

Задачи стохастического оптимального управления: Оптимизация, обучение и фильтрация

Галяев Андрей Алексеевич
Зав. лаб. 38, ИПУ РАН, Москва, Россия

Ученый Совет ИПУ РАН
Сентябрь, 21, 2026

Ключевые сотрудники ИПУ и их вклад (1950–1970-е)

Сотрудник	Направление	Период деятельности в ИПУ	Ключевой вклад
А.А. Фельдбаум	Оптимизация	1940-е–1969	Основоположник оптимального управления как вариационной задачи; автор теории дуального управления (сочетание обучения и управления)
Б.Н. Наумов	Оптимизация	1950–1967	Методы моделирования и оптимизации; принципы построения АСУ
Б.Т. Поляк	Оптимизация	1971-далее	Стохастическая оптимизация, теория управления
Б.Н. Петров	Оптимизация / управление	1947–далее	Теория беспоисковых самонастраивающихся (адаптивных) систем; теория координатно-параметрического управления; принцип двухканальности
А.М.Петровский, Е.П. Маслов	Оптимизация / фильтрация	1949-далее 1960-далее	Оптимальное управление наблюдениями; интеллектуальный анализ данных
Я.З. Цыпкин	Обучение / управление	1950-далее	Основоположник теории обучающихся и самообучающихся систем; основы теории релейных и импульсных автоматических систем
М.А. Айзерман	Обучение	1950–далее	Метод потенциальных функций в теории распознавания образов (совместно с Браверманом и Розоноэром); интеллектуальный анализ данных
А.Я.Червоненкис, В.Н. Вапник	Обучение	1960-далее	Метод обобщенного портрета; теория Вапника–Червоненкиса
В.С. Пугачев	Фильтрация	1956–далее	Теория условно-оптимальной фильтрации; статистические методы адаптивной фильтрации
Н.А. Кузнецов	Фильтрация / Оптимизация / Обучение	1960-далее	Оптимальная фильтрация; теория «рассинхронизованных» систем

Эволюция направлений теории управления (2011–2026)

Направление	2011–2015	2016–2020	2021–2026
Линейные и нелинейные системы, устойчивость (Ляпунов)	ядро дисциплины	ядро, медленный рост	ядро, незначительный рост
MPC / предиктивное управление	интенсивный рост	устойчивый рост, робастные варианты	зрелый, интеграция с ML
Сетевое и многоагентное управление	взрывной рост	стабилизация, зрелость	«распыление» по смежным секциям
Управление по данным (data-driven)	нишевое	интенсивный рост, «лемма Виллемса»	ведущее, DeePC/Купман/SINDy
Обучение с подкреплением, нейросетевые регуляторы	единичные работы	быстрый рост, Atari/AlphaGo	одно из ведущих, RL for control
Безопасное управление, CBF, верификация	зарождение	интенсивный рост, первые CBF-QP	ведущее, safe learning
Киберфизические системы, безопасность	зарождение	рост, атаки и защита	зрелое направление
Управление роботами и автономными системами	рост	взрывной locomotion/drones	рост, ведущее, VLA/foundation models
Управление в биомедицине, синтетическая биология	нишевое	рост, закрытый контур (ИПЖ)	значимое, синтетическая биология
Физически информированное ML, PINN	отсутствует	зарождение	быстрый рост
Оборона, многодоменное управление (JADC2)	ведомственная тематика	рост сетевых постановок	крупный сетевой драйвер

Дуальное управление

Задача дуального управления – это задача синтеза такого закона управления, который одновременно решает две взаимосвязанные цели: направляет объект к желаемому состоянию и активно собирает информацию о его неизвестных динамических свойствах. Феномен дуального управления был открыт А.А. Фельдбаумом в 1960 году.

Ключевая особенность дуального управления — двойственность управляющих воздействий. С одной стороны, они направлены на достижение основной цели — приведение объекта в требуемое состояние (**направляющая функция**). С другой стороны, эти же воздействия служат для активного изучения объекта: они специально формируются так, чтобы наилучшим образом выявить его характеристики и параметры модели, что в итоге повышает качество управления в будущем (**изучающая функция**).

Аналитические методы (СтОУ):

- 1 Динамическое программирование (поиск оптимальной стратегии по решению уравнения Беллмана, функция потерь одновременно учитывает и отклонение от цели, и ценность получаемой информации).
- 2 Теория статистических решений и байесовский подход (неопределённость параметров объекта как априорное распределение вероятностей).
- 3 Принцип максимума Понтрягина (поиск оптимальных траекторий при наличии ограничений и неопределённости).

