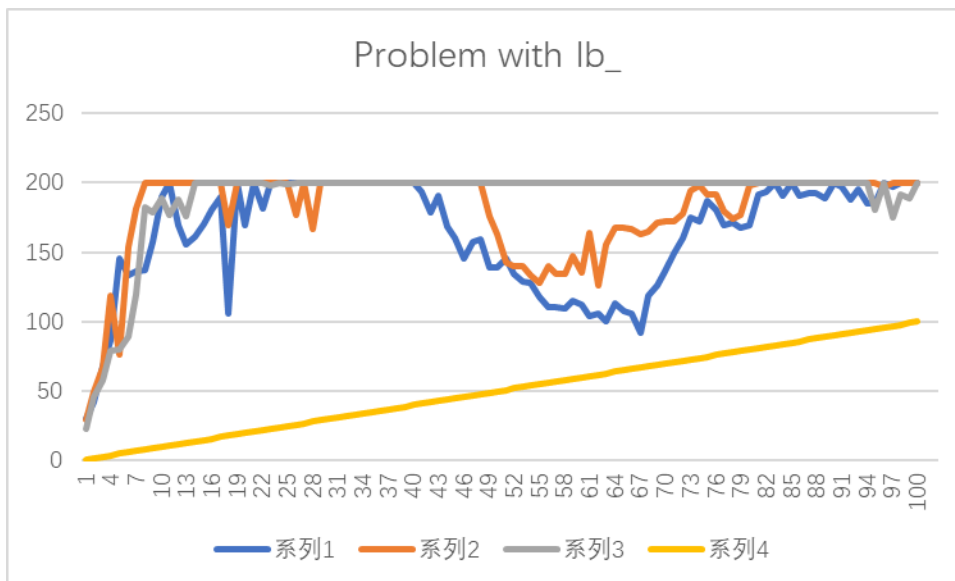
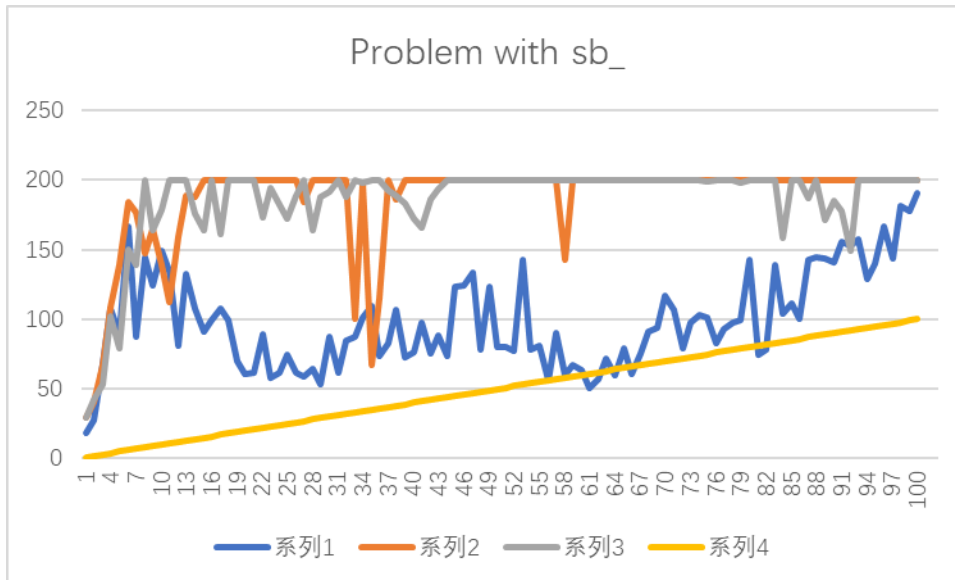


