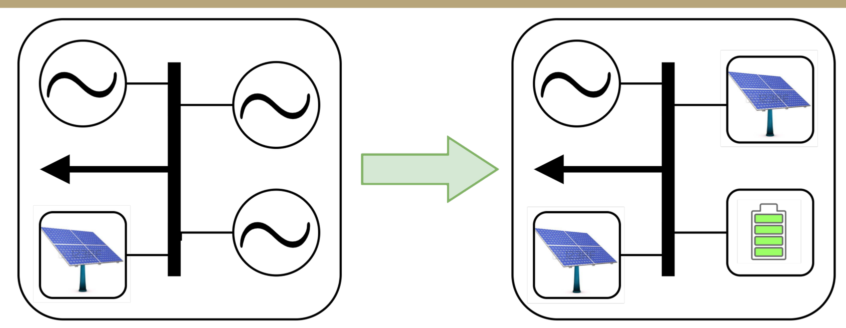
COMPUTATIONALLY EFFICIENT SAFE REINFORCEMENT LEARNING FOR POWER SYSTEMS

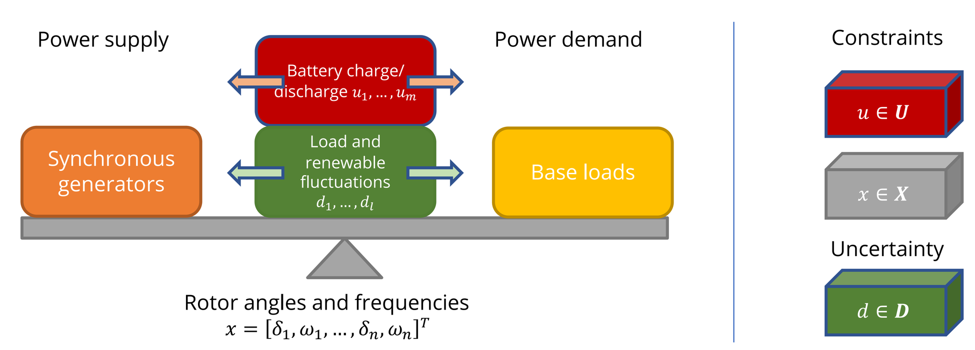
STUDENTS: DANIEL TABAS

Motivation



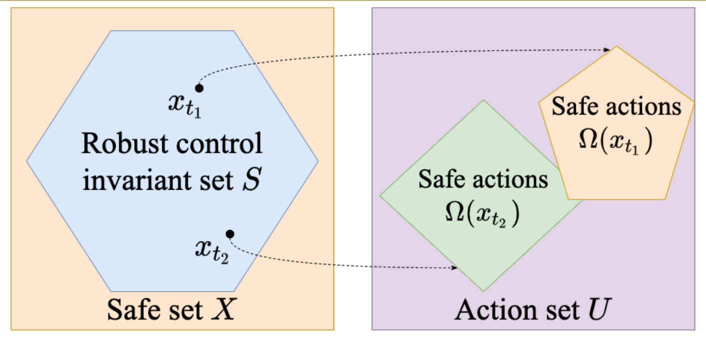
Power systems are transitioning from synchronous generator-based to inverter-based.
- More flexibility, less inherent stability
- Enables and necessitates new control techniques
- Increasing complexity → difficult to derive good policy
- Reinforcement learning (RL) can generate effective policies from data
- However, it is difficult to enforce **safety constraints** on learned polices or during training

Power system stability



Power system stability requires balancing power supply with power demand [1].
- Power imbalance causes frequency deviations
- Frequency deviations can lead to loss of stability or tripping of protective equipment
- Challenges: fluctuating input from renewable energy sources, hard limits on frequency deviations, actuation constraints, nonlinearities, computational power
- Opportunities: flexible energy input from inverter-based resources

