先固定agent2 policy，optimize agent1 policy。再固定agent1 policy，optimize agent2 policy。

一直重复迭代这个过程。