HW2 Policy Gradient Analysis
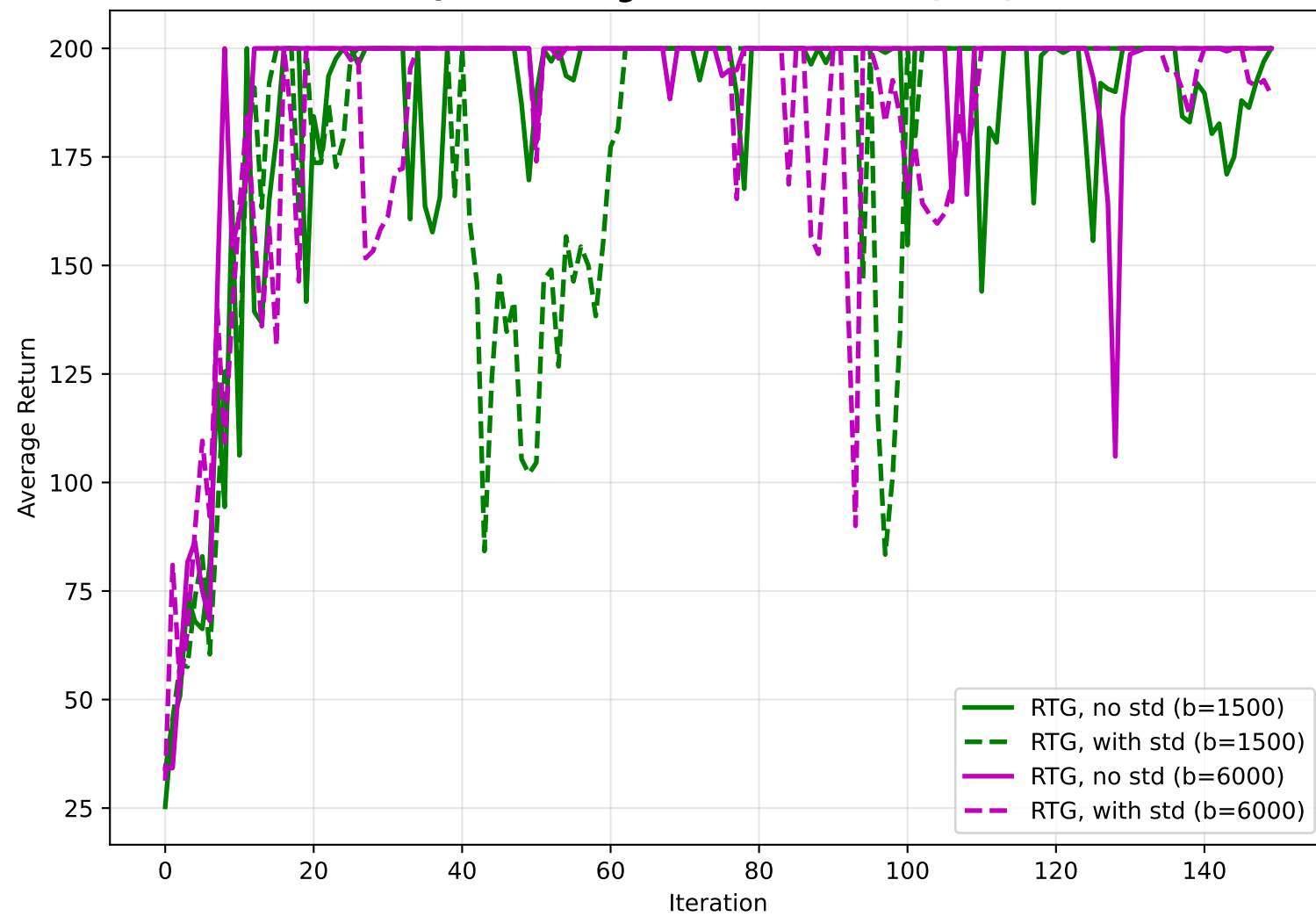

Q1: Value Estimator (Small Batch, b=1500)
Average Return / Iteration
- Trajectory-centric
- Reward-to-go
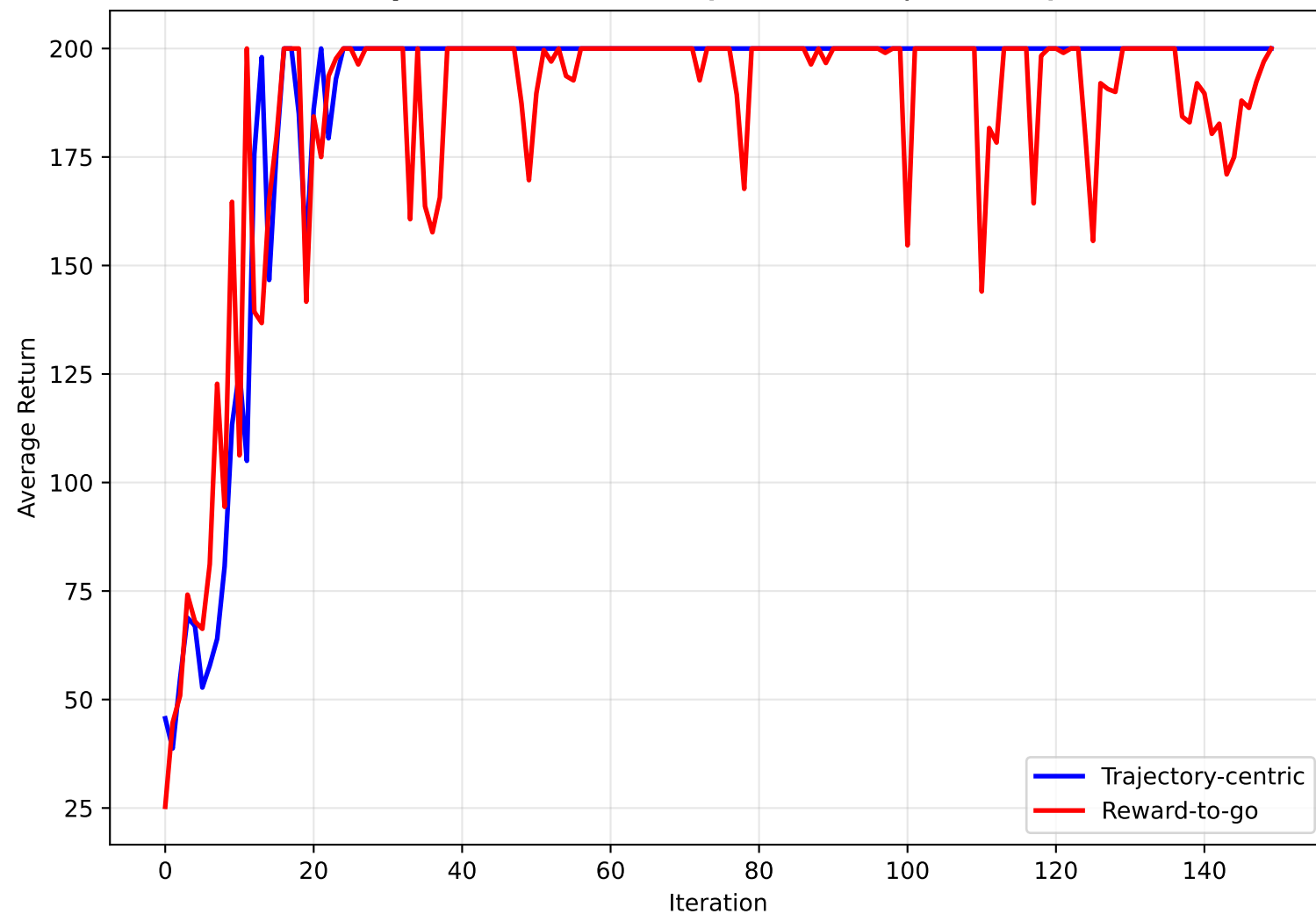
