Intelligent Control – Team Projects

Instructor: Konstantinos Chatzilygeroudis (costashatz@upatras.gr)

March 29, 2025

Each team (maximum 2 students) will work on one of the following projects. The deliverables are as follows:

- Fully documented code (with README and instructions)
- A written report detailing the methodology, experiments, and results
- A live demonstration of the system (simulation or hardware)
- A final presentation

I. Model-Based and Model-Free Reinforcement Learning

1. **Model-Based RL with Learned Dynamics and Model-Free Policy Optimization**
   Topics: Model-Based RL, Neural Network Dynamics, Policy Learning
   Description: Learn a predictive model of environment dynamics from data (e.g., neural network ensemble or probabilistic model), then use the model to generate rollouts for training a model-free policy (e.g., via Dyna-style updates or model-based imagination). Test on continuous control tasks.

   Bibliography:
   a. Chatzilygeroudis, K., Vassiliades, V., Stulp, F., Calinon, S. and Mouret, J.B., 2019. A survey on policy search algorithms for learning robot controllers in a handful of trials. IEEE Transactions on Robotics, 36(2), pp.328-347. pdf.
   b. Model-based Reinforcement Learning Seminar. pdf.

2. **Implementing Soft Actor-Critic (SAC) from Scratch**
   Topics: Actor-Critic, Maximum Entropy RL, Continuous Control
   Description: Implement the SAC algorithm from scratch using PyTorch or JAX and apply it to robotics tasks. Analyze the impact of entropy regularization, target networks, and Q-function stability.

   Bibliography:
   a. Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P. and Levine, S., 2018. Soft actor-critic algorithms and applications. pdf.

II. Exploration, Imitation, and Diversity-Driven Control

3. **Intrinsic Motivation in Continuous Control Environments**
   Topics: Reinforcement Learning, Intrinsic Rewards, Exploration
   Description: Implement an intrinsic motivation strategy such as curiosity-based learning (e.g., using prediction error or Random Network Distillation) to improve exploration in sparse-reward environments with continuous actions (e.g., Pendulum, Swimmer, or BipedalWalker). Compare performance with vanilla RL approaches.

   Bibliography:
   a. Yuan, M., 202b. Intrinsically-motivated reinforcement learning: A brief introduction. pdf.
   b. Forestier, S., Portelas, R., Mollard, Y. and Oudeyer, P.Y., 202b. Intrinsically motivated goal exploration processes with automatic curriculum learning. Journal of Machine Learning Research, 23(152), pp.1-4a. pdf.

4. **Imitation Learning with DAGGER or Privileged Information**
   Topics: Imitation Learning, Dataset Aggregation, Partial Observability
   Description: Train a policy using demonstrations from an expert with full-state access, while the policy has only partial or noisy observations (e.g., images or low-dimensional features). Implement DAGGER to iteratively improve the learner. Compare to behavior cloning and analyze generalization under noise.

   Bibliography:
   a. Ross, S., Gordon, G. and Bagnell, D., 2011, June. A reduction of imitation learning and structured prediction to no-regret online learning. In Proceedings of the fourteenth international conference on artificial intelligence and statistics (pp. 627-635). pdf.
   b. Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V. and Hutter, M., 2020. Learning quadrupedal locomotion over challenging terrain. Science robotics, 5(47), p.eabc5986. pdf.

5. **Quality-Diversity for Control Policy Repertoires**
   Topics: Evolutionary Algorithms, Behavior Space, Repertoire Learning
   Description: Implement MAP-Elites or a similar QD algorithm to generate a diverse repertoire of behaviors for a quadruped robot. The algorithm should output many different ways of walking/running on a straight line. Define a suitable behavior descriptor and analyze how diversity contributes to robustness or task adaptation.

   Bibliography:
   a. Cully, A., Clune, J., Tarapore, D. and Mouret, J.B., 2015. Robots that can adapt like animals. Nature, 521(7553), pp.503-507. pdf.
   b. Chatzilygeroudis, K.I., Tsakonas, C.G. and Vrahatis, M.N., 2023, July. Evolving dynamic locomotion policies in minutes. In 2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-8). pdf.
   c. Vassiliades, V., Chatzilygeroudis, K. and Mouret, J.B., 2017. Using centroidal voronoi tessellations to scale up the multidimensional archive of phenotypic elites algorithm. IEEE Transactions on Evolutionary Computation, 22(4), pp.623-630. pdf.

