CANDID DAC:

Coupled ActioN Dimensions with Importance Differences

Philipp Bordne[1], M Assif Hasan[1], Eddie Bergman[1], Noor, Awad[1], André Biedenkapp[1]

[1]University of Freiburg

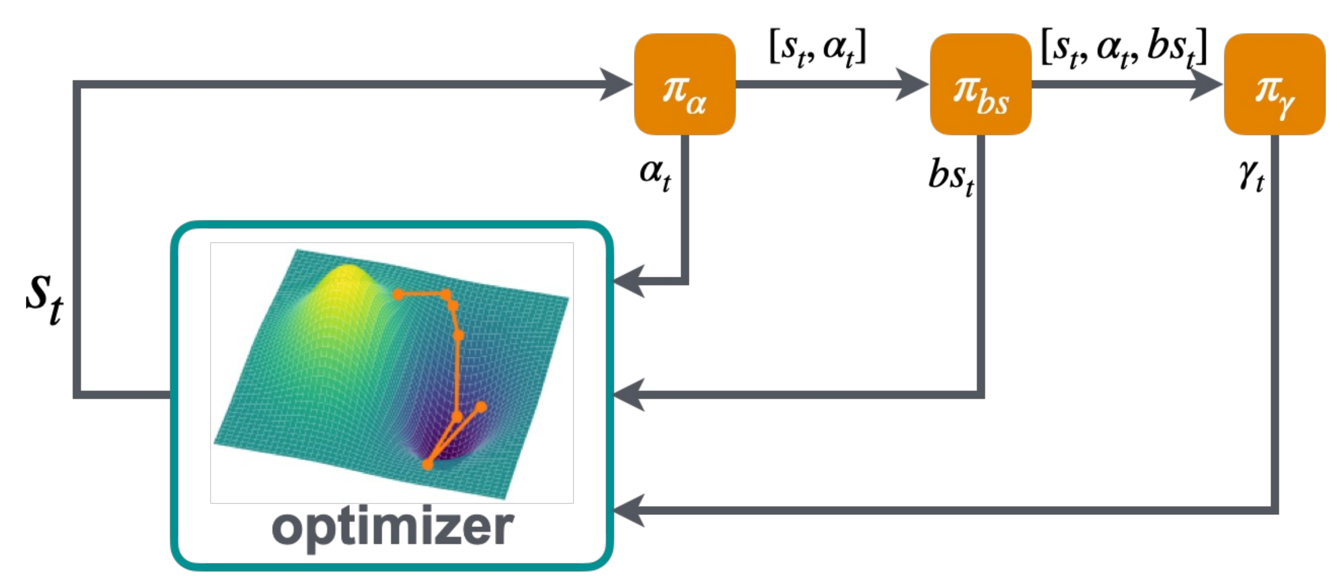
universität freiburg

small data MORE FROM LESS CRC 1597

In a Nutshell

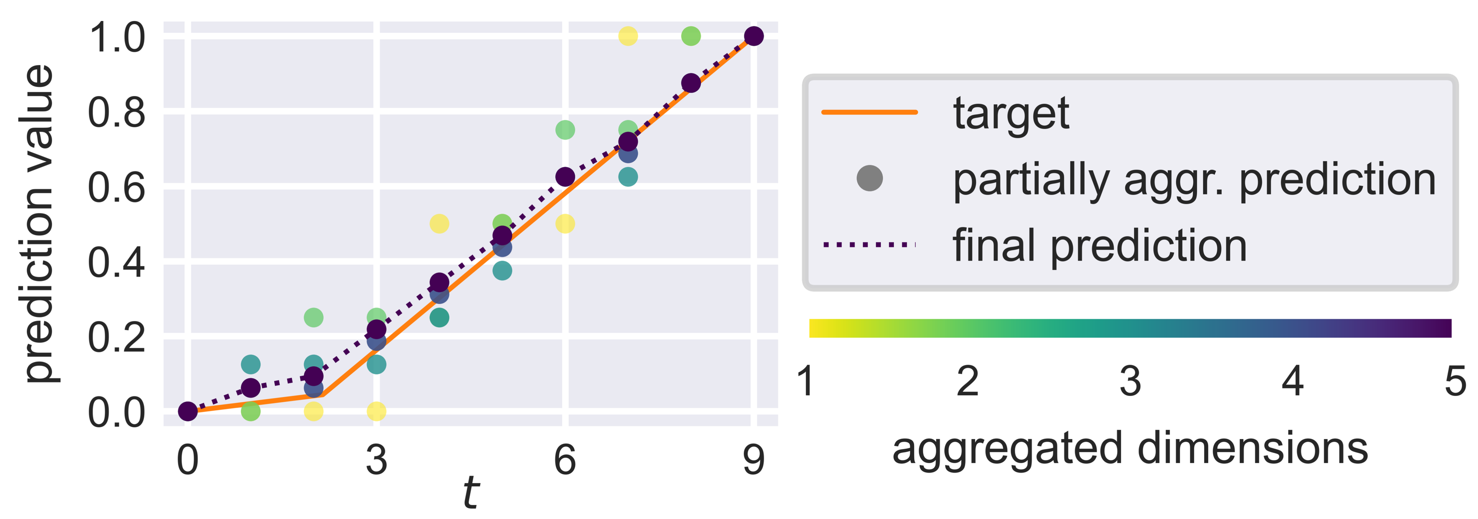
1. We factorize high-dimensional action-spaces **avoiding combinatorial explosion while preserving ability to coordinate**
2. We employ sequential policies to learn a policy **per action dimension (hyperparameter)**
   - Set hyperparameters in order of importance
   - Condition on already set hyperparameters
   - Propose new TD-update for sequential policies
3. We propose **a new toy-benchmark** to evaluate RL-algorithms under the CANDID setting
4. We evaluate DDQN-based **sequential policies on our new benchmark** against single policy and independent factorized policy baselines

