ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Disturbance-Adaptive Differentiable MPC for Underwater Structure Inspection using Underwater Robot

Minjong Kim[1], Bonchul Ku[1], Dongsub Kim[1], Young-woon Song[1],
Woojin Seol[2], Youhyun Jang[2], and Son-Cheol Yu[1]

[1]Department of Convergence IT Engineering, POSTECH, Pohang, Republic of Korea
[2]Central Research Institute, Korea Hydro & Nuclear Power (KHNP), Daejeon, Republic of Korea

Abstract— **Close-proximity inspection of underwater cylindrical structures is challenging due to nonlinear vehicle dynamics, flow disturbances, and payload uncertainty. Fixed-weight MPC provides structured constraint handling but lacks adaptivity, while model-free RL is adaptive but often unstable and unsafe under disturbances. We propose Marine AC-MPC, which combines a differentiable iLQR-based MPC layer with an actor-critic framework that learns time-varying MPC cost weights online. In MarineGym, the proposed method achieves more reliable orbit tracking and higher success rates than fixed-weight MPC and PPO baselines under disturbed conditions.**

I. INTRODUCTION

Cylindrical marine infrastructures such as subsea pipelines and offshore jacket legs accumulate corrosion and fatigue, requiring periodic close-proximity inspection [1]. Autonomous execution demands a controller that maintains a prescribed standoff radius and stable attitude while orbiting cylindrical structures under nonlinear 6-DOF hydrodynamics, time-varying currents, and parametric uncertainty (Fig. 1).

Two established control paradigms often fall short in this regime. Fixed-weight model predictive control (MPC) handles constraints explicitly but cannot adapt cost weights to non-stationary disturbances; its horizon-latency trade-off results in either late error correction or computational intractability [2]. Model-free reinforcement learning (RL) accommodates disturbance variability in principle but provides no safety guarantees and exhibits high-variance learning under non-stationary transitions, as observed with proximal policy optimization (PPO)-based controllers [3].

We propose Marine AC-MPC, a disturbance-adaptive control framework that embeds a differentiable iLQR-based MPC layer as the terminal layer of an actor-critic network, building on recent work on differentiable MPC for agile flight [4]. Our contributions are: (i) a real-time differentiable iLQR controller with constraint-aware optimization; (ii) an actor-critic architecture learning time-varying cost weights against flow and payload disturbances; and (iii) quantitative evaluation in MarineGym.

II. METHOD

A. Problem Formulation

Let $c \in \mathbb{R}^3$ denote the cylinder center and $p \in \mathbb{R}^3$ the robot position. With horizontal distance $d_{xy} = \|p_{xy} - c_{xy}\|_2$ and target standoff r_{orbit}, the controller must regulate the radial error $e_{\text{radial}} = d_{xy} - r_{\text{orbit}}$ and a target tangential speed $v_{\text{tan,ref}}$ while satisfying

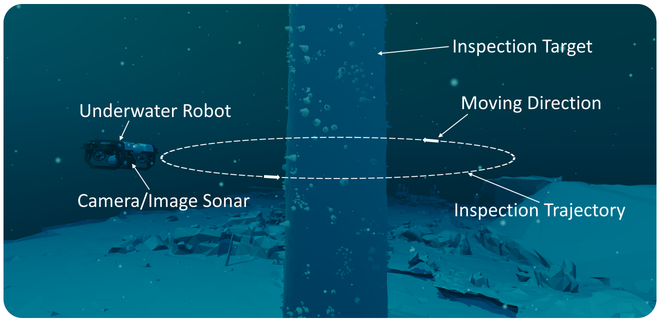

Fig. 1. An example of inspection task.

$$\begin{gathered} d_{\min} \le d_{xy} \le d_{\max}, \quad |\phi| \le \phi_{\max}, \quad |\theta| \le \theta_{\max}, \\ d_{xy} > r_{\text{cyl}} + \delta, \quad u \in [-1, 1]^{n_u}. \end{gathered} \tag{1}$$

An episode is successful if a full 2π orbit completes without violating any constraint.

B. Architecture

The actor π_θ maps each observation o_t to horizon-wise diagonal cost weights $\{Q^{(k)}, R^{(k)}\}_{k=0}^{H}$, mapped to positive ranges via a log-space sigmoid (Fig. 2). These parameterize a differentiable MPC layer that solves a finite-horizon constrained optimal control problem and outputs the control command u_t. Since the MPC layer is differentiable, the gradient is optimized at the control output and propagated back to θ, enabling end-to-end learning. A critic V_ψ estimates long-horizon value and supplies advantages for PPO updates with generalized advantage estimation (GAE) and a clipped value regression loss.

C. Differentiable iLQR Layer

At each time step t, the MPC layer solves the finite-horizon problem

$$\begin{aligned} u^\star_{t:t+H-1} = \arg\min_{u_{t:t+H-1}} \sum_{k=0}^{H-1} \Big[& e(x_{t+k})^\top W_e^{(k)} e(x_{t+k}) \\ & + u_{t+k}^\top W_u^{(k)} u_{t+k} \Big] \\ + \alpha_T \, e(x_{t+H})^\top & W_e^{(H)} e(x_{t+H}), \end{aligned} \tag{2}$$

---

*This work was supported by Korea Hydro & Nuclear Power Company Ltd., under Grant 2024-Robot Technology-001.

