



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(национальный исследовательский университет)»

Институт (Филиал) _____ №1 Авиационная техника _____ Кафедра _____ 106

Группа М10-204М-22 Направление подготовки 24.04.03 - Баллистика и гидроаэродинамика

Магистерская программа Динамика полёта и управление движением летательных аппаратов

Квалификация _____ магистр

**РАЗДАТОЧНЫЙ МАТЕРИАЛ
К
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЕ МАГИСТРА
(МАГИСТЕРСКОЙ ДИССЕРТАЦИИ)**

На тему: Использование методов обучения с подкреплением в задаче продольного управления ЛА

Автор диссертации _____ **Москвитин Андрей Семенович** _____ (_____)
(фамилия, имя, отчество полностью)
Научный руководитель _____ **Тюменцев Юрий Владимирович** _____ (_____)
(фамилия, имя, отчество полностью)
Рецензент _____ **Каганов Юрий Тихонович** _____ (_____)
(фамилия, имя, отчество полностью)

К защите допустить

Зав. кафедрой 106 _____ **Ефремов Александр Викторович** _____ (_____)
(№ каф) (фамилия, имя, отчество полностью)
30.05 _____ 2024 г.

Использование методов обучения с подкреплением в задаче продольного управления ЛА

Выполнил студент группы М1О-203М-22:

Москвитин Андрей Семенович

Научный руководитель:

Тюменцев Юрий Владимирович

1. Для эффективного управления ЛА требуются адаптивные контроллеры, один из подходов в его формировании это методы обучения с подкреплением.
2. Управление сложным нелинейными объектом с помощью линейного контроллера является не оптимальным.

1. Разработка подхода к решению задачи управления ЛА с помощью методов обучения с подкреплением (ОсП).
2. Провести анализ качества управления, полученного с помощью метода ОсП, а также сопоставить его с традиционным PI-контроллером

Система уравнений движения в изолированном продольном канале:

$$\dot{x} = \begin{cases} \dot{V}_y = V \cos \alpha, \\ \dot{V}_x = V \sin \alpha, \\ \dot{H} = V_x \sin \vartheta + V_y \cos \vartheta, \\ \dot{\vartheta} = \omega_z, \\ \dot{V} = g [n_{xa} - \sin (\vartheta - \alpha)], \\ \dot{\alpha} = -\frac{g}{V} [n_{ya} - \cos (\vartheta - \alpha)] + \omega_z, \\ \dot{\omega}_z = \frac{M_{Rza}}{J_z}. \end{cases}$$

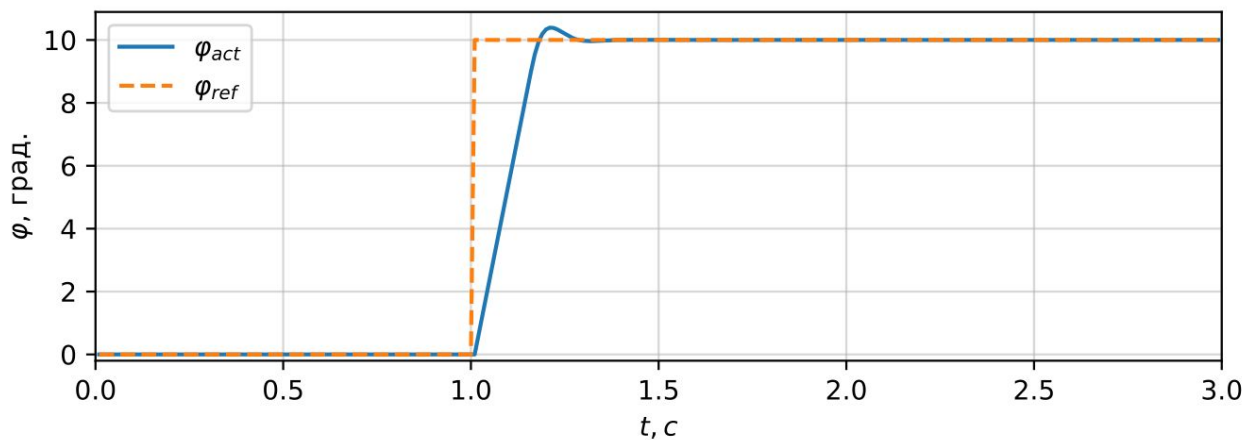
Нелинейность в модели определяется коэффициентами:

$$\begin{aligned} &m_z(\alpha, \omega_z, V, \varphi), \\ &C_{xa}(\alpha, \omega_z, V, \varphi), \\ &C_{ya}(\alpha, \omega_z, V, \varphi), \\ &P_x(H, M, P_a), \\ &W\left\{\frac{\varphi_{act}}{\varphi_{ref}}\right\}. \end{aligned}$$

Цельноповоротный стабилизатор:

$$W \left\{ \frac{\varphi_{act}}{\varphi_{ref}} \right\} = \frac{1}{T_{стаб}^2 p^2 + 2T_{стаб}\xi_{стаб}p + 1}$$

$T_{стаб} = 0.03$, $\xi_{стаб} = 0.707$, φ_{act} – фактическое положение, φ_{ref} – заданное положение



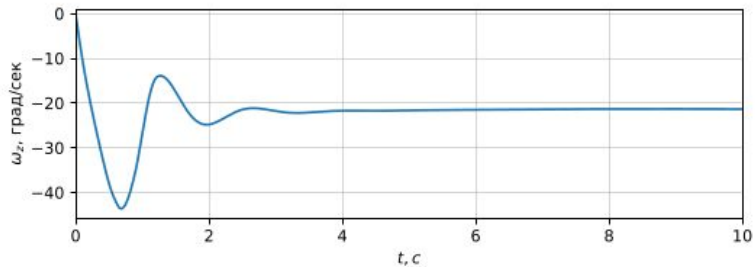
1. Скорость V принимается постоянной.
2. Ограничение на максимальную угловую скорость тангажа:

$$\omega_z = \begin{cases} -60 \text{ град/с} & \text{при } \omega_z \leq -60 \text{ град/с}, \\ \omega_z \text{ град/с} & \text{при } -60 \text{ град/с} < \omega_z < 60 \text{ град/с}, \\ 60 \text{ град/с} & \text{при } \omega_z \geq 60 \text{ град/с}. \end{cases}$$

