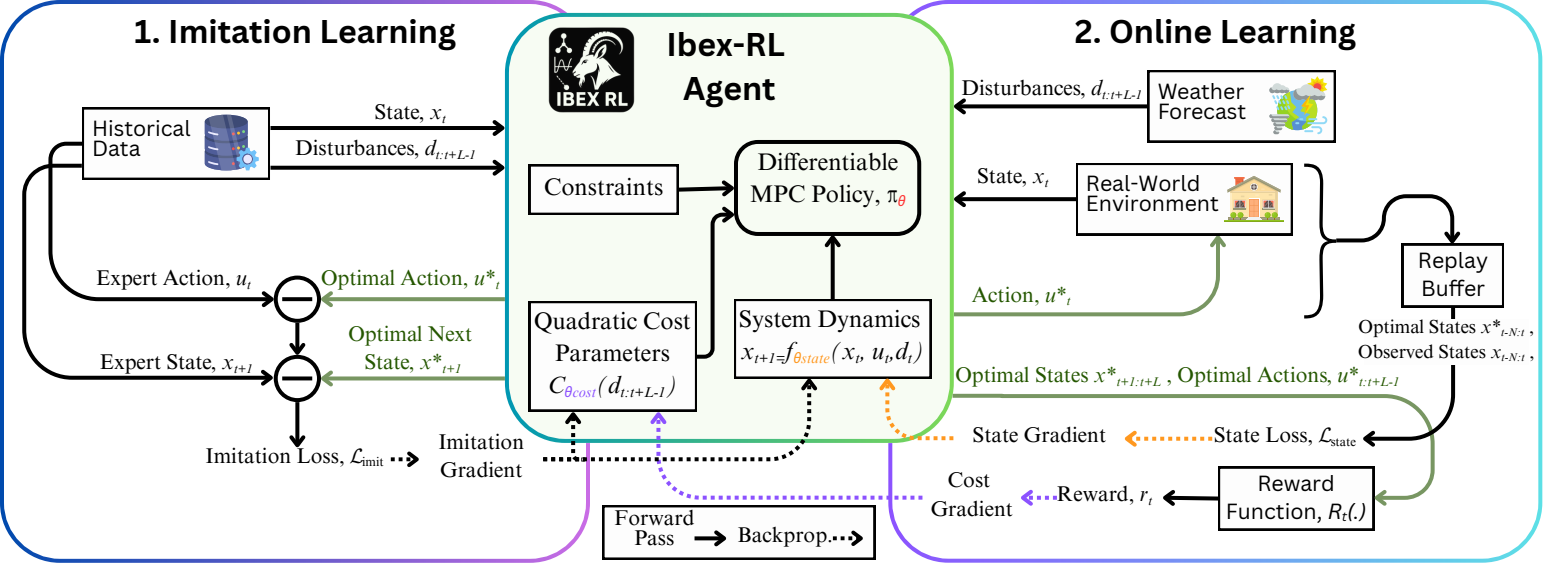
1. Imitation Learning
Historical Data
State, x_t
Disturbances, $d_{t:t+L-1}$
Expert Action, u_t
Optimal Action, u^*_t
Expert State, x_{t+1}
Optimal Next State, x^*_{t+1}
Imitation Loss, $\mathcal{L}_{imit}$
Imitation Gradient
Ibex-RL Agent
IBEX RL
Constraints
Differentiable MPC Policy, π_θ
Quadratic Cost Parameters $C_{\theta cost}(d_{t:t+L-1})$
System Dynamics $x_{t+1}=f_{\theta state}(x_t, u_t, d_t)$
Forward Pass
Backprop.
2. Online Learning
Disturbances, $d_{t:t+L-1}$
Weather Forecast
State, x_t
Real-World Environment
Action, u^*_t
Replay Buffer
Optimal States $x^*_{t-N:t}$, Observed States $x_{t-N:t}$,
Optimal States $x^*_{t+1:t+L}$, Optimal Actions, $u^*_{t:t+L-1}$
State Gradient
State Loss, $\mathcal{L}_{state}$
Cost Gradient
Reward, r_t
Reward Function, $R_t(.)$