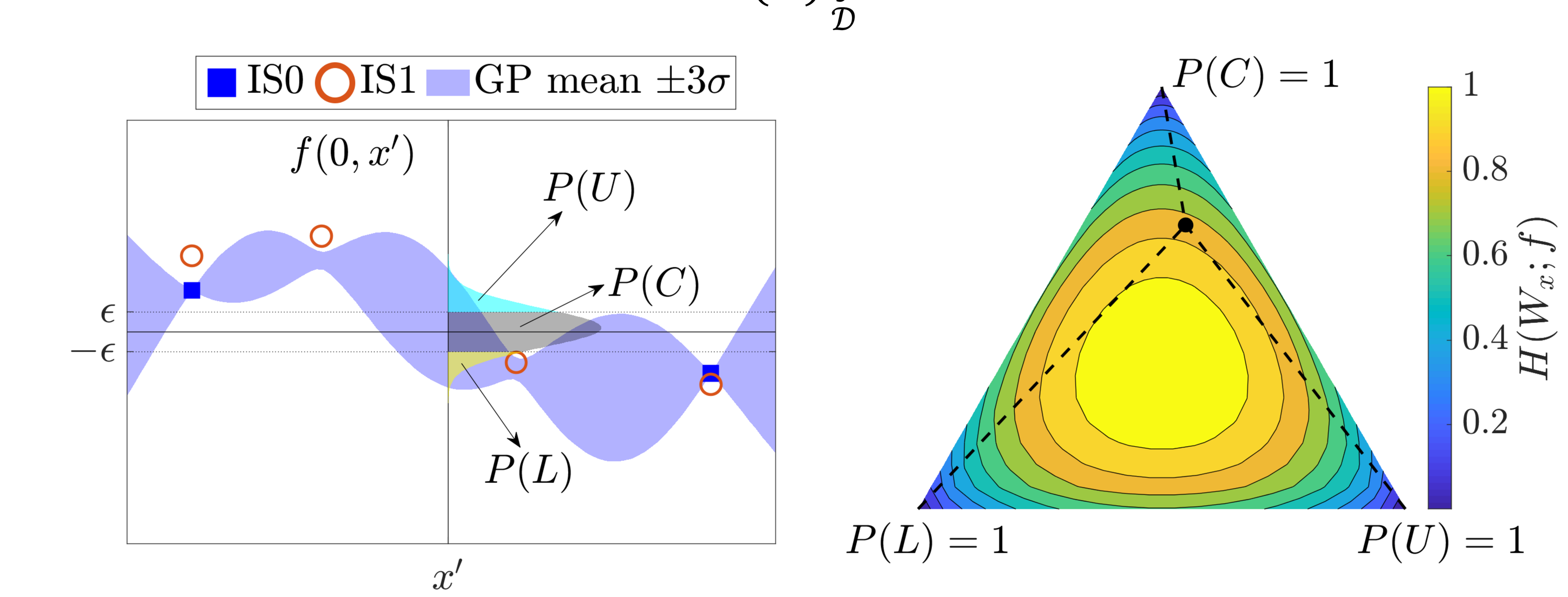
These events characterize whether the surrogate estimates $g_0(\boldsymbol{x})$ to lie below, within, or above a margin of width 2ϵ surrounding the zero contour.

We then define a discrete random variable W_x with probability mass $P(L)$, $P(C)$, and $P(U)$. The entropy of W_x represents a measure of uncertainty of the zero contour at $\boldsymbol{x}$:

$$H(W_{\boldsymbol{x}}; f) = -P(L)\ln P(L) - P(C)\ln P(C) - P(U)\ln P(U)$$

Contour entropy is the global measure of uncertainty of the zero contour:

$$\mathcal{H}(f) = \frac{1}{V(\mathcal{D})}\int_{\mathcal{D}} H(W_{\boldsymbol{x}}; f)\, d\boldsymbol{x}$$