Sequential Policies for DAC



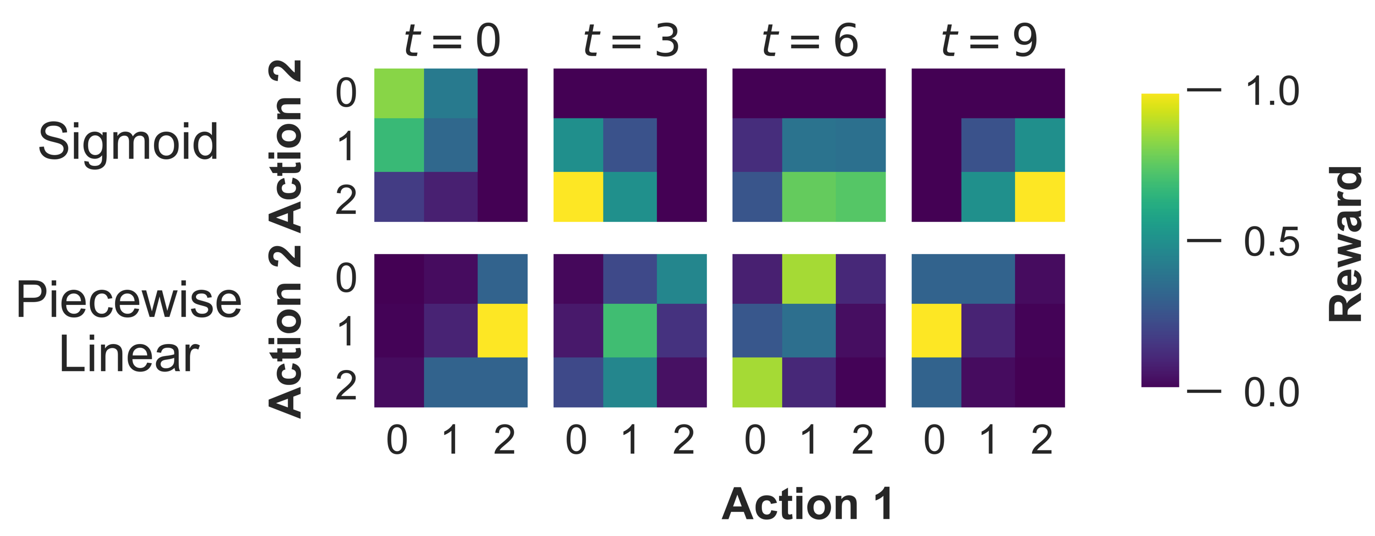
Piecewise Linear Benchmark

Example of Benchmark Instance (5D action space)



