Improvement-Based Acquisition Functions for Level Set Estimation

Anand Ravishankar
Dept. of Electrical Engineering
Stony Brook University
Stony Brook, New York

Fernando Llorente
Computing and Data Sciences Directorate
Brookhaven National Laboratory
Upton, New York

Petar M. Djurić
Dept. of Electrical Engineering
Stony Brook University
Stony Brook, New York

***Abstract*—Identifying regions where a function lies above or below a threshold is of significant interest to the scientific community. Estimating function values relative to a threshold is often framed in an active learning setting as a classification problem. Probabilistic models like Gaussian Processes (GPs) are often used to model the underlying function, and utility functions are used to infer the true class at the evaluation point. In this work, we introduce expected improvement-level set estimation (EI-LSE) and probability of improvement-level set estimation (PI-LSE), natural extensions of popular Bayesian optimization acquisition functions, to level set estimation (LSE). Besides providing theoretical guarantees on the misclassification rate, we evaluate these methods on synthetic and real-world datasets and compare them with other state-of-the-art LSE algorithms.**

***Index Terms*—Level Set Estimation, Bayesian Optimization, Acquisition Function**

I. Introduction

Level Set Estimation (LSE) involves using noisy evaluations of a function f to determine whether its true value exceeds a specified threshold h for various inputs. Accurately partitioning the function's domain is of considerable interest to many engineering and scientific applications. For example, in [1] the authors applied LSE algorithms to estimate geographical regions which have a high concentration of particulate matter, using sparse sensor data. In [2], the authors applied LSE to segment regions with high levels of chlorophyll and also to plan the placement of sensors to detect algae blooms. More recently, [3] proposed adaptive methods for identifying defective regions on material surfaces such as silicon wafers, in contrast to point-by-point physical examination. A key characteristic of LSE methods is their focus on achieving optimal performance with a minimal number of evaluations. This is essential in real-world scenarios where each evaluation carries an associated cost.

The idea of using sparse noisy evaluations of a black-box function is closely related to Bayesian Optimization (BO) [4], where one seeks to find the optimum of an unknown function. Formally, given noisy evaluations of a D-dimensional function f, we want to find $\mathbf{x}^* \in \arg\min_{\mathbf{x}} f(\mathbf{x})$. BO involves placing a prior over the underlying function and using a probabilistic surrogate model to improve the prior. An acquisition function (AF) suggests potentially information-rich points for evaluation, while maintaining a balance between exploring the parameter space and exploiting regions near the evaluated points.

This work was supported by NSF under Award 2212506.

Similar to BO, LSE also utilizes surrogate models and adaptive AFs to classify points with respect to h. Developing AFs for LSE generally involves borrowing concepts from the design of AFs for BO. Popular myopic AF designs for BO include GP-Upper Confidence Bound (GP-UCB) [5], Probability of Improvement (PI) [6], Expected Improvement (EI) [7] and Posterior Mean (PM) [8], among others. GP-UCB-based methods have been prominently developed for LSE [2] [9] [10]. These methods offer attractive theoretical guarantees in terms of the misclassification error or the convergence rate of the estimated level set to the true level set.

Initial work on LSE focused on placement of mobile sensors to estimate and track contours corresponding to ideal network conditions like minimum latency [11] [12] [13]. In [14], the authors introduced the "straddle" heuristic to classify points based on the distance to h and the uncertainty associated with the point. [2] formalized the problem of LSE and used a GP-UCB based AF to assign a confidence interval to each point, which is updated at each iteration until the algorithm is confident enough to classify the point in one region. The proposed method also returns an ϵ-accurate solution for the true set for some desired accuracy level. Since the proposed method was derived from GP-UCB for BO, it requires specifying a confidence parameter $\beta^{1/2}$, which adjusts the trade-off between exploration and exploitation. A popular heuristic dictates increasing this parameter with time, to promote exploring the entire parameter space. A more principled approach for selecting $\beta^{1/2}$ was suggested by [15] in a BO setting, which involves drawing it as a random sample from a chi-squared distribution with two degrees of freedom, instead of letting it be a function of time. [9] leveraged this approach for LSE and proposed a "randomized straddle" AF with theoretical bounds regarding misclassification loss in terms of information-theoretic terms. Other works at the intersection of BO and LSE are [16], [17].