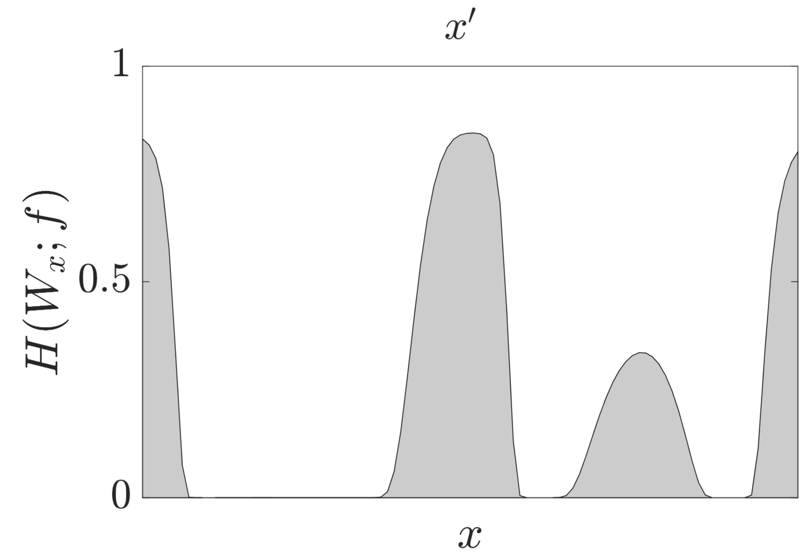


Figure 2: Top left: surrogate $f(0, x')$ at a particular x', and probability mass of events L, C, and U.

Top right: entropy $H(W_x; f)$ as a function of the probability masses. The black dot corresponds to $H(W_{x'}; f)$.

Bottom left: local entropy $H(W_x; f)$. Contour entropy is the area of gray shaded region.

Active learning algorithm

- After n function evaluations, CLoVER uses the current surrogate f^n as a generative model to solve the optimization problem:

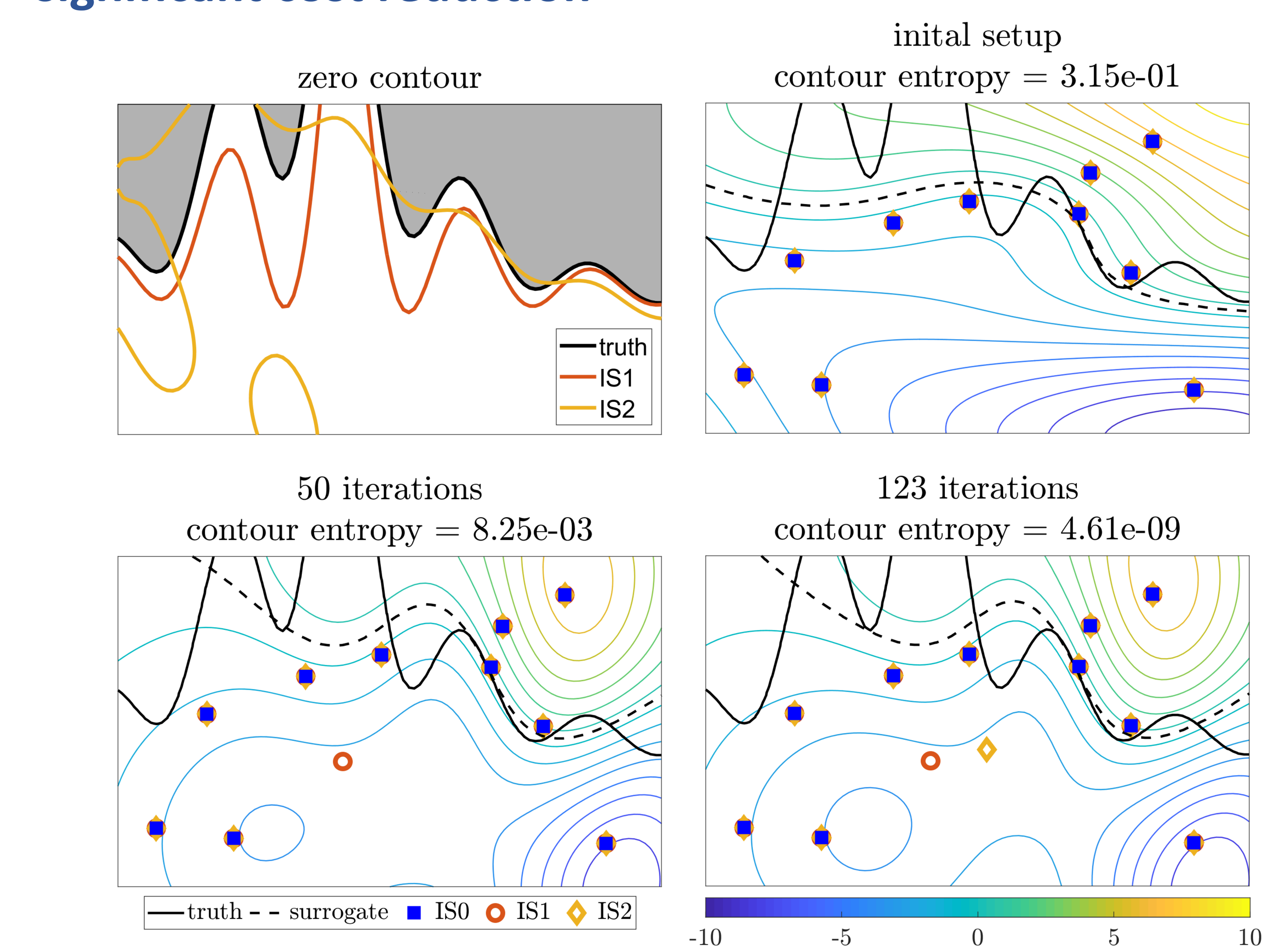
$$\underset{\ell \in \{0,\dots,M\},\, \boldsymbol{x} \in \mathcal{D}}{\text{maximize}}\; u(\ell, \boldsymbol{x}; f^n)$$

