Analysis of the Acquisition Function in the Paper

The acquisition function presented in the paper for Expected Improvement (EI) is described as:

$$EI_{s,t}(x) = \frac{\mathbb{E}_{p(y_{s,t})}\left[\max(y_{s,t} - y_t^*, 0)\right]}{\text{cost}(X_{s,t}, x_{s,t})}.$$

While this captures the general idea of improvement normalized by intervention cost, it lacks the essential components needed for a complete and practical implementation of Expected Improvement in Gaussian Process-based Bayesian Optimization. Here's an in-depth evaluation:

1. Limitations of the Paper's Formula

The formula provided in the paper has several critical omissions:

Absence of GP Mean and Variance

- The paper does not explicitly account for the posterior mean, $\mu_{s,t}(x)$, and variance, $\sigma_{s,t}(x)$, of the Gaussian Process (GP) model.
- These terms are fundamental for quantifying the expected improvement and the uncertainty associated with predictions.

Missing Standard Normal Terms

- The acquisition function does not include the cumulative distribution function (CDF, $\Phi(Z)$) or the probability density function (PDF, $\phi(Z)$) of the standard normal distribution, which are crucial for evaluating EI.

Ambiguity in Expectation

- The term $\mathbb{E}_{p(y_{s,t})}[\max(y_{s,t} - y_t^*, 0)]$ implies an expectation over the predictive distribution, but the paper does not detail how this is computed.
- In practice, this expectation is evaluated using GP posterior predictions.

2. Correct Formulation of the Acquisition Function

The standard Expected Improvement acquisition function, adapted for the problem setting, should be expressed as:

$$EI_{s,t}(x) = \frac{(\mu_{s,t}(x) - y_t^*)\,\Phi(Z) + \sigma_{s,t}(x)\phi(Z)}{\text{cost}(X_{s,t}, x_{s,t})},$$

where:

$$Z = \frac{\mu_{s,t}(x) - y_t^*}{\sigma_{s,t}(x)}.$$

Definitions:

- $\mu_{s,t}(x)$: GP posterior mean prediction at x.

- $\sigma_{s,t}(x)$: GP posterior standard deviation at x.
- y_t^*: Current best observed target value.
- $\Phi(Z)$: CDF of the standard normal distribution.
- $\phi(Z)$: PDF of the standard normal distribution.
- $\text{cost}(X_{s,t}, x_{s,t})$: Cost of the intervention.

3. Explanation of the Components

Exploration and Exploitation

- The term $\mu_{s,t}(x) - y_t^*$ encourages choosing points expected to yield improvements over the current best.
- The term $\sigma_{s,t}(x)$ incentivizes exploration of uncertain regions where the GP variance is high.

Probability Adjustment

- $\Phi(Z)$: Accounts for the probability that the improvement is non-negative.
- $\phi(Z)$: Adjusts for the magnitude of improvement under the GP posterior.

Cost Normalization

- Dividing by the intervention cost ensures that the optimization prioritizes feasible and cost-effective interventions.

4. Why the Paper's Formula is Incomplete

The paper's formula does not include:

Explicit Incorporation of $\mu_{s,t}(x)$ and $\sigma_{s,t}(x)$

- These are critical for the computation of EI, as they determine the expected improvement at any given point.

Role of $\Phi(Z)$ and $\phi(Z)$

- Without these standard normal distribution terms, the formula cannot account for the trade-off between exploration and exploitation.

Practical Computation

- The formula in the paper is theoretically suggestive but lacks implementation-ready details, making it less useful for replication or application.

5. Implications for Implementation

The corrected formula provides a robust framework for implementing Expected Improvement:

- It explicitly incorporates GP predictions ($\mu_{s,t}(x)$ and $\sigma_{s,t}(x)$).
- It balances exploration and exploitation using the standard normal terms.
- It adapts to real-world constraints by normalizing improvement with intervention costs.

In contrast, the formula in the paper is incomplete and could mislead practitioners unfamiliar with the details of acquisition function computation in Bayesian Optimization.

6. Conclusion

Yes, the acquisition function in the paper contains a typo and is incomplete. The corrected formula resolves these issues by explicitly including the GP posterior mean and variance, as well as the normal CDF and PDF. This correction aligns with standard practices in Bayesian Optimization and ensures the acquisition function is both theoretically sound and practically implementable.