Related work. While expected improvement (EI) and probability of improvement (PI) have been extensively studied in standard BO, generalized extensions to LSE remain under-explored, to the best of our knowledge. In [18], the authors assume a squared formulation for improvement, which, although natural for contour estimation, heavily penalizes points moderately further from h.

In this work, we introduce EI-LSE and PI-LSE, natural extensions of EI and PI in the context of contour estimation. We derive closed-form expressions and establish theoretical guarantees on the misclassification rate. In particular, we analyze their local behavior near h and prove improvement for uncertain points. Similar to [17], we include an exploratory term and discuss various heuristics for adjusting the associated confidence parameter.

The rest of the paper is structured as follows: Section II provides a brief introduction to GPs, which will be considered as the surrogate model of choice for the rest of the work. Section III introduces EI-LSE and PI-LSE and derives closed-form expressions. Information-theoretic based bounds for misclassification rates are also derived in this section. Section IV describes the implementation details along with results on synthetic and real-world data. Section V concludes the work with a brief note on future work.

II. Background

Let $f : \mathbb{R}^D \to \mathbb{R}$ be a black-box function that can be evaluated at $\mathbf{x} \in \mathbb{R}^D$ to obtain a noisy observation $y(\mathbf{x}) = f(\mathbf{x}) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and let $h \in \mathbb{R}$ be an explicit threshold level. The aim of LSE is to partition each point $\mathbf{x}$ into either a superlevel set $H = \{\mathbf{x} | f(\mathbf{x}) > h\}$ or a sublevel set $L = \{\mathbf{x} | f(\mathbf{x}) < h\}$.

Similar to BO, LSE uses probabilistic surrogate models like Gaussian Processes (GPs) [8] or Bayesian Neural Networks (BNNs) [19] to reason about f. In this work, we assume f is a sample path drawn from a zero-mean GP, $\mathcal{GP}(0, k)$, where k is a covariance function, which describes the statistical properties of the samples drawn from the GP. Given t observations $\mathbf{y}_t$ = $\{y_1, y_2, \dots, y_t\}$ made at points $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t\}$ with homogeneous noise variance σ_ϵ^2, the posterior over f is also a GP with the following mean and covariance functions:

$$\mu_t(\mathbf{x}) = \mathbf{k}_t^\top (\mathbf{K}_t + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{y}_t, \tag{1}$$

$$\sigma_t^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_t^\top (\mathbf{K}_t + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{k}_t(\mathbf{x}), \tag{2}$$

where $[\mathbf{K}_t]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $1 \leq i, j \leq t$, $\mathbf{k}_t(\mathbf{x}) = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}) & \dots & k(\mathbf{x}_t, \mathbf{x}) \end{bmatrix}^\top$.

III. Proposed Method

In this section, we derive closed form expressions for EI-LSE and PI-LSE. While in standard BO these AFs are used to guide the sampling for global optimum identification f^* [20], in the context of LSE we should seek to find points close to h, since identifying these points will implicitly partition the region into H and L. Formally, given a set of observed points $\{\mathbf{x}_i\}_{i=1}^t$ with corresponding function values, we should select points which maximally reduce the current best *gap* to the threshold h, which we define as follows:

$$g^* = \min_{1 \leq i \leq t} |f(\mathbf{x}_i) - h|.$$

A. Probability of Improvement for LSE

Let $g(\mathbf{x}) = |f(\mathbf{x}) - h|$ be our modified objective function, which we seek to minimize and let g^* be the current best function evaluation. We define PI-LSE($\mathbf{x}$) as

$$\begin{aligned}\text{PI-LSE}(\mathbf{x}) &= \mathbb{P}(g(\mathbf{x}) < g^*) = \mathbb{P}(|f(\mathbf{x}) - h| < g^*) \\ &= \mathbb{P}(h - g^* < f(\mathbf{x}) < g^* - h).\end{aligned} \tag{3}$$

Reparameterizing in terms of $z \sim \mathcal{N}(0, 1)$, we get

$$\text{PI-LSE}(\mathbf{x}) = \Phi_z\left(\frac{h + g^* - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right) - \Phi_z\left(\frac{h - g^* - \mu(\mathbf{x})}{\sigma(\mathbf{x})}\right), \tag{4}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote respectively the posterior mean and standard deviation functions in Eq (1), and $\Phi_z(\cdot)$ denotes the cumulative distribution function of the standard Gaussian distribution.