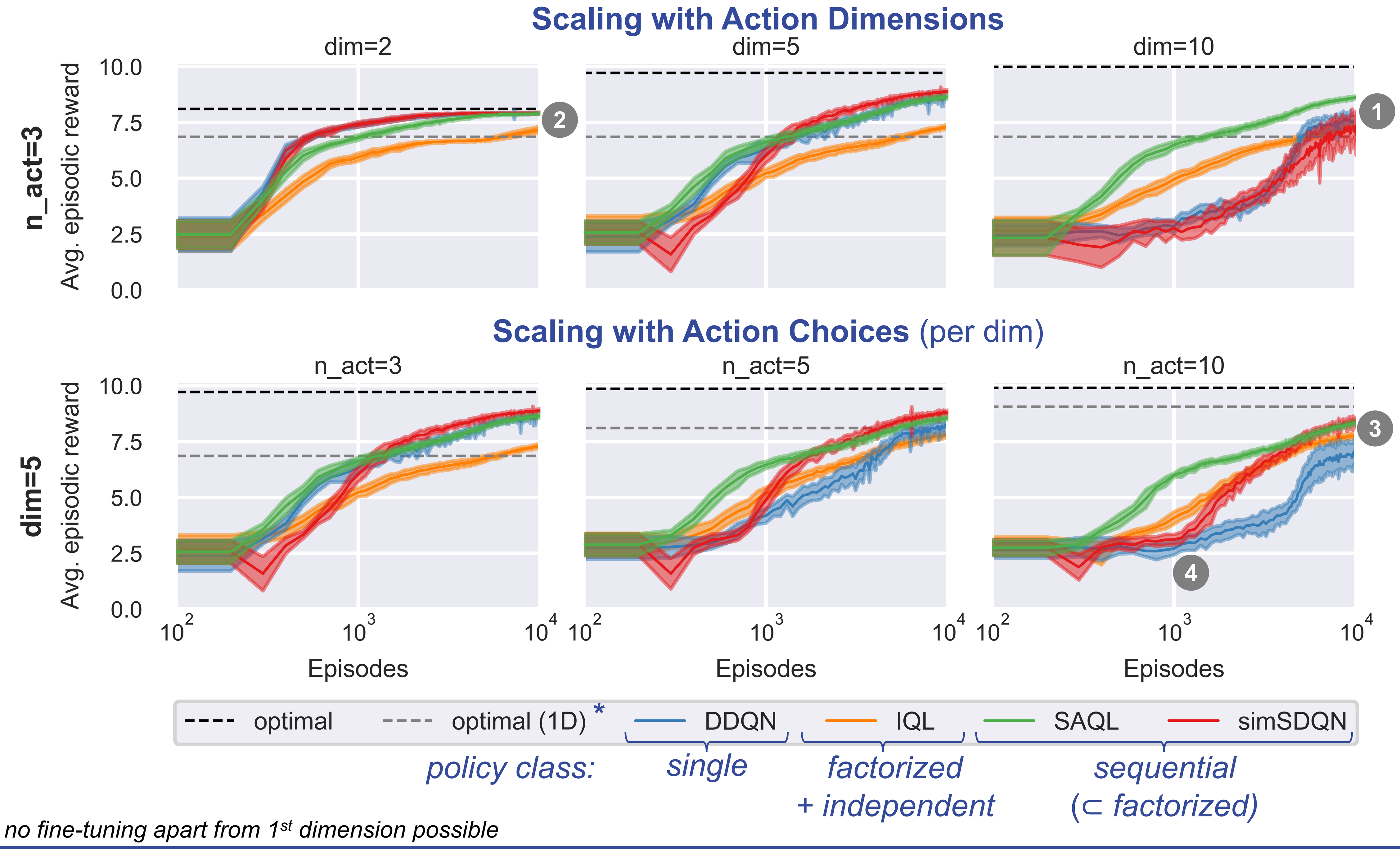
Task: Coordinate action dimensions to progressively fine-tune predictions on target function.