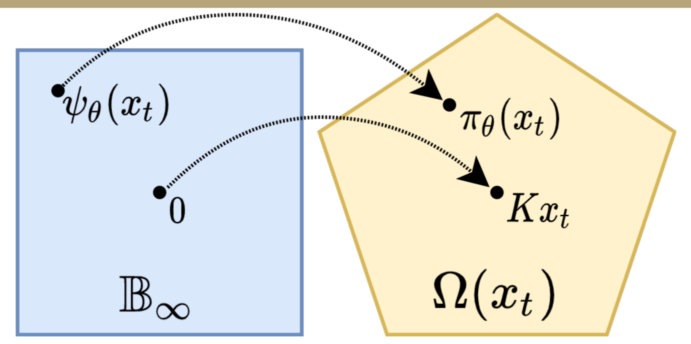
Set-theoretic control



We use set-theoretic control techniques [2] to ensure that a policy trained using deep RL guarantees constraint satisfaction.
- Dynamics: $x_{t+1} = Ax_t + Bu_t + Ed_t$
- Constraints: $x_t \in X, u_t \in U$ (X, U compact polytopes)
- Unknown-but-bounded disturbances/nonlinearities: $d_t \in D$
- Robust control invariant set: $S = \{x \in X: \exists\, u \in U \mid Ax + Bu + Ed \in S, \forall\, d \in D\}$
- Safe action set: $\Omega(x) = \{u \in U: Ax + Bu + Ed \in S, \forall\, d \in D\}$

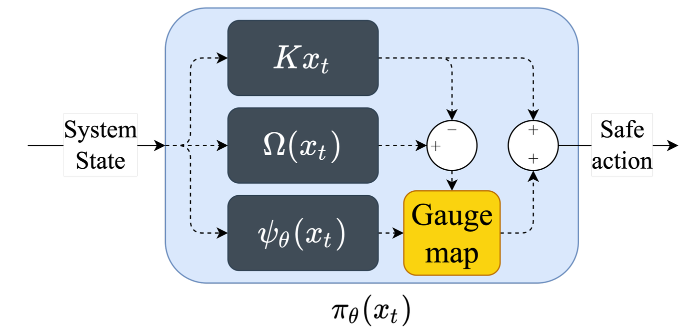
Goal: constrain policy to safe action set without solving projection (too slow)

Safety filter



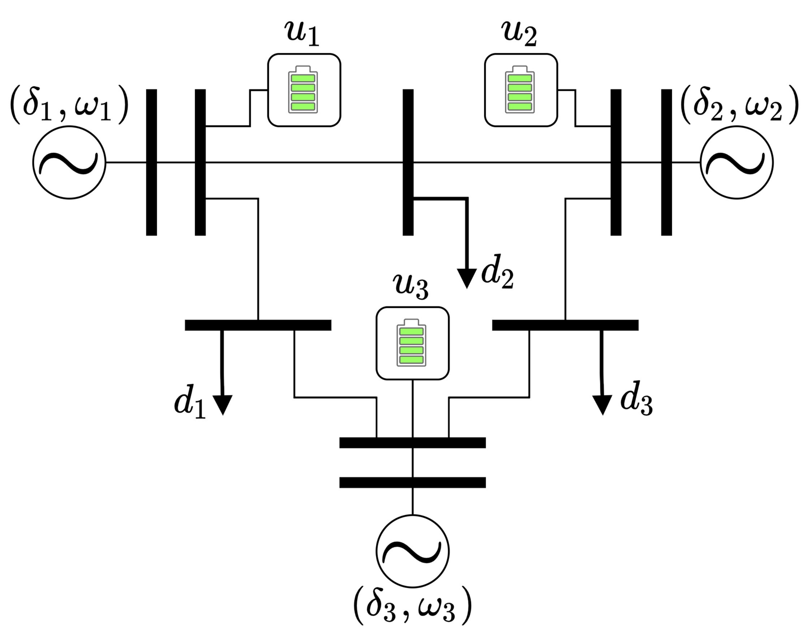
- We establish equivalence of any two convex, compact sets by constructing a closed-form, differentiable bijection between them
- We use this bijection to equate the unit hypercube B_∞ with the safe action set $\Omega(x_t)$
- It is easy to constrain the output of a neural network to B_∞ using tanh activation functions
- We train the neural network using standard policy gradient algorithms

