A Light-Weight Multi-Objective Asynchronous Hyper-Parameter Optimizer

Presentation of HOLA

March 24, 2025

Problem: Hyper-Parameter Optimization

- Task: Find optimal hyper-parameters $x \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_k$
- Multiple objectives to evaluate x:
  - $f_i : \mathcal{X} \to \mathbf{R}$ for $i = 1, \ldots, k$
  - Some to minimize, others to maximize
- Challenges:
  - Simulations are expensive (black-box)
  - No gradients available
  - Workers may be unreliable (stragglers, failures)
  - Need to balance exploration vs exploitation

Multi-Objective Scalarization

- Need to convert multiple objectives into single cost:

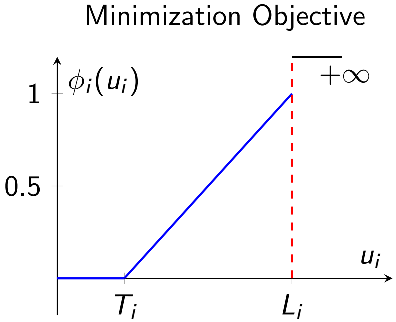
$$F(x) = \phi(f(x)) = \phi(f_1(x), \ldots, f_k(x))$$

- HOLA uses target-priority-limit scalarizer
  - Separable: $\phi(u) = \sum_{i=1}^{k} \phi_i(u_i)$
  - Each objective characterized by:
    - **Target** (T_i): Ideal value we aim for
    - **Limit** (L_i): Maximum/minimum acceptable value
    - **Priority** (P_i): Relative importance

Target-Priority-Limit Scalarizer

For minimization objective:

$$\phi_i(u_i) = \begin{cases} 0 & u_i \leq T_i \\ P_i \frac{u_i - T_i}{L_i - T_i} & T_i \leq u_i \leq L_i \\ +\infty & u_i > L_i \end{cases}$$



For maximization objective:

$$\phi_i(u_i) = \begin{cases} 0 & u_i \geq T_i \\ P_i \frac{T_i - u_i}{T_i - L_i} & L_i \leq u_i \leq T_i \\ +\infty & u_i < L_i \end{cases}$$

- Overall cost: $F(x) = \sum_{i=1}^{k} \phi_i(f_i(x))$
- Intuitively: 0 cost when better than target, infinite cost when worse than limit

Hyper-Parameter Attributes

Standardizing hyper-parameters for sampling:

- **Range**: Min/max values l_i and u_i
- **Scale**: Linear or logarithmic
- **Grid/discrete**: Finite set of values

Standardization: Map each parameter to $z_i \in [0, 1]$

- Linear scale: $z_i = \frac{x_i - l_i}{u_i - l_i}$
- Log scale: $z_i = \frac{\log x_i - \log l_i}{\log u_i - \log l_i}$

This standardizes the sample space, while respecting original constraints.

Optimization Method

Asynchronous exploration strategy:

1. **Initial phase**: Uniform sampling (or Sobol' sequence) until threshold
2. **Elite points**: Track top $r = \eta K$ simulations (e.g., $\eta = 0.2$)
3. **Statistical model**: Fit Gaussian mixture model to elite points
4. **Sample**: Generate new suggestions from the model

Comparison Groups and Pareto Ranking

- **Comparison Groups**: Group comparable objectives together
  - Each objective belongs to a comparison group
  - Within a group: Combine using priority-weighted sums
  - Example: $\Phi_g(x) = \sum_{i \in \text{group } g} \phi_i(f_i(x))$
- **Pareto/Nondominance Ranking**:
  - Calculate group scores: $\Phi(x) = (\Phi_1(x), \Phi_2(x), \ldots, \Phi_G(x))$
  - Solution x dominates y if:
    - For all groups g: $\Phi_g(x) \leq \Phi_g(y)$
    - For at least one group g: $\Phi_g(x) < \Phi_g(y)$
  - Solutions ranked by their Pareto front level
  - Elite set selected across multiple Pareto fronts