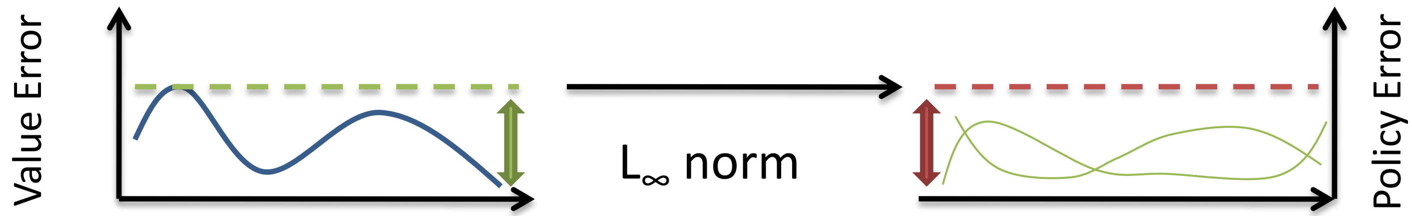
Robust Value Function Approximation Using Bilinear Programming

Marek Petrik, Shlomo Zilberstein

- New formulation of value function approximation for MDPs (ABP)
- Minimizes a bound on policy loss (PL)



LSPI

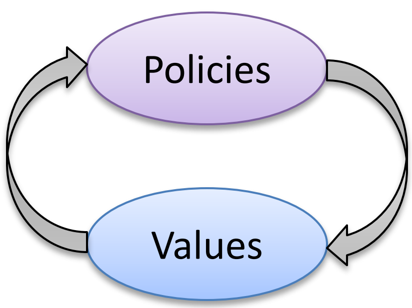
- Iterative optimization
- No convergence
- *Weak* PL guarantees: $\mathbf{L_2}$

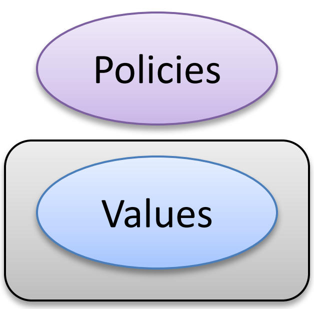
ALP

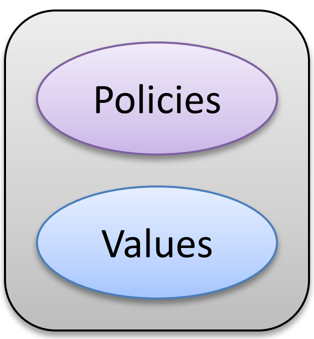
- Ignores policy
- Potentially large error
- *Weak* PL guarantees: $\mathbf{L_1}$

ABP

- Concurrently optimizes value and policy
- *Strong* PL guarantees: $\mathbf{L_\infty}$







- Approximate algorithm for solving ABP ≈ Convergent version of API