日期：2025.3.5

Q1：reward未与上一步step形成关联，需调整reward方式。

A1：仅考虑当前参数与目标的绝对误差，缺乏对参数变化方向的引导。

方案1：

| **奖励成分** | **计算公式** | **作用** |
| --- | --- | --- |
| 绝对误差奖励 | 100 \* exp(-MSE) | 保证最终收敛到目标参数 |
| 方向奖励 | sign(Δerror)\*50 | 增加方向引导 |  |  |
| 复合奖励 | 0.7绝对+0.3方向 | 平衡长期目标与短期优化 |  |  |

效果：曲线仍不收敛

问题分析：

1.改进动作空间映射：将Actor输出直接映射到参数有效范围，避免后续clip操作；替换原有Actor类，初始化时传入param\_ranges ×参数仍然趋于边界值

2. 增强奖励函数：引入滞回曲线形状相似性奖励（占比30%）；使用指数衰减形式强化形状匹配的重要性 ×参数仍然趋于边界值

3. 动态边界惩罚：当参数进入边界10%区域时施加二次函数增长的惩罚；惩罚最大值限制为50，防止奖励值域失控 ×参数仍然趋于边界值后在边界值附近跳跃并再次趋于边界值

可能是 train() 函数中对 Actor 或 Critic 的反向传播更新存在错误？并未采样 buffer 来更新网络？没有编写损失函数来对网络进行反向传播？

Q2：建立word，及时保存并更新问题与改进方案。

Q3：可以尝试先固定部分参数，观察reward收敛情况。

A3：固定第三个参数，其余保持不变。

Q4：滞回曲线绘制太粗糙，增加数据点。

A4：使用插值（np.interp）生成2000个点。

日期：2025.3.12