1. Уважаемые члены комиссии! Добрый день! Меня зовут Динь Нгок Туан. Сегодня я представляю мою работу на тему "Исследование алгоритмов оптимального управления, основанных на обучении с подкреплением". Мой руководитель Перегудин Алексей Алексеевич.
2. Сначала, я сделал обзор литературы, посвящённой обучению с подкреплением. Потом исследовал математическую модель регулятора на основе обучения с подкреплением для линейных и нелинейных систем. Наконец, провёл моделирование и анализ эффективности регуляторов и возможности их применения к физическим системам.
3. Оптимальные регуляторы обычно проектируются в автономном режиме путем решения уравнений Гамильтона Якоби-Беллмана (HJB), с использованием полных знаний о динамике системы. Многие методы решения основываются на решении уравнения HJB, но такие подходы требуют решения уравнений в обратном времени и не могут применяться онлайн.
4. На этом слайде представлена функция ценности в виде уравнения Беллмана. Стратегия, при которой значение функции ценности минимально, является оптимальной. Решение для поиска оптимальных стратегий состоит из двух этапов: Оценка стратегии и улучшение стратегии.
5. Структура Актор-Критик выглядит так. В критике происходит оценка текущей стратегии управления, а в акторе - её улучшение. Для поиска оптимальной стратегии можно использовать один из двух алгоритмов обучения с подкреплением: итерация по стратегии и итерация по критерию.
6. Рассмотрим динамическую систему в таком виде, где A – неизвестная матрица. Закон управления с обратной связью в таком виде [показать рукой на формулу]. Цель – синтезировать регулятор, который обеспечивает устойчивость системы, и также минимизацию квадратичного функционала качества [показать рукой на формулу].
7. *На этом слайде рассматривается алгоритм обучения с интегральным подкреплением для решения задачи оптимального управления. Для решения уравнения Беллмана можно использовать следующие два алгоритма обучения с подкреплением: итерация по критерию и итерация по стратегии. Отсюда мы можем найти оптимальную матрицу P.*
8. Рассмотрим неустойчивый объект, представленный на слайде. Здесь A – неизвестная матрица. Весовые матрицы Q и R выбраны единичными. Решение уравнения Риккати с использованием полных знаний о динамике выглядит так [показать рукой]. Нашему регулятору это решение неизвестно.

(Что касается матрицы Q: Моя задача не изучает выбор Q, R, а решает задачу оптимизации, когда A неизвестен. Конечно же, в реальной задаче следует выбирать Q и R исходя из желаемых переходных процессов)

1. Проведено сравнение трех алгоритмов: LQR, адаптивный регулятор и регулятор на основе обучения с подкреплением. Для регулятора на основе обучения с подкреплением значения коэффициентов регулятора K сходятся к оптимальным за короткое время, в то время как адаптивный регулятор не сходится к оптимальному. Как видно по графику, траектория сходимости и значение функции ценности регулятора на основе обучения с подкреплением и LQR практически совпадают.
2. Проведено сравнение влияния времени сбора данных на качество работы исследуемого регулятора. Мы видим что, чем меньше значение T, тем быстрее матрица регулятора K сходится к оптимальным значениям, и тем ближе поведение системы к оптимальному.
3. На этом слайде представлено применение предложенного алгоритма к линейной модели перевернутого маятника на тележке. Здесь A – неизвестная матрица.
4. Моделирование показало, что регулятор на основе обучения с интегральным подкреплением обеспечивает сходимость коэффициентов регулятора к оптимальному значению и значения функции ценности примерно равно значению в случае LQR.
5. Теперь рассмотрим управление **нелинейной** [выделить голосом] динамической системой с неизвестной функцией f(x). Целью управления вновь будет являться минимизация представленной на слайде функции ценности.
6. Для решения задачи я использовал алгоритм итерации по стратегии и алгоритм приближённого динамического программирования.
7. На этом слайде представлено моделирование регулятора для простой нелинейной системы с неизвестной функцией f(x). Результаты показывают, что веса сходятся к оптимальным значениям.
8. [думаю, этот слайд не нужен]
9. На этом слайде представлено применение предложенного алгоритма к нелинейной модели перевернутого маятника с неизвестными параметрами.
10. На этом слайде представлены результаты моделирования. ​​ На первом этапе работы регулятора в сигнал управления был добавлен шум, чтобы обеспечить наличие неисчезающего возбуждения. Как видно по графику, веса (а значит и коэффициенты регулятора) сходятся к оптимальным. Результаты сравнения также показывают преимущество рассмотренного подхода над использованием регулятора без обучения.
11. Заключение: …
12. Спасибо за внимание