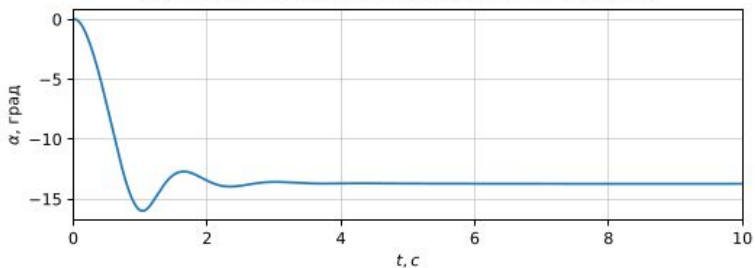
Реакция объекта управления на ступенчатое отклонение стабилизатора



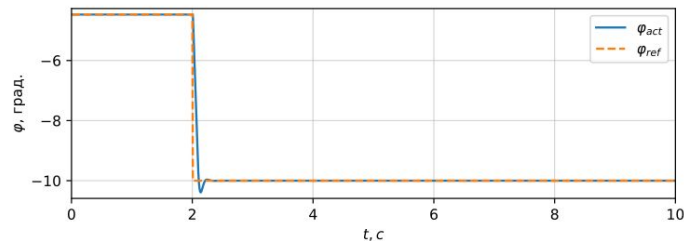
Расчет модели для начального состояния $H=2500$ м, $V=275$ м/с.



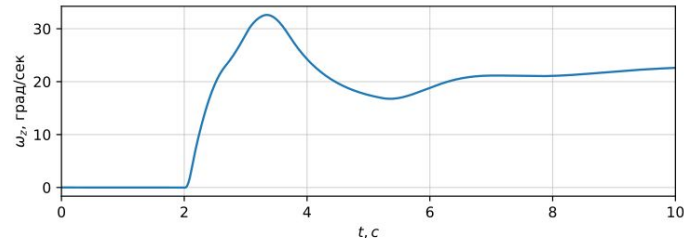
(a) Изменение угловой скорости тангажа ω_z



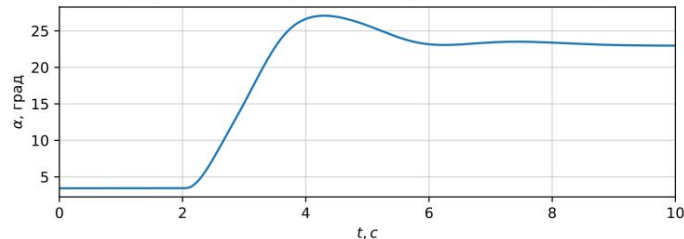
(b) Изменение угла атаки α



(a) Изменение командного сигнала φ_{ref} и φ_{act}



(b) Изменение угловой скорости тангажа ω_z



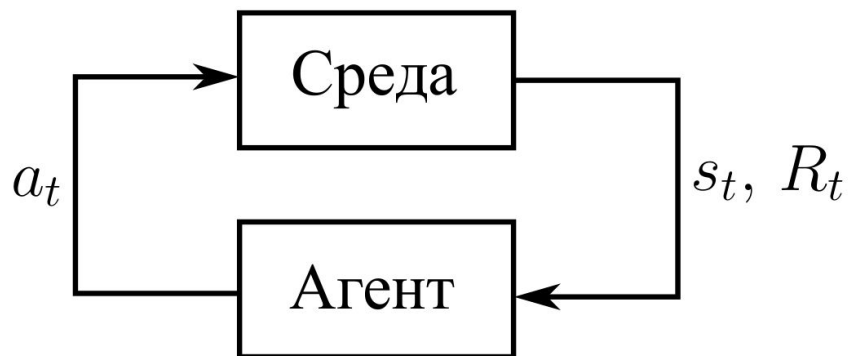
(c) Изменение угла атаки α

Взаимодействие агента со средой в терминах ОсП.

s_t – состояние среды;

R_t – награда на текущем шаге;

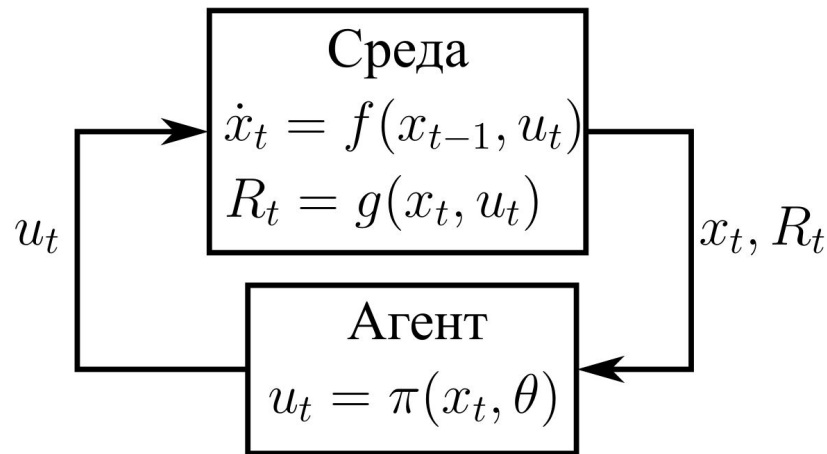
a_t – действие.



Взаимодействие агента со средой в терминах управления движением ЛА.