B. Expected Improvement for LSE

We formulate the EI-LSE acquisition function as

$$\text{EI-LSE}(\mathbf{x}) = \mathbb{E}\left[\max\left(g^* - |f(\mathbf{x}) - h|, 0\right)\right]. \tag{5}$$

Let us consider the following two improvement scenarios. (i) If $\mu(\mathbf{x})$ is above h, then the improvement can be achieved by selecting a point $\mathbf{x}^* \in [f^{-1}(h), f^{-1}(h + g^*)]$. (ii) If $\mu(\mathbf{x})$ is below h, then improvement can be achieved by selecting a point $\mathbf{x}^* \in [f^{-1}(h - g^*), f^{-1}(h)]$. Lastly, no improvement can be achieved if $\mu = h$. Equation (5) can be expressed as

$$\begin{aligned}&\text{EI-LSE}(\mathbf{x}) \\ &= \int_{-\infty}^{\infty} \max(g^* - |f(\mathbf{x}) - h|, 0) \mathcal{N}(\mathbf{x}; \mu(\mathbf{x}), \sigma^2(\mathbf{x})) df\end{aligned}$$

Splitting into the two improvement scenarios and reparameterizing, we get

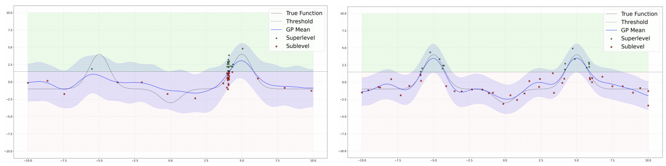
$$\begin{aligned}\text{EI-LSE}(\mathbf{x}) = &\underbrace{\int_{z_{\text{mid}}}^{z_{\text{high}}} (g^* + h - \mu(\mathbf{x}) - \sigma(\mathbf{x}) z) \cdot \phi(z) dz}_{\text{Case (i)}} \\ &+ \underbrace{\int_{z_{\text{low}}}^{z_{\text{mid}}} (g^* - h + \mu(\mathbf{x}) + \sigma(\mathbf{x}) z) \cdot \phi(z) dz}_{\text{Case (ii)}},\end{aligned}$$

$$\begin{aligned}\text{EI-LSE}(\mathbf{x}) = &(g^* + h - \mu(\mathbf{x}))\Phi(z_{\text{high}}) + 2(\mu(\mathbf{x}) - h)\Phi(z_{\text{mid}}) \\ &- (\mu(\mathbf{x}) + g^* - h)\Phi(z_{\text{low}}) \\ &+ \sigma(\mathbf{x})\left[\phi(z_{\text{low}}) + \phi(z_{\text{high}}) - 2\phi(z_{\text{mid}})\right].\end{aligned} \tag{6}$$

C. Inhibiting exploratory behavior

Many real-world and synthetic test functions are highly multimodal. Assuming a reasonable h, we would expect multiple regions where the points fall in both H and L. Given the current formulation of our AFs, once a region near h has been found, the methods would have no incentive to find other such regions.

As noted by [17], developing an intrinsically explorative strategy remains an open problem, without resorting to a



(a) (b)

Fig. 1. Illustration of the sampling trajectory mapped by PI-LSE with (a) β = 0.01 and (b) β = 0.2.

method such as the selection of points with maximum posterior variance. To this end, we adopt a similar approach and introduce a tuneable parameter β, which trades off with the posterior variance. Equations (6) and (4) are modified to get the final acquisition functions:

$$\text{PI-LSE}(\mathbf{x}) = \Phi_z\left(z_{\text{high}}\right) - \Phi_z\left(z_{\text{low}}\right) + \beta\sigma^2(\mathbf{x}) \tag{7}$$

and

$$\begin{aligned}\text{EI-LSE}(\mathbf{x}) = &(g^* + h - \mu(\mathbf{x}))\Phi(z_{\text{high}}) + 2(\mu(\mathbf{x}) - h)\Phi(z_{\text{mid}}) \\ &- (\mu(\mathbf{x}) + g^* - h)\Phi(z_{\text{low}}) \\ &+ \sigma(\mathbf{x})\left[\phi(z_{\text{low}}) + \phi(z_{\text{high}}) - 2\phi(z_{\text{mid}})\right] \\ &+ \beta\sigma^2(\mathbf{x}).\end{aligned} \tag{8}$$

Figure 1 shows the effect β has on the sampling trajectories. The test function in the figure is a multimodal function consisting of two Gaussian peaks and a central dip, with h = 1.5. Applying the proposed PI AF, with β = 0.01, results in the trajectory being concentrated toward the boundary found previously and the function missing out on one of the peaks completely. Upon increasing β to 2, we see that the trajectory is much more spread out and sampling ensures that the GP detects the true functional landscape.