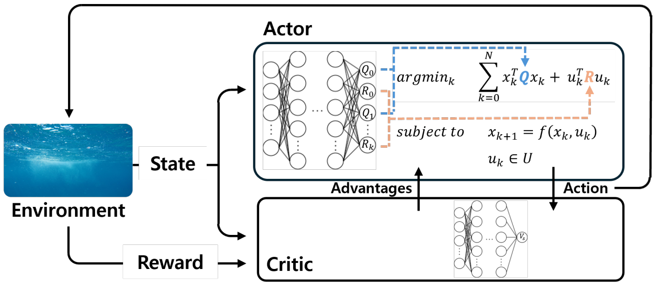

Fig. 2. A block diagram of the Marine AC-MPC.

subject to $x_{t+k+1} = f(x_{t+k}, u_{t+k})$ and $u_{\min} \leq u_{t+k} \leq u_{\max}$, where $e(x) \in \mathbb{R}^{10}$ encodes radial, depth, tangential, heading, and attitude errors, and $W_e^{(k)}, W_u^{(k)}$ are the learned diagonal weights. iLQR is warm-started by shifting the previous solution forward, and only $u_t = u_t^\star$ is applied in receding-horizon fashion. iLQR iterations are capped at 2-4 to bound runtime.

III. EXPERIMENTS

We evaluated in MarineGym [3] on an orbit-cylinder inspection task in which the vehicle must complete a 2π orbit at a prescribed standoff and tangential speed without violating safety constraints. Two disturbance sources are jointly randomized per episode: time-varying flow currents and payload mass, and inertia perturbations sampled at episode reset. Three controller families are compared using an identical evaluation protocol: fixed-weight MPC (acados [2]) at $H \in \{3, 6\}$; AC-PPO trained under random disturbances; and Marine AC-MPC at the same horizons. Performance is measured by success rate, radial RMSE, and tangential speed error.

TABLE I

COMPARISON OF CONTROL ACCURACY.

Model	Config	Success (%)	Radius RMSE (m)	Speed Error (m/s)
MPC	$H=3$	32	0.47	0.11
	$H=6$	36	0.41	0.13
AC-PPO		83	0.04	0.29
	w/o disturb.	80	0.17	0.11
Marine AC-MPC	$H=3$	83	0.14	0.26
	$H=6$	100	0.15	0.26

A. Quantitative Performance

Table I summarizes the comparison. Marine AC-MPC at H=6 achieves a 100% success rate with radial RMSE 0.15 m and speed error 0.26 m / s. Fixed-weight MPC achieves only 32-36% success, with a radial RMSE of 0.41-0.47 m, confirming that nominal MPC cannot maintain a safe orbit under random disturbances even when speed tracking is adequate. AC-PPO attains 83% success and the lowest radial RMSE (0.04 m) but the largest speed error (0.29 m/s), revealing that direct-action RL stabilizes the orbit at the expense of velocity tracking. Marine AC-MPC reaches 83% success even at H=3, demonstrating substantially reduced sensitivity to horizon tuning relative to fixed-weight MPC.

B. Trajectory Robustness

Fig. 3 compares orbit trajectories. Fixed-weight MPC at H=6 produces visible local irregularities under disturbance, while AC-PPO trained under random disturbance maintains a stable orbit but exhibits residual jitter and at least one episode terminated by aggressive recovery. Marine AC-MPC produces clean orbits at both horizons with no observed safety terminations. Increasing the horizon from H=3 to H=6 refines smoothness rather than enabling stability, indicating that learned cost adaptation compensates for short prediction windows.

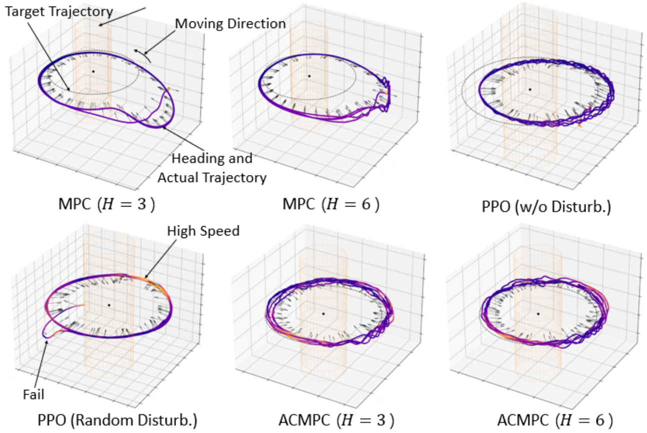

Fig. 3. Orbit trajectories under disturbed conditions.

IV. CONCLUSIONS

Marine AC-MPC reconciles the complementary deficiencies of fixed-weight MPC and model-free RL by combining prediction-based constraint enforcement with learned, state-conditioned cost adaptation. Empirically, it achieves perfect success under randomized flow and payload disturbances with zero safety terminations, providing a viable foundation for safe autonomous underwater inspection under real-time constraints.

REFERENCES

[1] F. Nauert and P. Kampmann, "Inspection and maintenance of industrial infrastructure with autonomous underwater robots," *Frontiers in Robotics and AI*, vol. 10, 2023.

[2] R. Verschueren et al., "acados-a modular open-source framework for fast embedded optimal control," *Mathematical Programming Computation*, vol. 14, no. 1, pp. 147-183, 2022.

[3] S. Chu, Z. Huang, Y. Li, M. Lin, D. Li, I. Carlucho, Y. R. Petillot, and C. Yang, "MarineGym: A high-performance reinforcement learning platform for underwater robotics," arXiv:2503.09203, 2025.

[4] A. Romero et al., "Actor-critic model predictive control: Differentiable optimization meets reinforcement learning for agile flight," *IEEE Transactions on Robotics*, vol. 42, pp. 673-692, 2025.