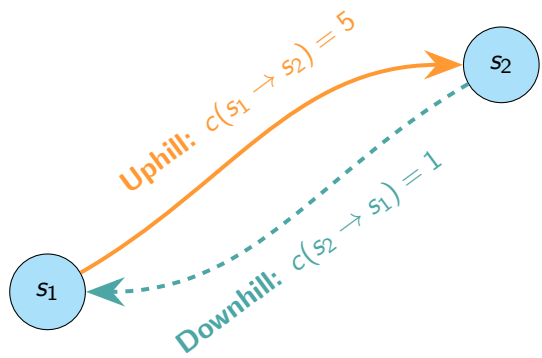
QPRL: Learning Optimal Policies with Quasi-Potential Functions for Asymmetric Traversal

Jumman Hossain[1], Nirmalya Roy[1]

[1]University of Maryland, Baltimore County, USA

Motivation: Asymmetric Traversal Costs

- Real-world robotic navigation often has **direction-dependent** and **irreversible** traversal costs.
- Traditional RL algorithms typically assume symmetry in costs.
- **Asymmetric costs**: uphill vs. downhill, irreversible transitions.
- Recent **quasimetric RL** approaches relax symmetry assumptions.
- However, they often neglect:
  - Explicit path-dependent cost modeling.
  - Rigorous safety guarantees.



Our Contributions: Quasi-Potential Reinforcement Learning

Novel Decomposition: $d(s, g) = \Phi(g) - \Phi(s) + \Psi(s \to g)$

- **Path-Independent Potential (Φ)**: Reusable costs, analogous to gravitational potentials.
- **Path-Dependent Residual (Ψ)**: Irreversible or dissipative costs, like friction.

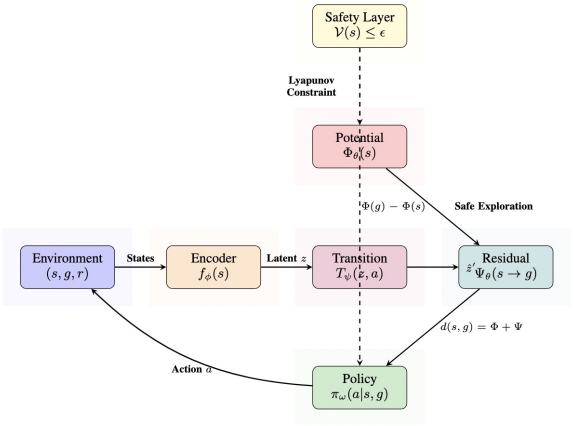
Benefits:

- Clear interpretability and accurate modeling of directionality.
- Enhanced exploration and efficient policy optimization.
- Safety via Lyapunov stability constraints.

Theoretical Advances:

- Improved convergence rate: $\tilde{\mathcal{O}}(\sqrt{T})$ **vs. previous** $\tilde{\mathcal{O}}(T)$.
- Lyapunov-based safety guarantees ensure minimal constraint violations.

QPRL Framework



- Decomposes asymmetric traversal costs into:
  - Path-independent potential Φ.
  - Path-dependent residual Ψ.
- Integrates **Lyapunov safety constraints** for stable exploration.

Algorithm: QPRL

Algorithm 1 Quasi-Potential Reinforcement Learning (QPRL)

Input: Replay buffer $\mathcal{D}$, learning rates $\alpha_\phi, \alpha_\psi, \alpha_\theta, \alpha_\omega$, threshold ϵ
for iteration $= 1$ to N **do**
  Sample batch $\{(s_i, a_i, s'_i, c_i, g_i)\}_{i=1}^{B} \sim \mathcal{D}$
  Update Encoder & Transition Model:
  $z_i = f_\phi(s_i)$, $\hat{z}'_i = T_\psi(z_i, a_i)$
  Update ϕ, ψ minimizing $\|\hat{z}'_i - f_\phi(s'_i)\|^2$
  Update Quasi-Potential Function Φ, Ψ:
  Update θ minimizing cost reconstruction and constraint losses
  Update Policy with Safety Layer:
  Update ω minimizing quasi-potential cost with safety constraints
end for

Algorithm: QPRL (Cont'd)

- Encoder f_ϕ and transition model T_ψ:
  - Compress state representation.
  - Predict next latent state efficiently.
- Quasi-potential Φ, Ψ:
  - Reconstruct asymmetric costs.
  - Ensure quasimetric constraints (triangle inequality, non-negativity).
- Policy optimization (safety enforced):

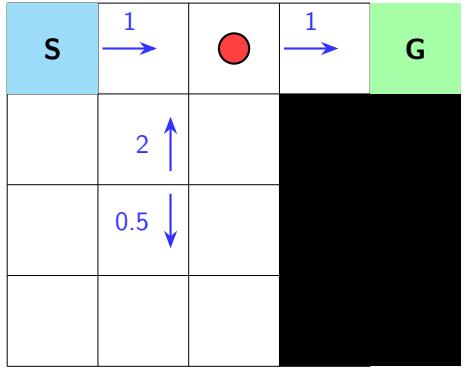
$$\mathbb{E}[\Phi_\theta(s')] \leq \Phi_\theta(s) + \epsilon$$

- Safety penalty in policy loss:

$$\mathcal{L}_\pi = \frac{1}{B} \sum_{i=1}^{B} [\hat{d}_i + \lambda \cdot \text{ReLU}(\Phi_\theta(\hat{z}_i') - \Phi_\theta(s_i) - \epsilon)]$$

- Dynamic Lagrange multiplier λ enforces safe transitions.