Методы идентификации и адаптации:

- 1 Методы параметрической идентификации (МНК, рекуррентные алгоритмы).
- 2 Непараметрические методы идентификации (ядерные оценки, на основе рядов Вольтерра, локально-аппроксимационные подходы).
- 3 Адаптивное управление с идентификатором.
- 4 MPC (Model Predictive Control) с адаптацией.
- 5 Обучение с подкреплением (RL) (баланс между исследованием (exploration) и применением (exploitation)).
- 6 Фильтры Калмана и их расширения (ЕКЕ, UKF).

Группировка по информационной структуре

Группа	Доступная информация	Управление
Полная информация	Состояние X_t известно точно в каждый момент.	Обратная связь по состоянию (feedback control).
Неполная информация	Доступны только шумные измерения $Y_t = h(X_t, v_t)$.	Управление на основе фильтрации (разделение или совместное управление).
Неизвестная модель	Параметры модели неизвестны и оцениваются в процессе.	Дуальное управление или адаптивное управление.
Распределенная информация	Несколько агентов с частичной информацией.	Кооперативное или децентрализованное управление.

Систематизация математических постановок задач СтОУ

1. Классификация по математическому аппарату

Группа	Математический аппарат	Типичные постановки	Примеры задач
А. Динамическое программирование	Уравнения в частных производных (PDE) первого и второго порядка. Принцип оптимальности Беллмана.	Задачи с непрерывным временем, полной наблюдаемостью состояния, аддитивным шумом.	Задача Мертона, LQR, управление запасами, управление процессом генерации в диффузионных моделях.
В. Стохастическое управление с PDE	Уравнения Колмогорова-Фоккера-Планка. Сопряженные системы.	Управление ансамблями частиц, задачи с распределенными параметрами, управление популяциями.	Управление роем беспилотников, макроэкономическое моделирование, управление дорожным трафиком.
С. LQR / LQG	Уравнения Риккати (дифференциальные и алгебраические). Фильтр Калмана.	Линейная динамика, квадратичная стоимость, гауссовский шум.	Слежение за целью, наведение ЛА, управление мощностью в сетях связи.
Д. Марковские процессы принятия решений (MDP)	Уравнения Беллмана для полумарковских процессов. Процессы рождения-гибели.	Дискретные состояния и/или действия, марковское свойство, пуассоновские потоки событий.	Управление очередями, управление зарядкой аккумуляторов, адаптивное управление светофорами.
Е. Дуальное управление	Байесовское обновление распределения параметров. Уравнение Беллмана с расширенным состоянием.	Неопределенные параметры модели, управление выполняет двойную функцию.	Активное обучение, баланс «исследования-эксплуатации», управление при неполных данных.
Е. Распределенно-робастное управление (DRSC)	Теория мер, оптимальный транспорт, минимизация на множестве распределений.	Неизвестное точное распределение шума, но известен класс допустимых распределений.	Управление запасами, портфельное инвестирование, управление цепочками поставок.
Г. Задача Шрёдингера	Энтропийно-регуляризованное оптимальное управление, теория больших отклонений.	Связь двух распределений за конечное время.	Обучение диффузионных моделей.

2. Классификация по структуре неопределенности

Группа	Тип неопределенности	Характер управления	Примеры задач
I. Аддитивный гауссовский шум	Шум в динамике, независимый от состояния и управления.	Управление по принципу разделения (оценка и управление разделены).	LQG, слежение, управление дозированием.
II. Мультипликативный шум	Шум пропорционален состоянию.	Управление зависит от текущего состояния, нет разделения.	Задача Мертона, управление ветроэнергетикой.
III. Неизвестные параметры модели	Неопределенность в параметрах f, Σ, r .	Управление двойное (исследование + эксплуатация) или робастное.	Дуальное управление, робастное управление при неполных данных.
IV. Неизвестное распределение шума	Неопределенность типа распределения (не только параметры).	Распределенно-робастное управление.	Управление запасами, портфельное инвестирование.
V. Стохастическая среда с дискретными событиями	Случайные моменты времени (пуассоновские потоки), скачкообразные изменения.	Управление на основе текущего состояния (реактивное).	Управление очередями, управление светофорами.

3. Группировка по типу целевого функционала

Группа	Целевой функционал	Примеры задач
1. Минимизация дисперсии / момен-та второго порядка	$\mathbb{E}[\ X_T - X^*\ ^2]$ или интегральная квадратичная ошибка.	LQR, слежение, квадратичное хеджирование.
2. Максимизация ожидаемой полезности	$\mathbb{E}[U(X_T)]$ или $\int U(c_t)dt$ (CRRA, CARA).	Задача Мертона, управление популяцией.
3. Минимизация риска с учетом ограничений	Риск-чувствительный критерий (например, $\mathbb{E}[e^{\theta X_T}]$) или задача с ограничениями на вероятность невыхода за границы.	Управление кредитным риском, управление вентиляцией.
4. Минимизация суммы затрат + регуляризация	Квадратичная стоимость управления и терминальный член.	LQR, управление мощностью.
5. Энтропийная регуляризация	Минимизация дивергенции Кульбака–Лейблера между траекториями.	Генеративный ИИ, диффузионные модели.

