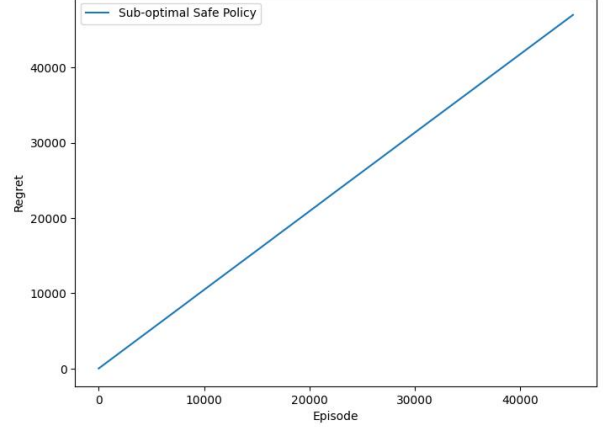
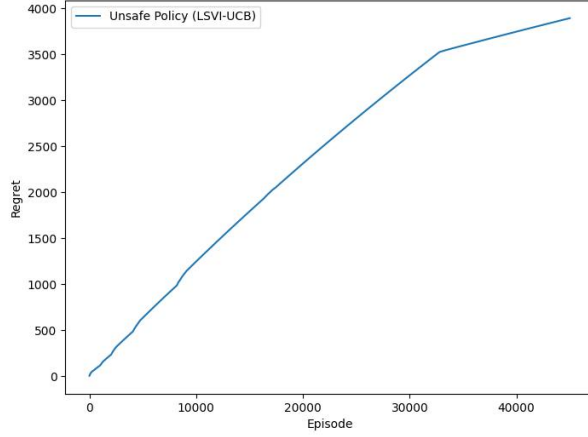


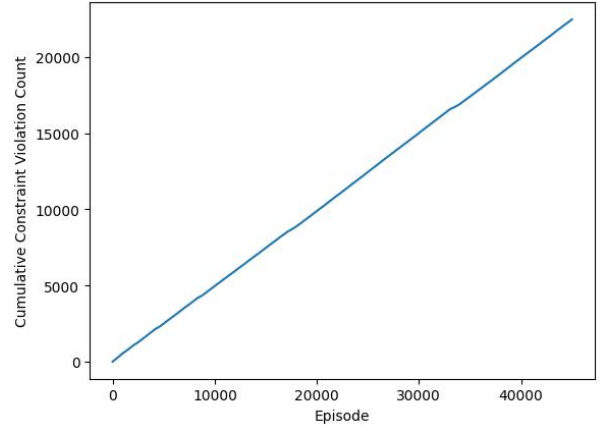
(a) Regret of our method (NCS-LSVI) over more episodes, showing sublinear behavior for  $K' = 2000$ .



(b) Regret of a sub-optimal but safe baseline that stays within an  $\epsilon$ -neighborhood of the initial safe policy.



(c) Regret of LSVI-UCB (Jin et al., 2020), which achieves low regret but violates constraints.



(d) Cumulative constraint violations for LSVI-UCB, showing linear growth. The other two methods have zero violations, so no violation plots are included.