Experimental Evaluation: Asymmetric GridWorld



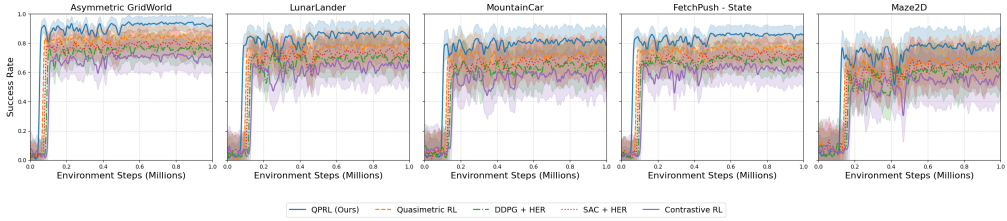
- Agent must navigate from start (**S**) to goal (**G**).
- Horizontal moves cost **1**.
- Climbing upward costs **2**, descending costs **0.5**.
- Walls are impassable, illustrating direction-dependent navigation.
- Evaluates QPRL's handling of asymmetric costs and safety constraints.

Experimental Results: Performance Comparison

Environment	Metric	QPRL (Ours)	QRL	Contrastive RL	DDPG+HER	SAC+HER
Asymmetric GridWorld	Success Rate (%)	**92.5 ± 2.2**	87.3 ± 3.0	82.4 ± 3.5	78.9 ± 4.2	80.3 ± 4.0
MountainCar	Normalized Return	**-95.6 ± 4.1**	-108.4 ± 6.7	-118.3 ± 8.1	-125.5 ± 7.6	-121.2 ± 7.0
FetchPush	Success Rate (%)	**91.2 ± 3.0**	85.5 ± 3.6	79.3 ± 4.1	73.8 ± 4.5	77.0 ± 4.3
LunarLander	Success Rate (%)	**88.9 ± 3.4**	81.4 ± 4.0	76.7 ± 4.5	72.5 ± 5.0	74.2 ± 4.8
Maze2D	Success Rate (%)	**85.3 ± 3.7**	78.1 ± 4.3	72.6 ± 4.7	68.9 ± 5.2	70.1 ± 4.9

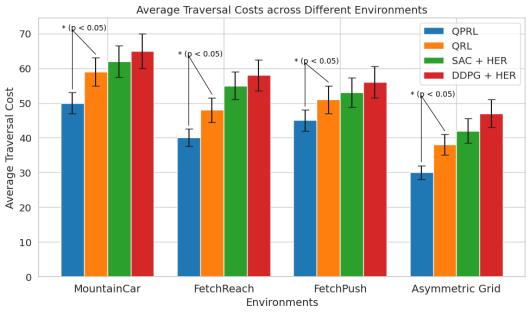
- QPRL consistently achieves **highest success rates** and **best returns**.
- Notably reduces variance across multiple random seeds.
- Demonstrates clear empirical advantage in asymmetric environments.

Performance Analysis: Learning Curves



- QPRL (blue line) achieves **high performance** earliest.
- Maintains **highest asymptotic success rates**.
- Shows **lower variance** in performance, indicating stability.
- Results statistically significant ($p < 0.01$, paired t-test).

Traversal Cost Comparison



Average traversal cost comparison. QPRL demonstrates the **lowest cost**, showing its advantage in exploiting asymmetric dynamics.

Env.	Method	Sym. (%)	Asym. (%)	Gap (%)
GridWorld				
	QPRL	94.1 ± 1.8	88.7 ± 2.5	5.4
	QRL	92.3 ± 2.0	83.5 ± 2.8	8.8
	SAC+HER	90.2 ± 2.3	81.0 ± 3.2	9.2
	DDPG+HER	89.8 ± 2.5	80.5 ± 3.5	9.3
MountainCar				
	QPRL	-90.5 ± 4.3	-98.2 ± 5.0	7.7
	QRL	-88.2 ± 4.1	-96.5 ± 5.2	8.3
	SAC+HER	-87.0 ± 4.0	-95.8 ± 5.3	8.8
	DDPG+HER	-86.5 ± 4.2	-94.5 ± 5.1	8.0
FetchPush				
	QPRL	92.0 ± 2.2	85.3 ± 3.1	6.7
	QRL	90.5 ± 2.3	81.0 ± 3.2	9.5
	SAC+HER	89.8 ± 2.5	79.8 ± 3.5	10.0
	DDPG+HER	88.5 ± 2.4	78.5 ± 3.4	10.0
LunarLander				
	QPRL	88.6 ± 3.4	82.4 ± 3.7	6.2
	QRL	87.0 ± 3.5	80.0 ± 4.0	7.0
	SAC+HER	85.5 ± 3.8	77.5 ± 4.2	8.0
	DDPG+HER	84.0 ± 3.6	76.0 ± 4.1	8.0

Conclusion

ICML International Conference On Machine Learning

- We proposed **Quasi-Potential Reinforcement Learning (QPRL)**, a RL framework tailored for **asymmetric traversal costs**.
- QPRL decomposes cost into **path-independent potentials** and **path-dependent residuals**, enabling efficient and interpretable learning.
- Achieves **state-of-the-art performance** in multiple tasks with improved **sample efficiency** and **reduced traversal costs**.
- Integrates **Lyapunov-based safety constraints** to avoid irreversible transitions during learning.
- Future work includes real-world deployment in:
  - **Topological navigation** with sparse rewards.
  - **Multi-agent systems** for safe coordination .