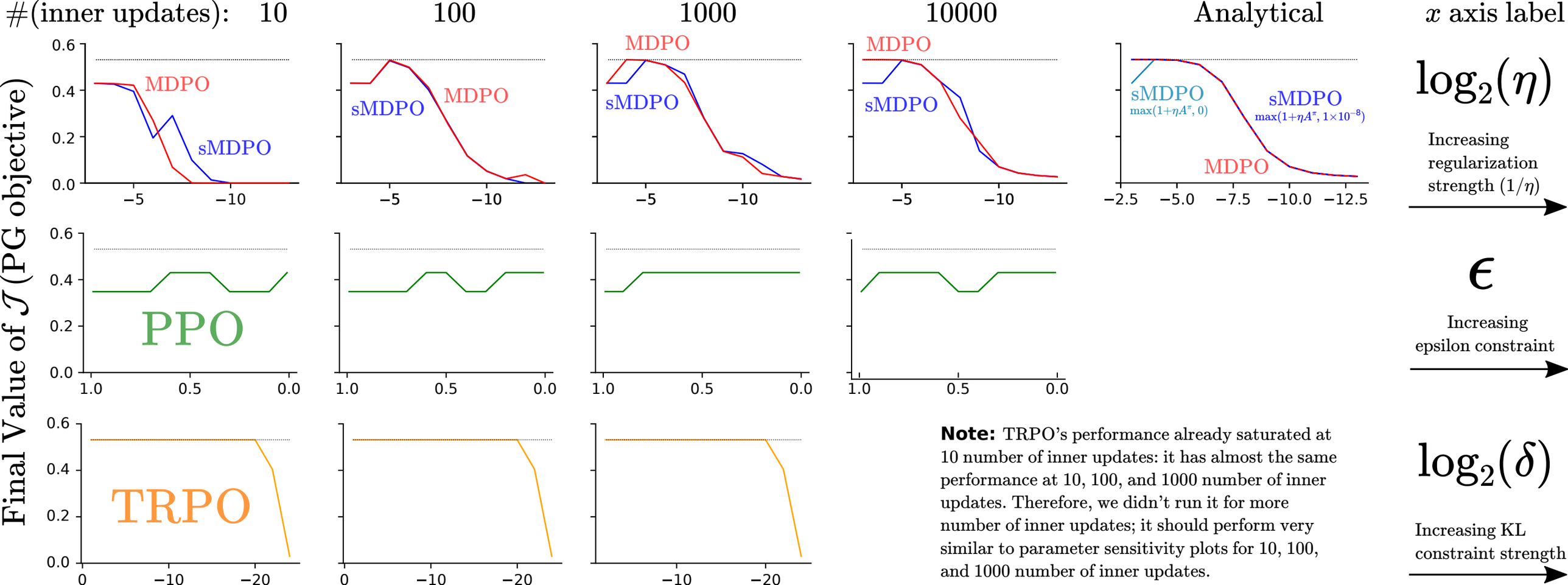
#(inner updates): 10
100
1000
10000
Analytical
x axis label
Final Value of $\mathcal{J}$(PG objective)
MDPO
sMDPO
PPO
TRPO
$\log_2(\eta)$
Increasing regularization strength $(1/\eta)$
ϵ
Increasing epsilon constraint
$\log_2(\delta)$
Increasing KL constraint strength
sMDPO $\max(1+\eta A^\pi, 0)$
sMDPO $\max(1+\eta A^\pi, 1\times10^{-8})$
Note: TRPO's performance already saturated at 10 number of inner updates: it has almost the same performance at 10, 100, and 1000 number of inner updates. Therefore, we didn't run it for more number of inner updates; it should perform very similar to parameter sensitivity plots for 10, 100, and 1000 number of inner updates.