The introduction of an exploratory term presents a challenge in selecting an appropriate value for β. We investigate three distinct heuristics for this parameter. (1) **Constant**: keeping β constant throughout the process. (2) **Incremental**: [5] suggest increasing β_t logarithmically and provide theoretical guarantees for regret bounds. We consider linearly increasing β_t with each iteration to gradually promote more exploration. (3) **Stochastic draw**: as shown by [15] for GP-UCB, selecting the confidence parameter from distributions like the χ^2_k or $\Gamma(\alpha, \lambda)$ leads to sub-linear Bayesian cumulative regret bounds. Following the same outline, we propose drawing β from a chi-squared distribution with two degrees of freedom, χ^2_2.

D. Theoretical Analysis

Level set estimation is inherently a classification problem with the aim of assigning points to H or L. We define the point-wise misclassification error as

$$\varepsilon_n(\mathbf{x}) = \mathbf{1}_{\{f(\mathbf{x})>h,\ \mu_n(\mathbf{x})\le h\}} + \mathbf{1}_{\{f(\mathbf{x})\le h,\ \mu_n(\mathbf{x})>h\}},$$

and the overall misclassification error as

$$\mathbb{E}[\varepsilon_n] = \int_{\mathcal{X}} \mathbb{P}(\varepsilon_n(\mathbf{x}) = 1)\, d\mathbf{x}.$$

The following theorem provides upper bounds on the overall misclassification error for EI-LSE and PI-LSE for scalar-valued inputs x, with the same bounds holding for vector-valued inputs.

Theorem 1 (Convergence of Misclassification Error for EI-LSE)**.** *For a discretized set of evaluation points $\{x_i\}_{i=1}^N$, with probability at least $1-\delta$, the expected misclassification error satisfies*

$$\mathbb{E}[\varepsilon_n] \le C\sqrt{\frac{\gamma_n}{n}} \tag{9}$$

for some constant $C > 0$.

Let $f : \mathcal{X} \to \mathbb{R}$ be an unknown function defined on a compact domain $\mathcal{X} \subset \mathbb{R}^d$. We assume f is a member of RKHS $\mathcal{H}_k$ associated with kernel k, with RKHS norm $\|f\|_{\mathcal{H}_k} \le B$. We model f with a GP and estimate the super-level set $\hat{H} : x \in \mathcal{X} : \mu_n(x) \ge h$, given the true super-level set $H : x \in \mathcal{X} : f(x) \ge h$.

Lemma 1. *With probability at least 1 - δ, $\forall x \in \mathcal{X}$*

$$|f(x) - \mu_n(x)| \le \sqrt{\beta_n}\sigma_n(x)$$

where $\beta_n = B + \sqrt{2(\gamma_n + \log(1/\delta))}$ (Lemma 5.1, [5]).

Lemma 2. *Let $x_1, x_2, \ldots x_{t-1}$ be the points evaluated by f. Then*

$$\sum_{t=1}^{n} \sigma^2_{t-1}(x_t) \le C_1\gamma_n,$$

where γ_n is the maximum information gain after n evaluations and $C_1 \ge 0$ (Lemma 4, [21]). Assume that at time t, the maximum variance is $\sigma^2_{max,t}$. Since Eq. (8) explicitly encourages exploration, EI-LSE will prioritize points with $\sigma^2_{t-1}(x) \approx \sigma^2_{max,t}$ and we get

$$\max_x \sigma^2_n(x) \le \frac{C_1\gamma_n}{n}.$$

The proof for Theorem 1 proceeds as follows: A misclassification at point x occurs when $(f(x)-h)(\mu_n(x)-h) \le 0$ or equivalently $|f(x)-\mu_n(x)| > |f(x)-h|$. Using Lemma 1, we get $|f(x)-h| \le \beta_n\sigma_n(x)$. So the probability of misclassification at x is bounded by:

$$P(\text{misclassification at } x) \le P(|f(x)-h| \le \sqrt{\beta_n}\sigma_n(x))$$

with expected misclassification rate over the domain as:

