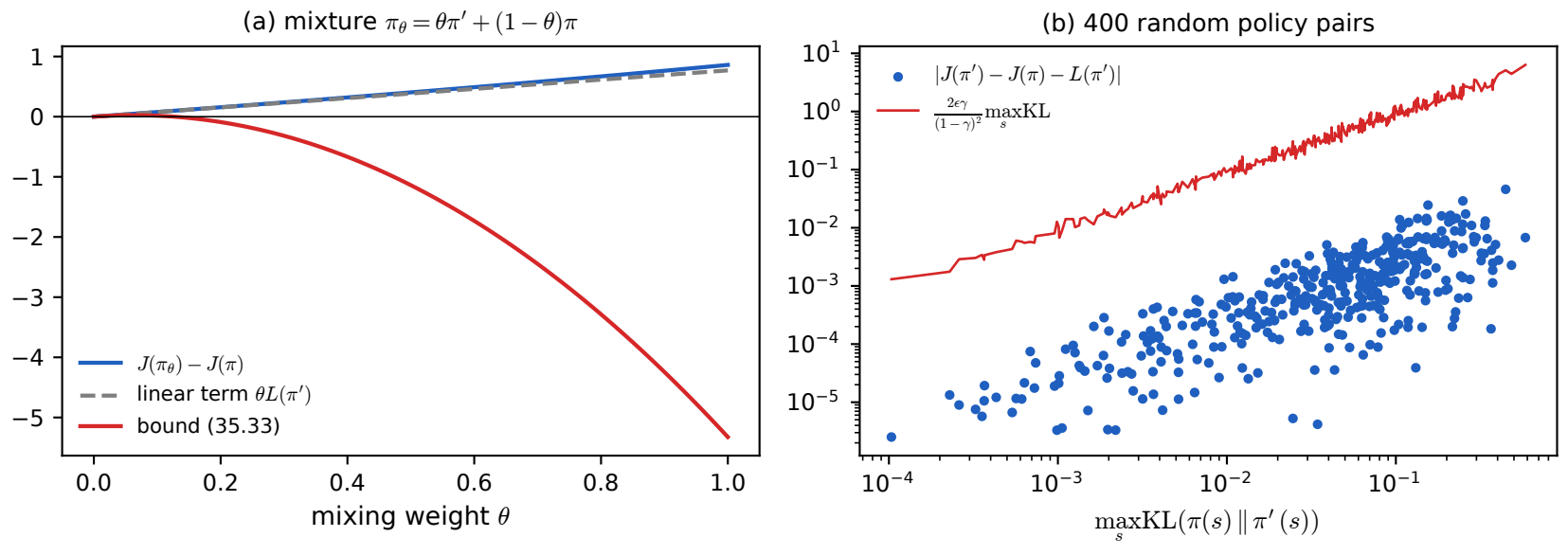

(a) mixture $\pi_\theta = \theta\pi' + (1-\theta)\pi$
1
0
−1
−2
−3
−4
−5
$J(\pi_\theta) - J(\pi)$
linear term $\theta L(\pi')$
bound (35.33)
0.0
0.2
0.4
0.6
0.8
1.0
mixing weight θ
(b) 400 random policy pairs
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
$|J(\pi') - J(\pi) - L(\pi')|$
$\frac{2\epsilon\gamma}{(1-\gamma)^2}\max_s \mathrm{KL}$
10^{-4}
10^{-3}
10^{-2}
10^{-1}
$\max_s \mathrm{KL}(\pi(s) \,\|\, \pi'(s))$