\mathcal{X} – вектор состояния ЛА;

\mathcal{U} – вектор управления.



Максимизация ожидаемой награды

$$\mathbb{E}_{\pi} [R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \dots],$$

путем выбора параметра θ для стратегии

$$\pi_{\theta}(a_t|s_t, \theta) = P[a_t = a|s_t = s, \theta_t = \theta],$$

на основе оптимизации целевой функции

$$J(\theta) \approx R(s, \pi_{\theta}).$$

Состояние среды:

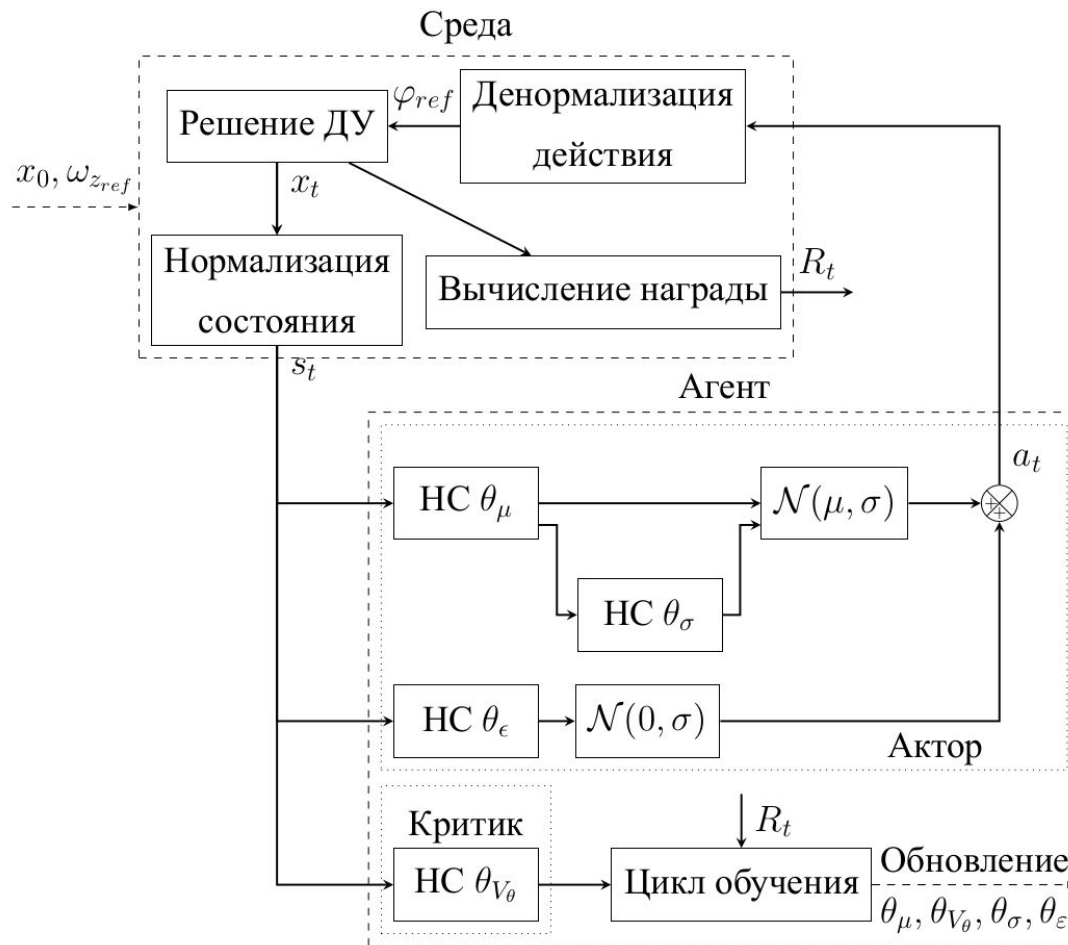
$$s = (H, \omega_z, V, \omega_{z_{ref}}, \omega_{z_{err}}).$$

Действие:

$$a_t = \varphi_{ref}.$$

Цель свести к минимуму ошибку по угловой скорости:

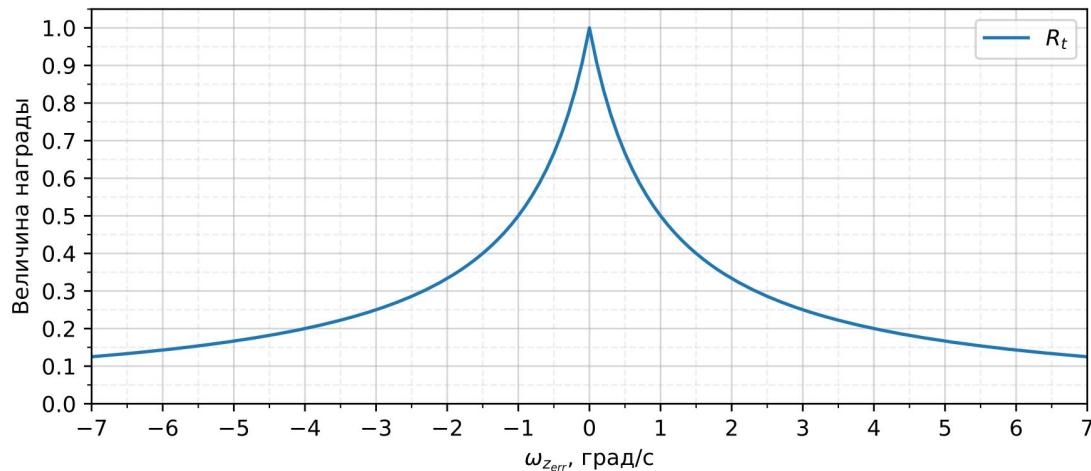
$$\omega_{z_{err}} = \omega_z - \omega_{z_{ref}}.$$