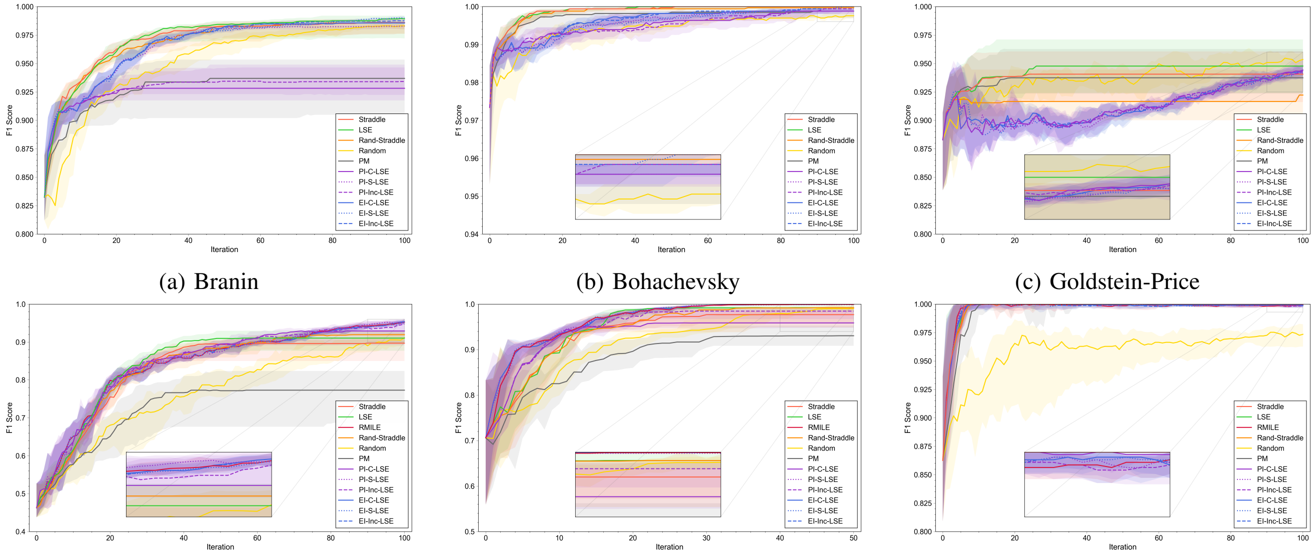
$$\mathbb{E}[\varepsilon_n] = \int_{\mathcal{X}} P(\text{misclassification at } x)dx. \tag{10}$$

Hence the expected error is bounded as follows:

$$\mathbb{E}[\varepsilon_n] \le \mu(\{x \in \mathcal{X} : |f(x)-h| \le \sqrt{\beta_n}\sigma_n(x)\}). \tag{11}$$

Assuming standard reproducing kernel k being Lipschitz continuous in its arguments, we consider f to be Lipschitz continuous with constant L, and the measure of this region is bounded by:

$$\mu(\{x \in \mathcal{X} : |f(x)-h| \le \sqrt{\beta_n}\sigma_n(x)\}) \le C_2 \cdot \sqrt{\beta_n}\max_{x\in\mathcal{X}}\sigma_n(x)$$



(a) Branin (b) Bohachevsky (c) Goldstein-Price

(d) Himmelblau (e) Rastrigin(1D) (f) Rastrigin(2D)

Fig. 2. Performance comparison of acquisition functions across nine benchmark functions. Each plot shows the F_1 score (vertical axis) progression over iterations (horizontal axis). The inset plot shows a zoomed-in region of the final F_1 score.

for some $C_2 > 0$. Using Lemma 2, we know that

$$\max_{x \in \mathcal{X}} \sigma_n(x) \leq \sqrt{\frac{C_1 \gamma_n}{n}}.$$

This allows us to finally bound the misclassification error using the maximum posterior variance as:

$$\begin{aligned} \mathbb{E}[\varepsilon_n] &\leq C_2 \cdot \sqrt{\frac{\beta_n C_1 \gamma_n}{n}}, \\ \mathbb{E}[\varepsilon_n] &\leq C \sqrt{\frac{\gamma_n}{n}}. \end{aligned} \tag{12}$$

For different kernels, γ_n values have been derived previously [5]. This also serves as an upper bound for PI, since it lacks an additional posterior standard deviation term in Equation 7. We also note that this bound is equal to the bound derived for Randomized Straddle (Theorem 4.2, [9]), further proving the effectiveness of EI-LSE and PI-LSE.