Comparing Rewards per Action Combination (2D action space)



Challenge for factorized policies: Independent optimization of action dimensions not possible in Piecewise Linear benchmark.

Experiments on Piecewise Linear Benchmark



SAQL scales both with dimensionality and no. of action choices (1+3)

simSDQN takes longer to learn (4) and is negatively impacted by dimensionality (1)

IQL fails to coordinate even in simplest case (2)

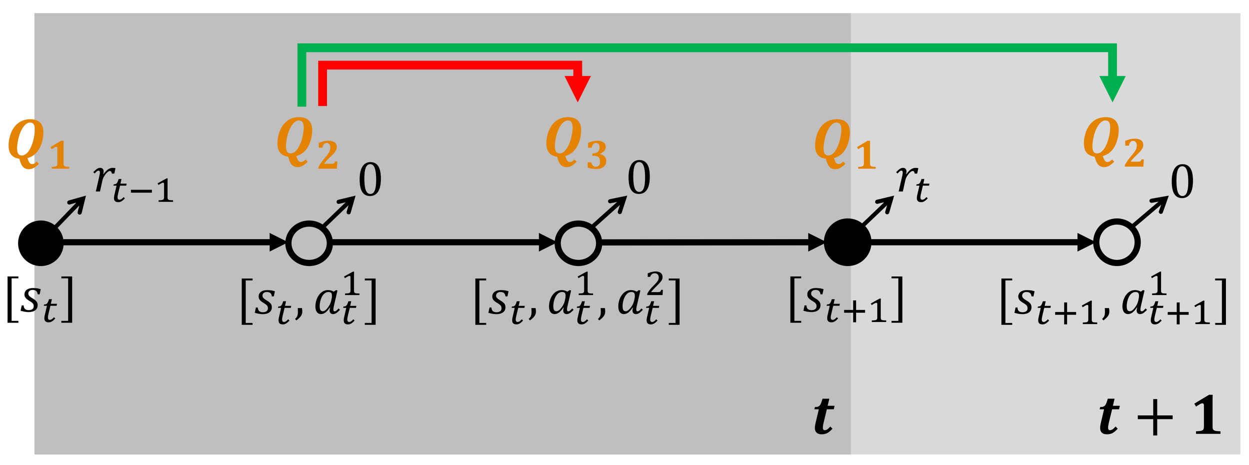
DDQN does not scale with action space size (dimensionality AND no. of choices, 1+3)

→ Sequential policies promising to better solve DAC (**SAQL** seems more scalable)

*: *no fine-tuning apart from 1st dimension possible*

Sequential Policy Variants

Differ by TD-Updates:



simSDQN: solve extended MDP explicitly

SAQL: sequential game

Future Work

- Extended evaluation of sequential policies:
  - In real world settings
  - Compare against MARL baselines (VDN/QMIX)
- Advancement of framework:
  - Advanced communication (e.g., learned message passing)
  - Combine with value function factorization
  - Joint exploration schemes

https://github.com/PhilippBordne/candidDAC.git