В нашей задаче функция награды имеет вид $R_t = 1 - \frac{\frac{|\omega_{zerr}|}{k}}{1 + \frac{|\omega_{zerr}|}{k}}$,

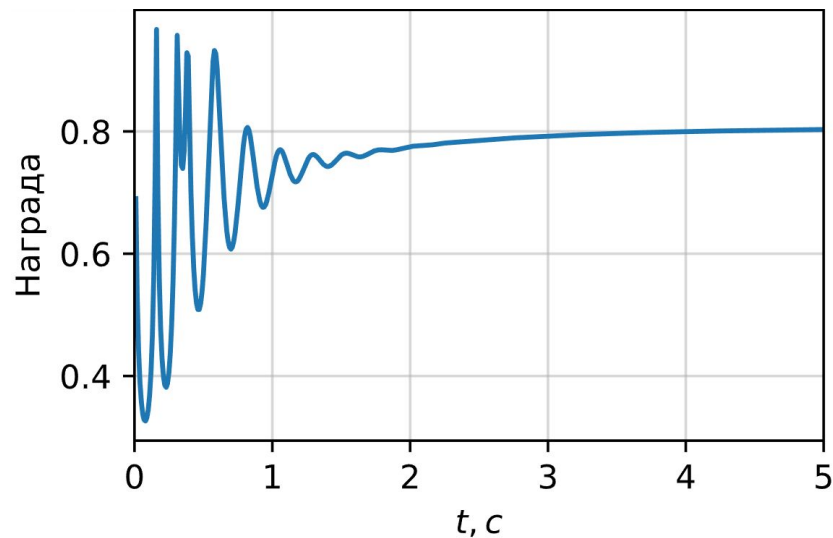
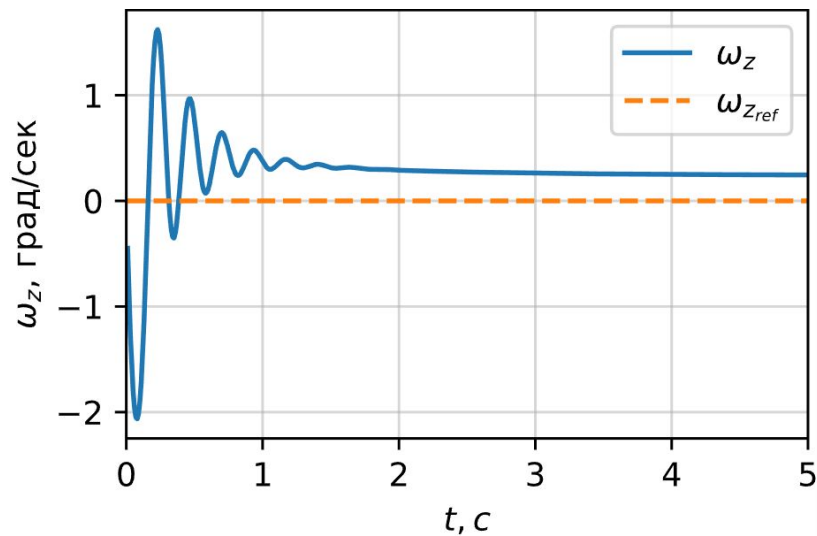
где:

$$\omega_{zerr} = \omega_z - \omega_{zref}.$$

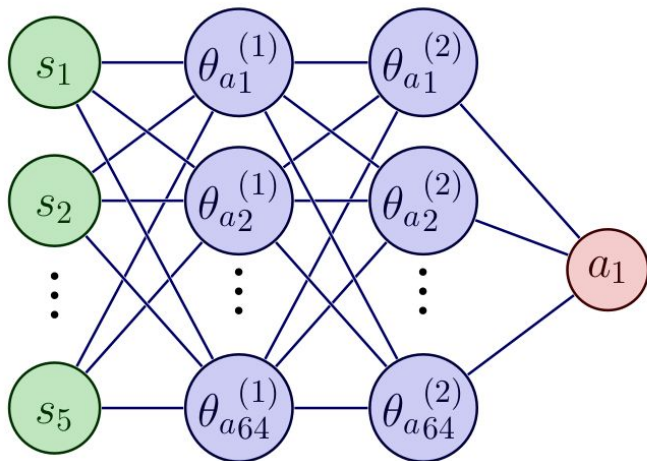


Значение данной функции в каждый момент времени, определяет успешность выполнения цели.

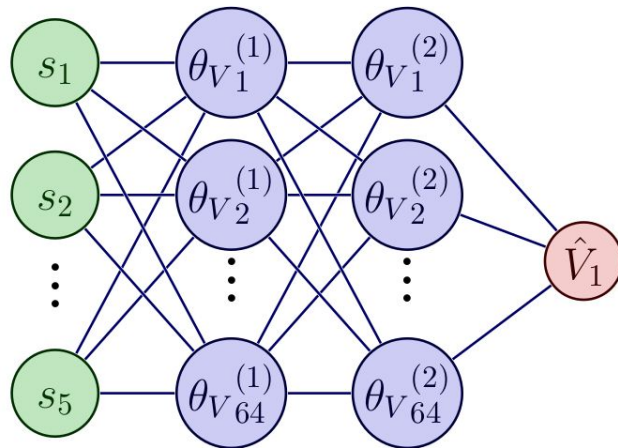
Пример изменения получаемой награды



Структура сети для формирования действия:

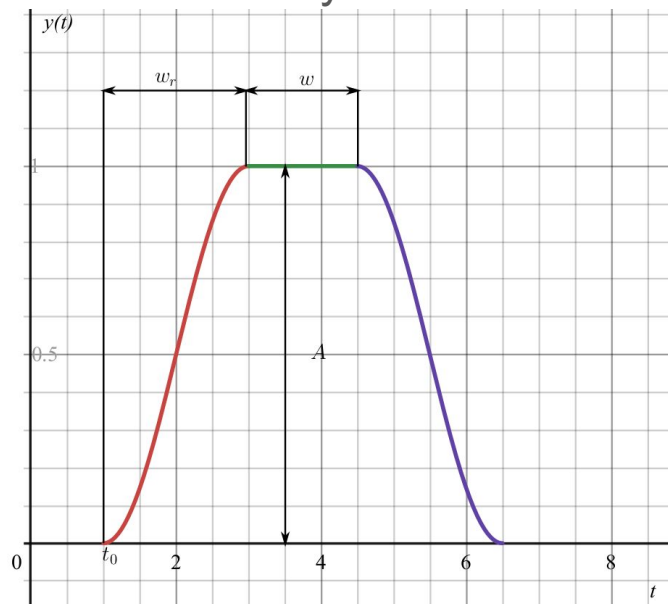


Структура сети для формирования оценок ожидаемой награды:

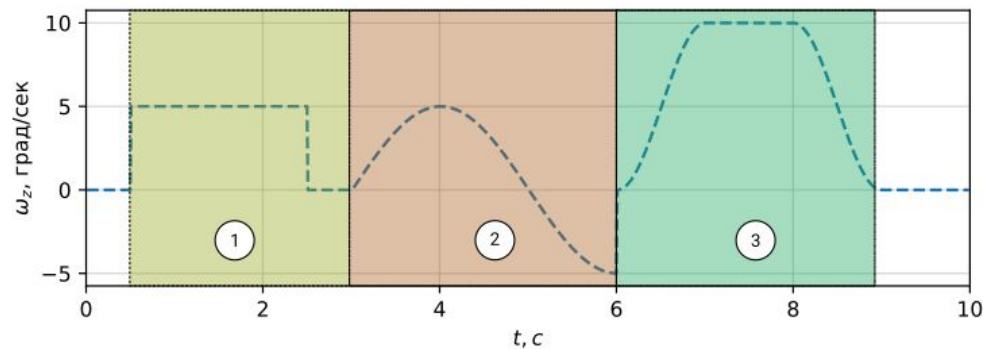


Эти сети необходимы для нахождения градиента целевой функции ∇J

1. Ступенчатый сигнал со случайной амплитудой в диапазоне $-20 \dots 20$ град/с
2. Синусоидальный сигнал со случайной частотой от $0.125 \dots 0.75$ Гц и амплитудой $-20 \dots 20$ град/с
3. Плавный ступенчатый сигнал вида:

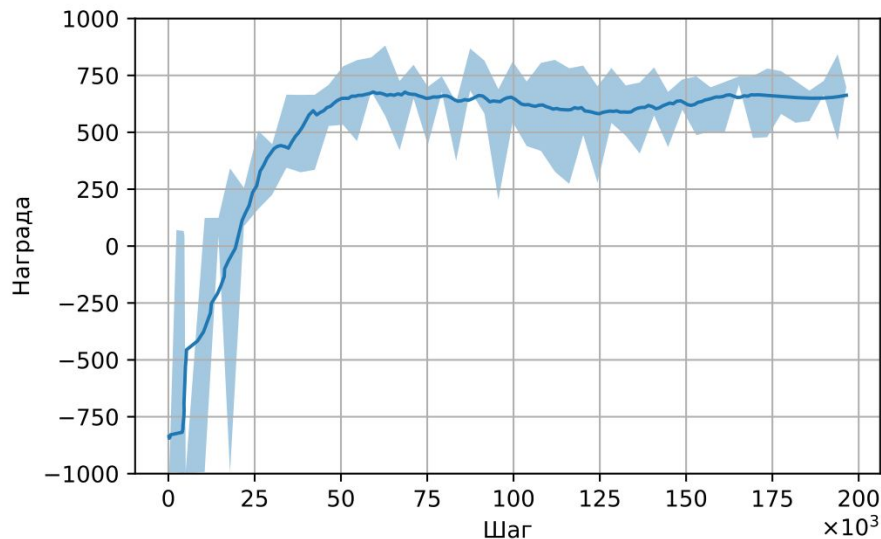


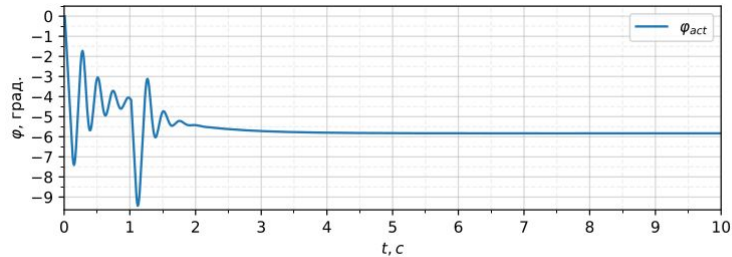
Пример обучающего сигнала как комбинация 3-ех сигналов



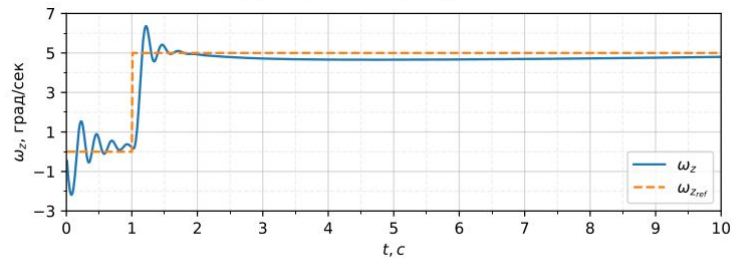
Обучение агента проводилось методом непосредственной оптимизации стратегии (PPO)