III. Learning with Perception and Dynamics Variation

6. **Vision-Based RL for Goal-Conditioned Tasks**
   Topics: Image-Based Observations, Representation Learning, Goal-Conditioning
   Description: Train a policy using raw pixel observations from a simulated camera to solve a continuous control task (e.g., visual reaching). Use convolutional encoders and optionally include auxiliary tasks (e.g., reconstruction or contrastive learning). Use goal images for goal-conditioned policies.

   Bibliography:
   a. Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O. and Zaremba, W., 2017. Hindsight experience replay. Advances in neural information processing systems. pdf.
   b. Chane-Sane, E., Schmid, C. and Laptev, I., 2021, July. Goal-conditioned reinforcement learning with imagined subgoals. In International conference on machine learning (pp. 1430-1440). PMLR. pdf.
   c. Lu, C., Ball, P.J., Rudner, T.G., Parker-Holder, J., Osborne, M.A. and Teh, Y.W., 202b. Challenges and opportunities in offline reinforcement learning from visual observations. pdf.

7. **Learning Locomotion for a Quadruped Robot (Simulated)**
   Topics: Deep RL, Reward Design, Stability, Continuous Control
   Description: Train a policy using reinforcement learning (e.g., PPO or SAC) to enable a simulated quadruped robot to walk stably. Explore different reward functions and optionally apply curriculum learning or terrain variation for robustness.

   Bibliography:
   a. Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V. and Hutter, M., 2020. Learning quadrupedal locomotion over challenging terrain. Science robotics, 5(47), p.eabc5986. pdf.

8. **Domain Randomization for Sim-to-Sim Transfer**
   Topics: Domain Randomization, Generalization, Robust Policy Learning
   Description: Train a policy in a simulated environment with randomized physics parameters (e.g., mass, friction, sensor noise), and test its generalization to new unseen dynamics. Evaluate zero-shot transfer.

   Bibliography:
   a. Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W. and Abbeel, P., 2017, September. Domain randomization for transferring deep neural networks from simulation to the real world. In 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS) (pp. 23-30). pdf.

9. **Reward Shaping and Curriculum Learning for Complex Tasks**
   Topics: Reward Design, Curriculum Learning, Reinforcement Learning
   Description: Investigate how reward shaping or task curricula impact learning on difficult or sparse-reward tasks (e.g., standing up, object pushing, or balancing). Compare learning curves across different shaping strategies and optionally explore automated curriculum generation.

Bibliography:
a. Atanassov, V., Ding, J., Kober, J., Havoutis, I. and Della Santina, C., 2024. Curriculum-based reinforcement learning for quadrupedal jumping: A reference-free design. IEEE Robotics & Automation Magazine. pdf.
b. Portelas, R., Colas, C., Weng, L., Hofmann, K. and Oudeyer, P.Y., 2020. Automatic curriculum learning for deep rl: A short survey. pdf.
c. Xie, Z., Ling, H.Y., Kim, N.H. and van de Panne, M., 2020, December. Allsteps: curriculum-driven learning of stepping stone skills. In Computer Graphics Forum (Vol. 39, No. 8, pp. 213-224). pdf.

10. **Learning Agile Maneuvers for a 3D Quadrotor**
Topics: Reinforcement Learning, Agile Control, Trajectory Learning
Description: Use reinforcement learning to train a simulated quadrotor to perform agile maneuvers such as flips, dives, or slalom navigation. Focus on designing a reward function, selecting an appropriate policy architecture, and applying curriculum learning or shaping to improve convergence and robustness.

Bibliography:
a. Bellegarda, G., Nguyen, C. and Nguyen, Q., 2024. Robust quadruped jumping via deep reinforcement learning. Robotics and Autonomous Systems, 182, p.104799. pdf.
b. Coates, A., Abbeel, P. and Ng, A.Y., 2009. Apprenticeship learning for helicopter control. Communications of the ACM, 52(7), pp.97-105. pdf.
c. Song, Y., Romero, A., Müller, M., Koltun, V. and Scaramuzza, D., 202c. Reaching the limit in autonomous racing: Optimal control versus reinforcement learning. Science Robotics. pdf.