Policy network



The structure of the policy comprises:
- Kx_t, a safe linear feedback used for numerical stability
- $\Omega(x_t)$, the set of safe actions from state x_t
- $\psi_\theta(x_t)$, a feedforward neural network with parameter θ and tanh activation functions in the output layer
- A closed-form, differentiable *gauge map* that takes the outputs of ψ_θ and maps them one-to-one to elements of $\Omega(x_t)$

Test system



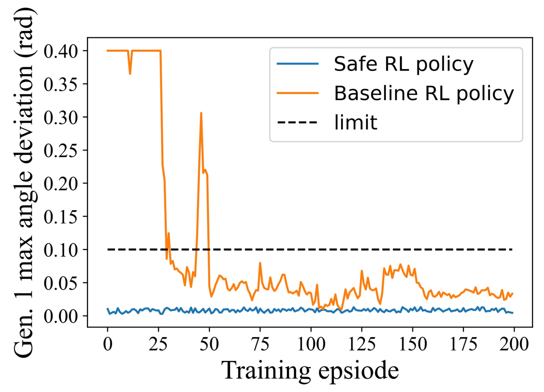
The test system is a 9-bus power system with 3 synchronous generators, 3 uncertain loads, and 3 controllable energy storage systems. The power system dynamics are modeled using the linearized swing equations and DC power flow equations. At each time step, the stage cost is given by

$$l(x_t, u_t) = ||x_t||^2 + ||u_t||^2.$$

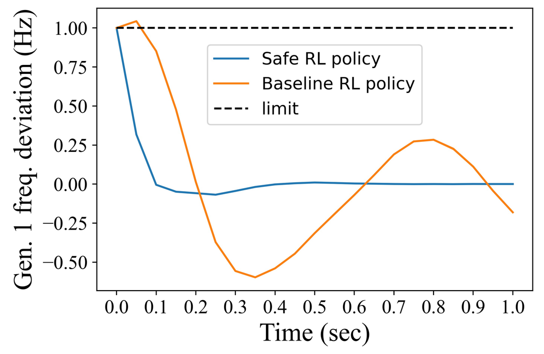
Results

Compared performance of our method to a policy network trained with a soft penalty on constraint violations. Our proposed method demonstrated:

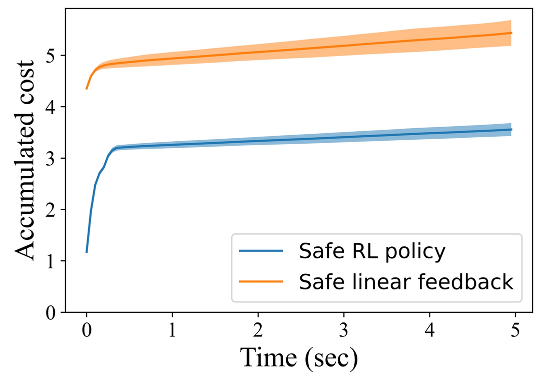
- Better rewards
- No constraint violations throughout training
- Constraint satisfaction during testing



Throughout training, the proposed safe RL policy avoids constraint violations by keeping the rotor angle deviations of all generators below their allowed limits. The baseline policy trained using a constraint violation penalty only learns to satisfy the constraints in the long run.



Even after training, there exist initial conditions for which the baseline RL policy does not always satisfy the constraints of the system.



Since the safe RL policy inherits the safety guarantees from a safe linear feedback controller, we compared the costs accumulated along a trajectory to show that the flexibility of the RL policy enables lower operating costs.

Future Work, References, and Acknowledgments

Future work:
- Investigate robustness of learned policies to topology changes
- Apply the technique to other problems in power systems control and optimization
- Investigate multi-agent setting

References:

[1] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control*, 2nd ed. Wiley, 2008.
[2] F. Blanchini and S. Miani, *Set-theoretic methods in control*. Birkhauser, 2015.

This work is partially supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114 and NSF grant ECCS-1930605. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY *of* WASHINGTON

ADVISORS: BAOSEN ZHANG

SPONSORS: NATIONAL SCIENCE FOUNDATION