## Homework 2

### 管政 2018211600

#### 1、Problem3 (1)

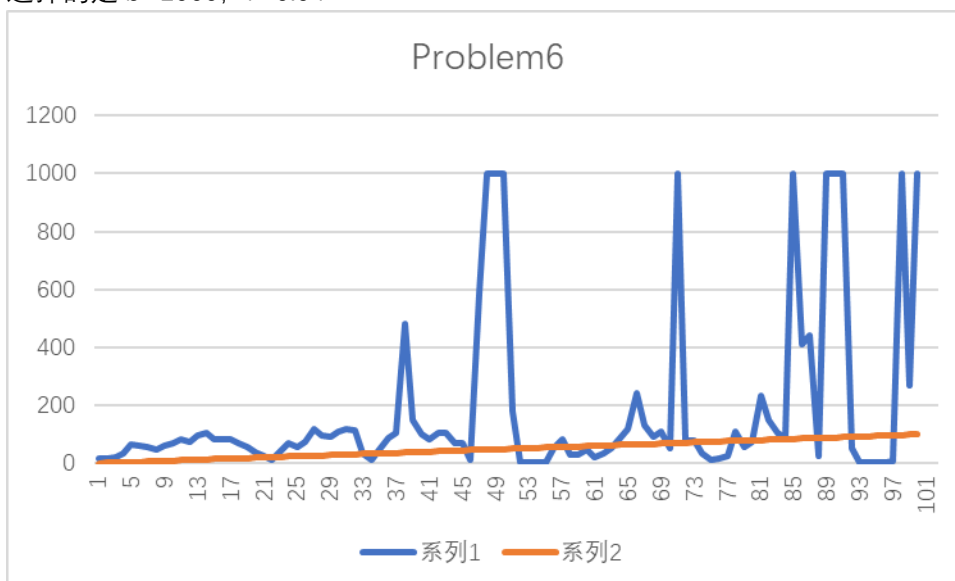


#### (2)

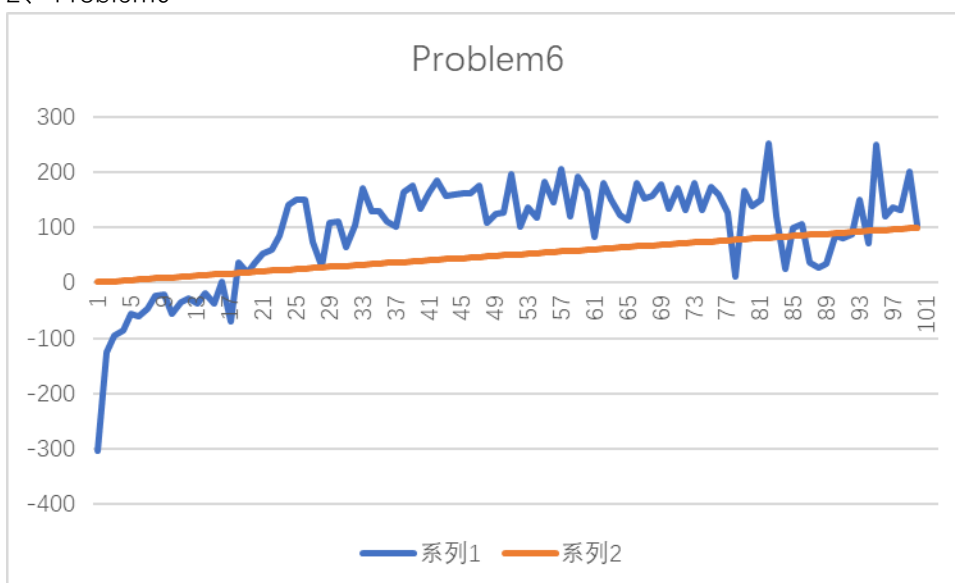
从图上可以看出系列3表现得最好, 系列3是使用reward\_to\_go 和 standardize\_advantages 的策略。并且在 large batch 的时候收敛的比 small batch 时要快。

### 1、Problem4

选择的是  $b=1000$ ,  $r=0.04$

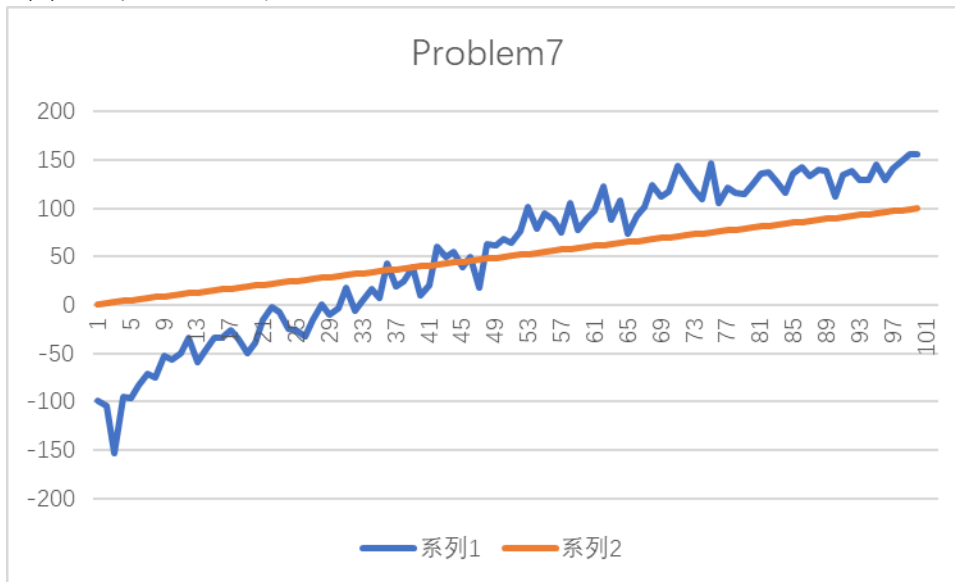


### 2、Problem6



### 3、Problem7

(1) 选择了  $b=5000$ ,  $r=0.01$



(2)