Изменение награды в процессе обучения агента

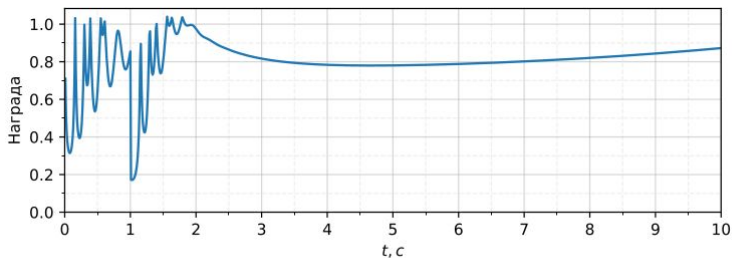




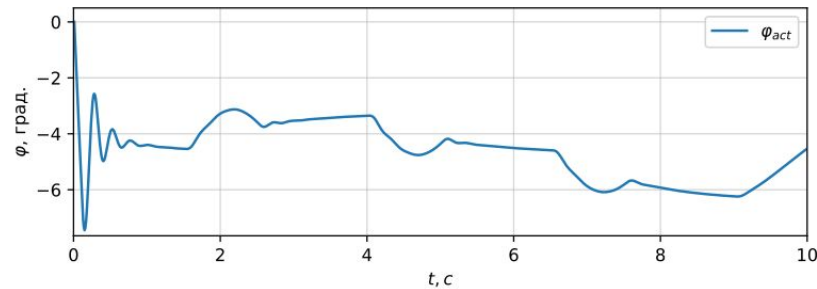
(a) Управляющие действия



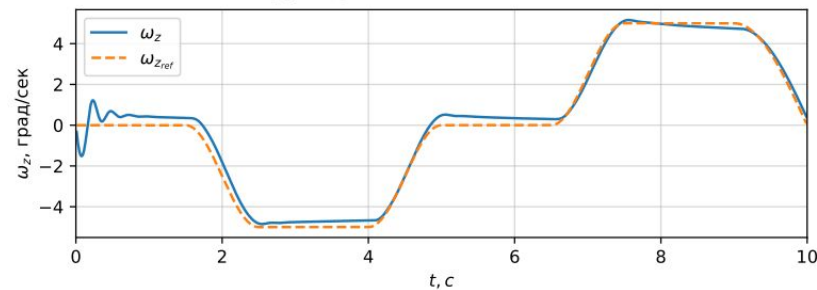
(b) Изменение угловой скорости тангажа ω_z



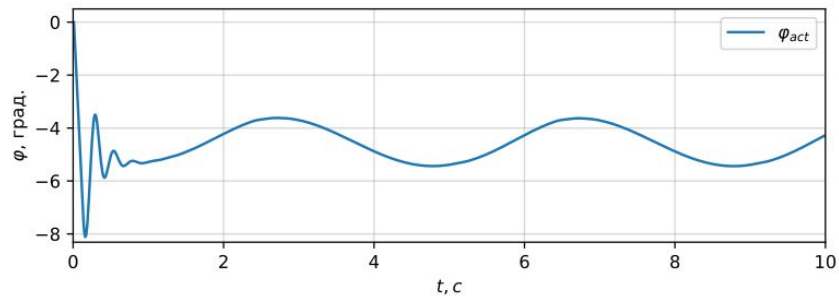
(c) График награды



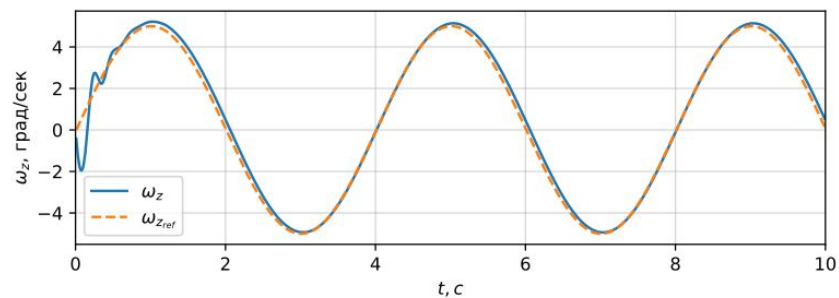
(a) Управляющие действия



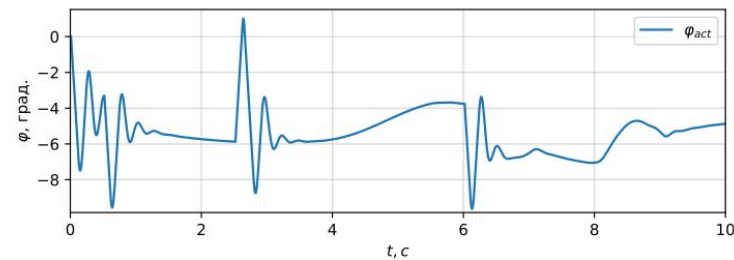
(b) Изменение угловой скорости тангажа ω_z



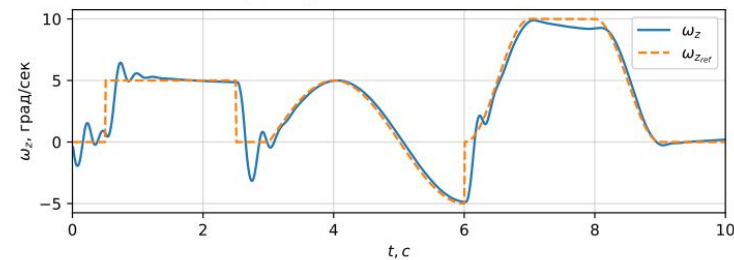
(а) Управляющие действия



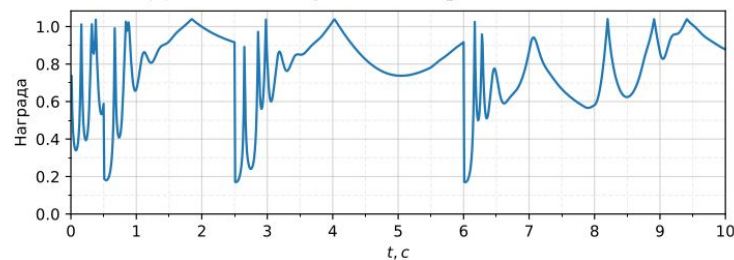
(b) Изменение угловой скорости тангажа ω_z