Систематизация математических постановок задач СтОУ

4. Сравнительная таблица математических постановок

Постановка	Уравнение	Решатель	Применимость
Уравнение Гамильтона–Якоби–Беллмана (HJB)	$\rho V = \max_a \{r + \mathcal{L}^a V\}$	Численное решение PDE (нейронные сети, сеточные методы), траекторный интеграл	Нелинейные задачи, непрерывное состояние и действие
Уравнение Риккати (LQR/LQG)	$PA + A^\top P - PBR^{-1}B^\top P + Q = 0$	Аналитическое решение	Линейная динамика, квадратичная стоимость
Уравнение Колмогорова–Фоккера–Планка	$\partial_t p = -\nabla \cdot (fp) + \frac{1}{2} \nabla^2 (\Sigma \Sigma^\top p)$	Методы частиц, конечно-разностные схемы	Задачи с ансамблями частиц
Дуальное управление (адаптивное)	Расширенное HJB с распределением параметров	Монте-Карло, нейронные сети	Неизвестные параметры
Распределенно-робастное (DRSC)	$\inf_{\mathbb{P} \in \mathcal{U}} \sup_a \mathbb{E}[J]$	Полуопределенное программирование	Неизвестное распределение шума
Задача Шрёдингера	$\min_u \int \frac{1}{2} \ u\ ^2 dt + KL(p_T \ p_{\text{data}})$	Стохастическое управление (SDE)	Генеративный ИИ

5. Итоговая трехмерная систематизация

❶ Математический аппарат (A–G)

❷ Структура неопределенности (I–V)

❸ Целевой функционал (1–5)

Каждая конкретная задача — это точка в этом пространстве. Например:

- LQR: $[A, I, 1]$, Задача Мертона: $[A, II, 2]$
- Безопасное обучение (CBF, нейросетевые функции Ляпунова): $[A, III, 3]$
- Физически информированное обучение (PINN, SINDy): $[B, III, 1]$
- Скрытые атаки в киберфизических системах: $[A/C/E, III/IV/V, 3/4]$
- Управление светофорами: $[D, V, 4]$
- Обучение с подкреплением (RL) как приближенное DP: $[E, III, 1/5]$
- Обучаемая робототехника (sim-to-real, VLA/LLM): $[E/A, IV, 5/3]$
- Дуальное управление с активным обучением: $[E, III, 3]$
- Управление по данным (DeePC, Купман): $[F, IV, 4]$
- Диффузионная модель (Мост Шредингера): $[G, I, 5]$

Систематизация математических постановок задач СтОУ

Управление по данным (DeePC, поведенческий подход) [F, IV, 4]

Ось	Обоснование
Математический аппарат (F)	DeePC (Data-enabled Predictive Control) строится на фундаментальной лемме Виллемса и не требует явной модели. Это распределённо-робастный подход, где мы работаем с множеством траекторий, а не с одной моделью.
Структура неопределённости (IV)	Мы не знаем точной модели системы, но у нас есть данные. Это классическая <i>неизвестная модель / неизвестное распределение</i> , которую мы заменяем ансамблем траекторий.
Целевой функционал (4)	DeePC использует квадратичную регуляризацию, чтобы сделать решение задачи устойчивым.

Обучение с подкреплением (RL) как приближенное DP [E, III, 1/5]

Ось	Обоснование
Математический аппарат (E)	Глубокое RL — это, по сути, <i>дуальное управление</i> в самом широком смысле. Агент не просто управляет, но и активно «исследует» (зондирует), обновляя политику. Оно решает HJB приближённо, но делает это через обучение, а не аналитику (A).
Структура неопределённости (III)	Модель среды <i>неизвестна</i> ; агент учится на опыте (траекториях).
Целевой функционал (1 или 5)	В базовом RL (например, DDPG) мы минимизируем ожидаемую сумму наград (стохастическая оптимизация — тип 1). В расширенных версиях, где поощряется исследование (например, через максимизацию энтропии), цель смещается к энтропийной регуляризации (тип 5).

