Key Issues in the Paper: Dynamic Causal Bayesian Optimization

1. Prior Construction

- Section 3.3 introduces the dynamic causal GP model with a complex prior construction but does not provide a clear explanation of the practical computation of:

$$\hat{E}\left[f_Y^{NY}\left(x_{PY}, i_{PY}, w\right)\right].$$

- The paper mentions the use of do-calculus with observational data but omits the computational procedure.

2. Graph Evolution

- While Fig. 2 provides an overview of graph evolution over time, there is some ambiguity in how the algorithm handles transitions between states, particularly concerning the "square nodes" of previous interventions.
- The relationship between $G_{0:T}$ and G_t is not clearly defined.

3. Mathematical Derivations

- The proof of Corollary 1 and its derivation from Theorem 1 are not included in the paper.
- The expansion of:

  $$E_{p(w|do(X_{s,t}=x),i)}\left[f_Y^{NY}\left(x_{PY}, i_{PY}, w\right)\right]$$

  is missing intermediate steps.

4. Data Requirements

- The minimum required size of the observational dataset D_O is not specified.
- It is unclear how much historical interventional data is needed for effective optimization.

5. Model Updates

- The procedure for updating GP models with new interventional data is not described.
- The algorithm's method for maintaining and updating multiple GP models for different intervention sets is not explained.

6. Optimization Process

- The exact process for selecting the next intervention point using the acquisition function is omitted.
- Constraints on intervention values are not discussed.

7. Temporal Dependencies

- Handling different time scales or irregular temporal sampling is not addressed.
- The paper does not discuss how to account for delayed effects of interventions.

8. Practical Implementation

- The construction of the "dynamic causal prior" is not explained in the code.
- The paper does not provide guidance on selecting appropriate kernel functions for different types of variables.

9. Expected Improvement with Costs

The expected improvement function with costs, as implied by the paper, can be expressed as:

$$EI_{s,t}(x) = \frac{\sigma_{s,t}(x)\,[z\Phi(z) + \phi(z)]}{cost(X_{s,t}, x_{s,t})},$$

where:

- $\sigma_{s,t}(x)$ is the standard deviation of the GP posterior at x.
- z is the normalized improvement defined as:

  $$z = \frac{f(x^*) - \mu_{s,t}(x)}{\sigma_{s,t}(x)},$$

  where $f(x^*)$ is the best observed value and $\mu_{s,t}(x)$ is the GP mean at x.
- $\Phi(z)$ is the cumulative distribution function of the standard normal distribution.
- $\phi(z)$ is the probability density function of the standard normal distribution.
- $cost(X_{s,t}, x_{s,t})$ is the cost function associated with the intervention at $x_{s,t}$.