



Figure 1: The training curves of EVO using different sample sizes. In `SafetyPointCircle1-v0`, EVO maintains strong constraint satisfaction and performance even with limited 10 samples. In `SafetyPointGoal1-v0`, constraint satisfaction maintains once the sample size exceeds 20.

055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107
108
109

Environment	CPO	EVO(10)	EVO(20)	EVO(50)	EVO(100)
SafetyPointCircle1	11h 23m 28s	11h 53m 19s (8s)	11h 52m 28s (8s)	11h 53m 31s (8s)	11h 55m 22s (9s)
SafetyPointGoal1	11h 29m 42s	11h 58m 51s (8s)	11h 59m 47s (7s)	11h 58m 33s (8s)	11h 59 m 8s (9s)

Table 1: EVO training time with different sample sizes compared to CPO. The time data includes the total training time for EVO, with the GPD fitting time shown in parentheses "()" . EVO increases limited training time compared to CPO. GPD fitting takes only a few seconds in total.