Систематизация математических постановок задач СтОУ

Безопасное обучение (CBF, нейросетевые функции Ляпунова) [А, III, 3]

Ось	Обоснование
Математический аппарат (А)	Управляющие барьерные функции (CBF) — это прямой аналог функций Ляпунова, которые являются частью классического аппарата НЖВ. Их синтез (особенно через нейросети) — это задача <i>оптимального управления с ограничениями</i> .
Структура неопределённости (III)	Модель объекта или внешней среды может быть <i>неизвестна</i> (обучаемый регулятор).
Целевой функционал (3)	Основная цель — <i>безопасность</i> , то есть гарантия того, что состояние системы не покинет заданное безопасное множество. Это классическая <i>задача минимизации риска</i> (вероятностные или детерминированные гарантии).

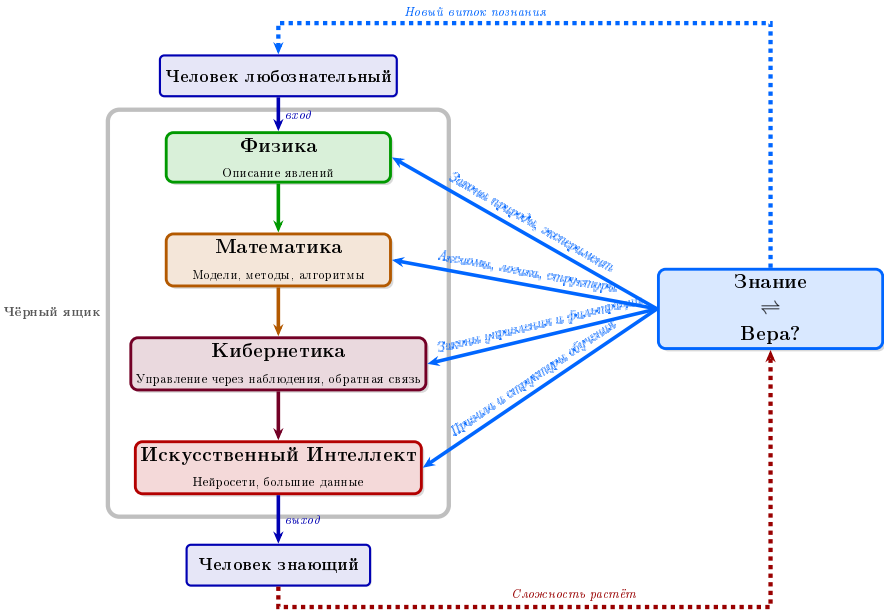
Обучаемая робототехника (sim-to-real, VLA/LLM) [Е/А, IV, 5/3]

Ось	Обоснование
Математический аппарат (Е или А)	Sim-to-real — задача переноса политики, обученной в одной среде (симулятор), на другую (реальный мир). Это требует <i>дуального подхода</i> (Е). Для целостного управления гуманоидами (whole-body control) часто используется аппарат НЖВ (А) для решения задачи оптимального управления сложными многосвязными системами.
Структура неопределённости (IV)	Мы не знаем точного распределения динамики реального мира, только симулятор. Это <i>неизвестное распределение</i> .
Целевой функционал (5 или 3)	Чтобы политика была устойчива к вариативности реального мира, используется регуляризация, часто основанная на <i>энтропии</i> (5). Если политика должна гарантировать безопасность касания или баланса, цель становится <i>риск-чувствительной</i> (3).

Управление движением и навигация

7. Трёхмерная систематизация

№	Математический аппарат	Неопределённость	Целевой функционал	Объединённые задачи
1	A (HJB)	I (аддитивный шум)	1 (дисперсия)	Слежение, следование по пути, стабилизация, SMPC-слежение
2	A (HJB)	I (аддитивный шум)	1 (время)	Навигация в поле дрейфа, управление антенной
3	A (HJB)	I (аддитивный шум)	3 (ошибка, риск)	Наведение ракеты, управление с ограничениями
4	A (HJB)	III (неизв. параметры)	1 (дисперсия)	Сближение с целью, сопровождение
5	A (HJB)	III (неизв. параметры)	3 (риск)	PP/PT (МПО)
6	B (ФПК/ фильтрация)	III (неизв. параметры)	1 (дисперсия)	Пассивная навигация, SLAM
7	B (ФПК)	III (неизв. параметры)	3 (риск)	Мягкая посадка, ковариационное наведение
8	B (ФПК), D (MDP)	IV (неизв. распределение)	1 (дисперсия)	Планирование пути, уклонение от столкновений
9	C (LQR/LQG)	I (аддитивный шум)	1 (дисперсия)	LQR-стабилизация, линейное слежение
10	E (Дуальное управление)	III (Неизвестные параметры модели)	5 (Энтропийная регуляризация)	Поиск пути в неизвестной области



Чем точнее модели → тем сложнее системы → чем сложнее системы → тем больше приходится доверять