IV. Experiments and Results

In this section, we assess the performance of EI-LSE and PI-LSE on synthetic and real-world datasets. We compare our methods against the following state-of-the-art AFs: (i) Straddle, (ii) Level Set Estimation algorithm (LSE), (iii) Robust Maximum Improvement for Level-set Estimation (RMILE) and (iv) Randomized Straddle (Rand-Straddle). We also include two naive strategies for baseline performance: Random selection of points (Random) and selecting points with the highest posterior mean (PM). We denote the different heuristics for selecting β in EI-LSE and PI-LSE as follows: Constant $\beta = 0.1$ (EI / PI-C-LSE), incremental β_t (EI / PI-Inc-LSE), and stochastic draw (EI / PI-S-LSE). In our experiments, we considered a standard RBF kernel for all GP evaluations.

A. Synthetic Data

We considered six synthetic datasets for assessing the performance of EI-LSE and PI-LSE. Table (I) provides a brief description of each function and the associated h considered, with the mathematical formulation omitted due to space constraints. For each method, 20 independent simulations were carried out to account for simulation variability for 50 iterations for 1-D functions and 100 iterations for 2-D functions. The optimizers were initialized with 10 randomly drawn points for 1-D functions and 20 points for 2-D functions. To quantify the classification accuracy of the methods, we used the F_1-score. Figure 2 compares the performance of the AFs on the benchmark functions. It is evident that all variants of EI/PI-LSE deliver competitive results, with them notably outperforming other methods on functions like Himmelblau, Rastrigin-1D and 2D. It is also of key importance to note that the EI-LSE variants generally outperform the PI-LSE variants, which can be attributed to the additional exploratory term in Equation 8. Among the different heuristics for selecting β, we observe that drawing β from a χ^2_2 distribution generally offers a slightly better F_1 score. This improvement can be theoretically justified by considering the relationship between the χ^2_2 distribution and the confidence regions of GPs. When employing GP models, the squared Mahalanobis distance between a prediction and its mean follows a chi-squared distribution. By drawing β from χ^2_2, we are effectively sampling from the natural distribution of confidence levels around our predictions rather than using a fixed, potentially sub-optimal value. The code for all experiments can be found in this Github link.

TABLE I
BENCHMARK FUNCTIONS FOR LEVEL SET ESTIMATION

Function	Dim.	Domain	Threshold (h)
Rastrigin	1D	$[-5.12, 5.12]$	5
Himmelblau (neg.)	2D	$[-5, 5]^2$	-50
Rastrigin	2D	$[-5.12, 5.12]^2$	5
Bohachevsky	2D	$[-5, 5]^2$	5
Branin	2D	$[-5, 10] \times [0, 15]$	20
Goldstein-Price	2D	$[-2, 2]^2$	50

B. Real-world data

We consider the problem of quantifying the quality of silicon ingots in solar cells by classifying the carrier lifetime value as above or below a threshold. Additional information about the data can be found in [9] [22]. The function values are offset to set the threshold to 0. To simulate results similar to Randomized-Straddle, we use a Matern 3/2 kernel and initialize the optimizer with a singular point. Optimization is carried out for 200 iterations and repeated for 50 independent trials. Figure 3 shows the F_1 score against iterations. It is evident that EI-Inc-LSE performs comparatively well, as compared to the other methods, further proving the efficacy of using improvement-based methods for level set estimation.