(а) Управляющие действия



(b) Изменение угловой скорости тангажа ω_z



(с) График награды

Сценарий	Средняя абсолютная ошибка, град/с	Среднеквадратичная ошибка, (град/с) ²
Ступенчатый сигнал	0.328	0.324
Серия ступенчатых сигналов	0.258	0.122
Синусоида	0.212	0.125
Обучающий сигнал	0.526	0.959

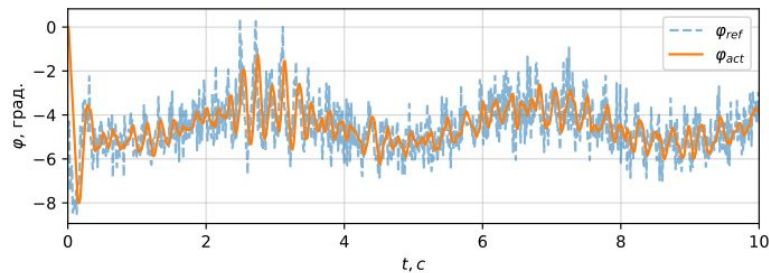
Особые случаи. Влияние шума

Шум на основе нормального распределения с дисперсией 0.5 и мат. ожиданием 0 ($\mathcal{N}(0, 0.5)$)

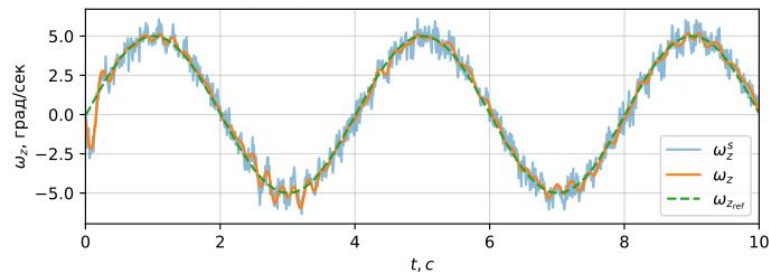
Выбранная величина дисперсии шума в 15 раз больше чем у реальных датчиков.

5.3 нулевой сигнал, °/с, не более	± 0.1
5.4 уровень шумовой составляющей (3σ), °/с, не более	0.1
5.5 порог чувствительности, °/с, не более	0.03

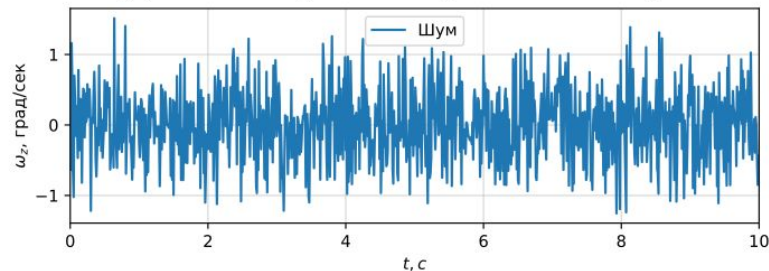
Основные физические и электрические характеристики ИПДММ-2-1
Таблица 1 (продолжено)



(а) Управляющие действия



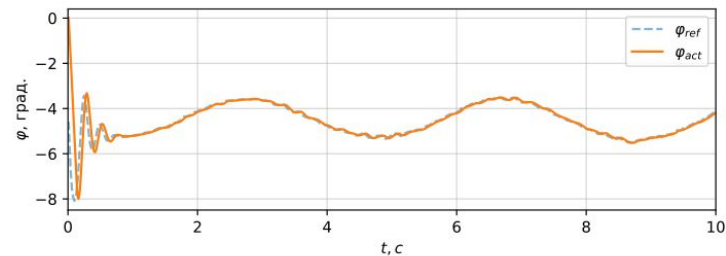
(б) Изменение угловой скорости тангажа ω_z



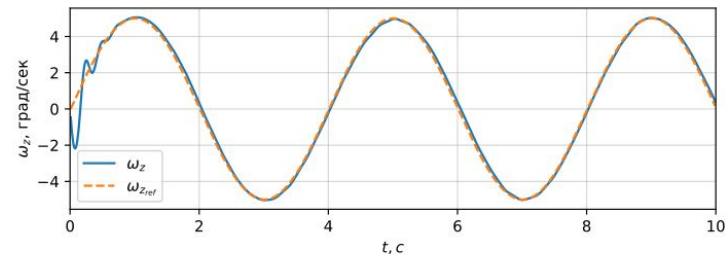
(с) Величина добавленного шума ($\omega_z^s - \omega_z^x$)

Модель турбулентности фон Кармана:

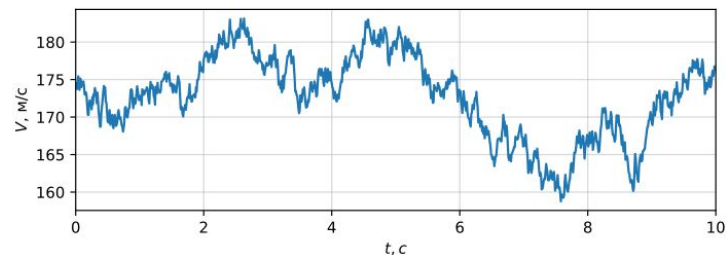
$$\Phi_u(\Omega) = \frac{\sigma_u^2 L_u}{\pi} \frac{1}{(1 + (1.339 L_u \Omega)^2)^{\frac{5}{6}}},$$