Q1: Value Estimator (Large Batch, b=6000)
Average Return / Iteration
- Trajectory-centric
- Reward-to-go
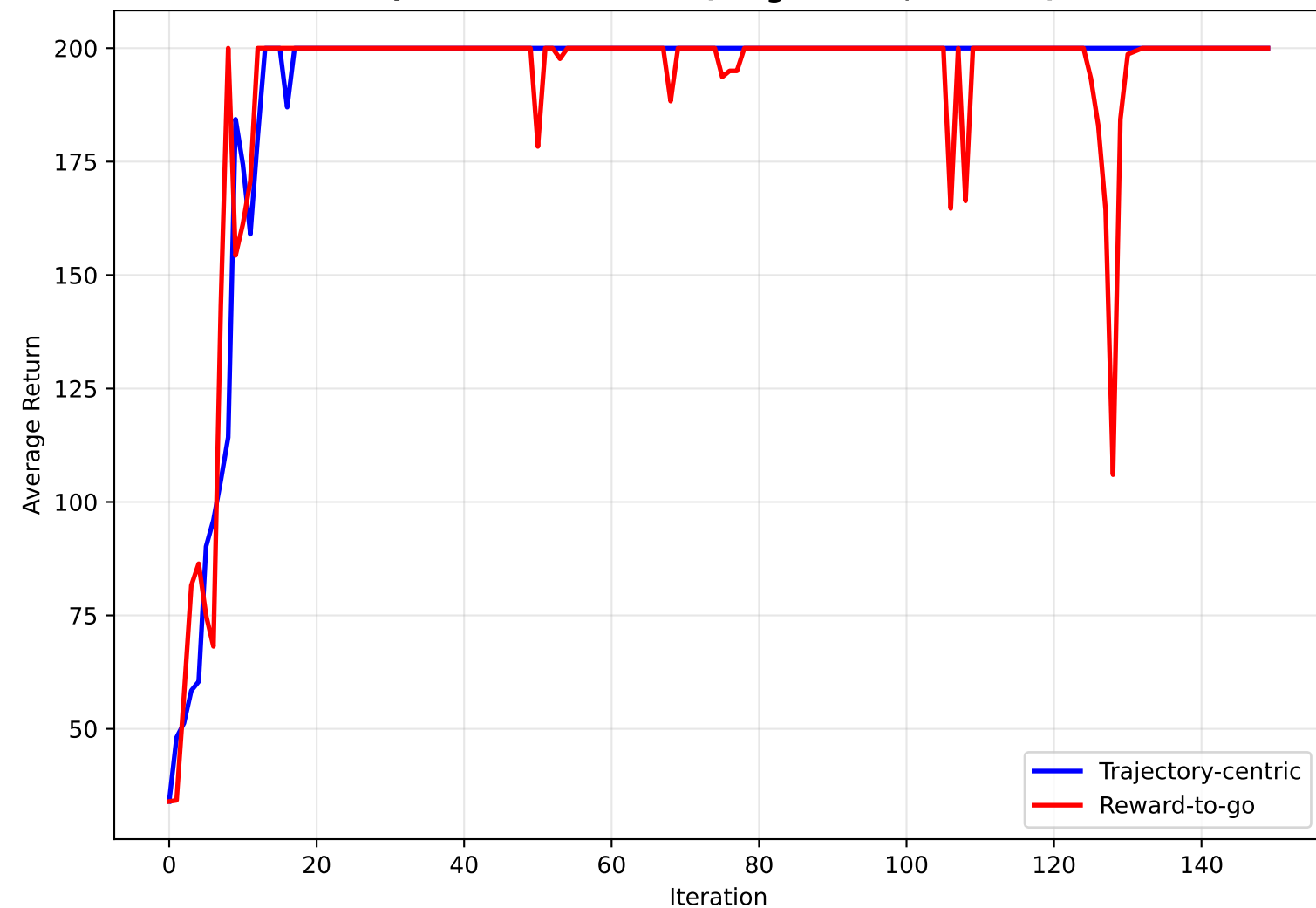
Q2: Advantage Standardization (RTG)
Average Return / Iteration
- RTG, no std (b=1500)
- RTG, with std (b=1500)
- RTG, no std (b=6000)
- RTG, with std (b=6000)

Q3: Batch Size Impact (RTG, no adv std)
Average Return / Iteration
- b=1500
- b=6000