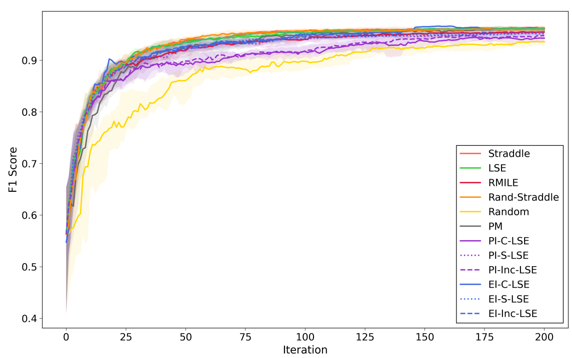


Fig. 3. F_1 score against iterations for classifying of silicon ingot carrier lifetime value

V. CONCLUSION

In this work, we have addressed the LSE problem from the perspective of Bayesian optimization and derived novel improvement-based AFs, PI-LSE and EI-LSE, for LSE. We also derive closed-form expressions and provide theoretical bounds on misclassification rate. The experiments show the effectiveness of PI-LSE and EI-LSE on synthetic and real-world datasets. Future directions could involve more principled ways of balancing exploration around the boundary regions.

REFERENCES

[1] P. Kearns, B. Jedynak, and J. Lipor, "A finite-horizon approach to active level set estimation," *arXiv preprint arXiv:2310.11985*, 2023.

[2] A. Gotovos, "Active Learning for Level Set Estimation," Master's thesis, Eidgenössische Technische Hochschule Zürich, Department of Computer Science,, 2013.

[3] S. Hozumi, K. Kutsukake, K. Matsui, S. Kusakawa, T. Ujihara, and I. Takeuchi, "Adaptive defective area identification in material surface using active transfer learning-based level set estimation," *arXiv preprint arXiv:2304.01404*, 2023.

[4] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[5] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*. Omnipress, 2009, pp. 1015–1022.

[6] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[7] J. Mockus, V. Tiesis, and A. Zilinskas, "Towards global optimization: A Bayesian approach," in *Proceedings of the 2nd International Symposium on Global Optimization*. Elsevier, 1978, pp. 117–121.

[8] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[9] Y. Inatsu, S. Takeno, K. Kutsukake, and I. Takeuchi, "Active learning for level set estimation using randomized straddle algorithms," 2024. [Online]. Available: https://arxiv.org/abs/2408.03144

[10] B. Bryan and J. Schneider, "Actively learning level-sets of composite functions," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 80–87. [Online]. Available: https://doi.org/10.1145/1390156.1390167

[11] K. Dantu and G. S. Sukhatme, "Detecting and tracking level sets of scalar fields using a robotic sensor network," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3665–3672.

[12] S. Srinivasan, K. Ramamritham, and P. Kulkarni, "Ace in the hole: Adaptive contour estimation using collaborating mobile sensors," in *2008 International Conference on Information Processing in Sensor Networks (ipsn 2008)*, 2008, pp. 147–158.

[13] A. Singh, R. Nowak, and P. Ramanathan, "Active learning for adaptive mobile sensing networks," in *2006 5th International Conference on Information Processing in Sensor Networks*, 2006, pp. 60–68.

[14] B. Bryan, R. C. Nichol, C. R. Genovese, J. Schneider, C. J. Miller, and L. Wasserman, "Active learning for identifying function threshold boundaries," *Advances in neural information processing systems*, vol. 18, 2005.

[15] S. Takeno, Y. Inatsu, and M. Karasuyama, "Randomized Gaussian process upper confidence bound with tighter Bayesian regret bounds," 2023. [Online]. Available: https://arxiv.org/abs/2302.01511

[16] I. Bogunovic, J. Scarlett, A. Krause, and V. Cevher, "Truncated variance reduction: A unified approach to Bayesian optimization and level-set estimation," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[17] A. Zanette, J. Zhang, and M. J. Kochenderfer, "Robust super-level set estimation using gaussian processes," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 276–291.

[18] P. Ranjan, D. Bingham, and G. Michailidis, "Sequential experiment design for contour estimation from complex computer codes," *Technometrics*, vol. 50, no. 4, pp. 527–541, 2008.

[19] D. J. C. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1992.

[20] R. Garnett, *Bayesian optimization*. Cambridge University Press, 2023.

[21] S. R. Chowdhury and A. Gopalan, "On kernelized multi-armed bandits," 2017. [Online]. Available: https://arxiv.org/abs/1704.00445

[22] K. Kutsukake, M. Deura, Y. Ohno, and I. Yonenaga, "Characterization of silicon ingots: Mono-like versus high-performance multicrystalline," *Japanese journal of applied physics*, vol. 54, no. 8S1, p. 08KD10, 2015.