- **Acquisition function** measures expected reduction in contour entropy per unit cost:

$$u(\ell, \boldsymbol{x}; f^n) = \frac{\mathbb{E}[\mathcal{H}(f^n) - \mathcal{H}(f^{n+1}) \mid \ell^{n+1} = \ell,\ \boldsymbol{x}^{n+1} = \boldsymbol{x}]}{c_\ell(\boldsymbol{x})}$$

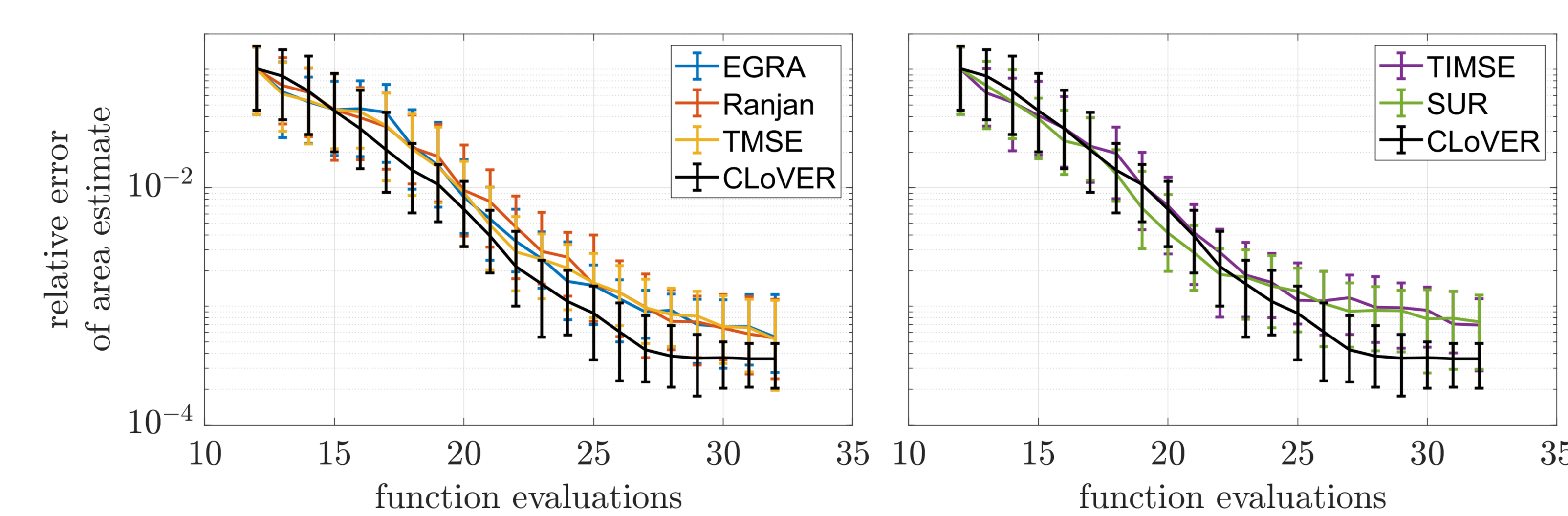
- Expectation is computed approximately with analytical expression (no numerical integration)
- Optimization is solved by a search over a discrete set of points

Considering multiple information sources can lead to significant cost reduction



- Multimodal function (Bichon et al., AIAA J., 2008)
- Results over 100 random initial evaluations:
  - Median cost of multi-information algorithm: 18.1
  - Median cost of single information source algorithm: 38.0

CLoVER outperforms similar algorithms



- Surrogate is used to compute area of Branin-Hoo function over the threshold $g_0 = 80$. Area is estimated with 10^6 Monte-Carlo samples
- **CLoVER achieves a smaller error on average when compared to similar algorithms available for single information source (R package KrigInv)**

Acknowledgements

This work was supported in part by the U.S. Air Force Center of Excellence on Multi-Fidelity Modeling of Rocket Combustor Dynamics, Award FA9550-17-1-0195, and by the AFOSR MURI on Managing Multiple Information Sources of Multi-Physics Systems, Awards FA9550-15-1-0038 and FA9550-18-1-0023.