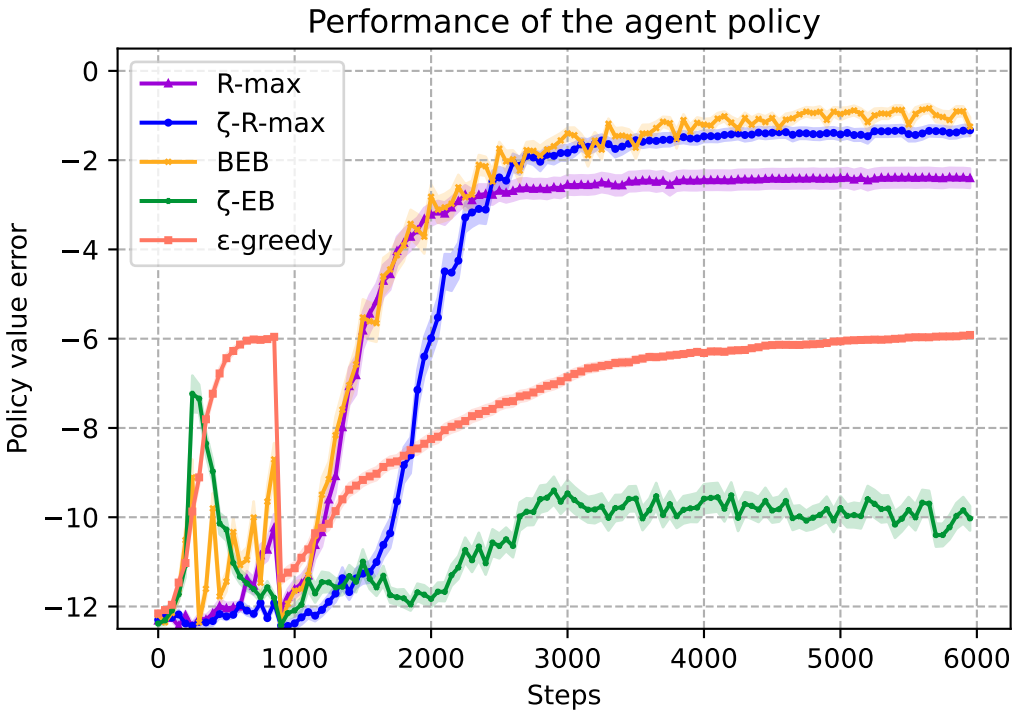
Performance of the agent policy
R-max
ζ-R-max
BEB
ζ-EB
ε-greedy
Policy value error
Steps
0
−2
−4
−6
−8
−10
−12
0
1000
2000
3000
4000
5000
6000