(a) Изменение угловой скорости тангажа ω_z



(b) Изменение угловой скорости тангажа ω_z

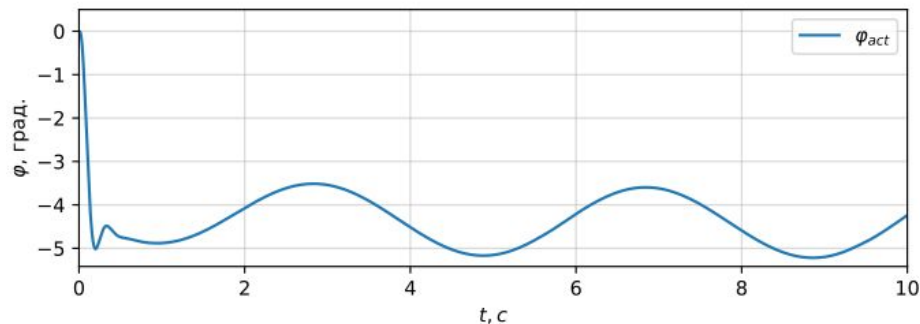


(c) Изменение скорости V

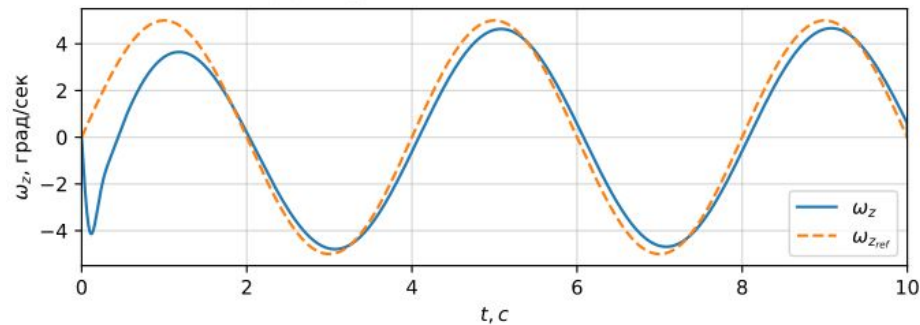
Сценарий	Средняя абсолютная ошибка, град/с	Среднеквадратичная ошибка, (град/с) ²
Турбулентное воздействие	0.169	0.101
Шумовое воздействие	0.271	0.167
Без внешнего воздействия	0.212	0.125

Результат работы PI-контроллера:

Средняя абсолютная ошибка PI-контроллера	0.216
Средняя абсолютная ошибка PPO-контроллера	0.212

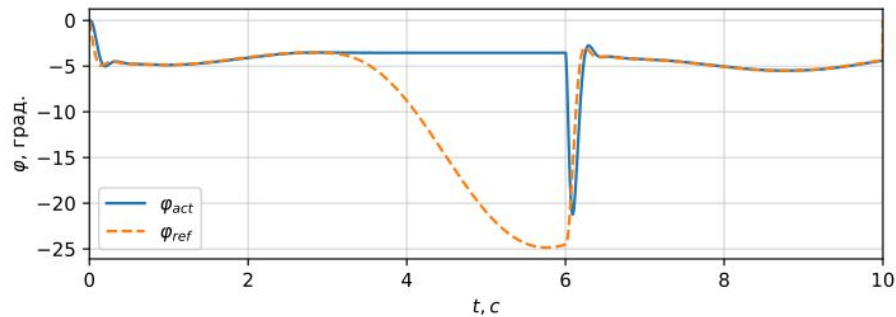


(a) Управляющие действия

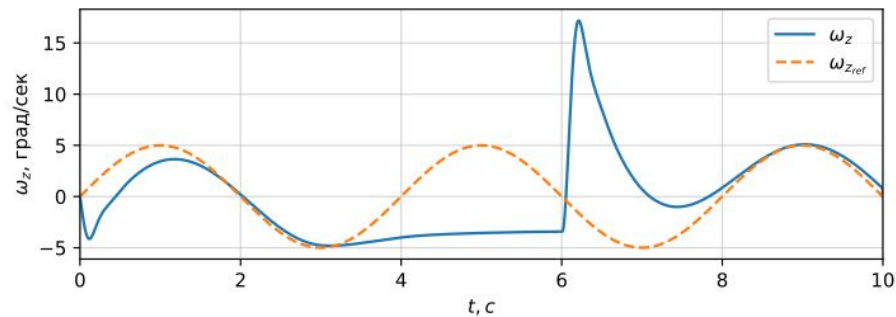


(b) Изменение угловой скорости тангажа ω_z

PI-контроллер

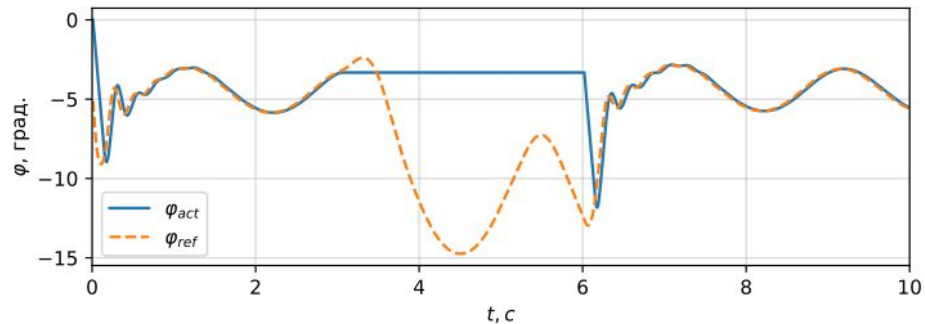


(a) Управляющие действия

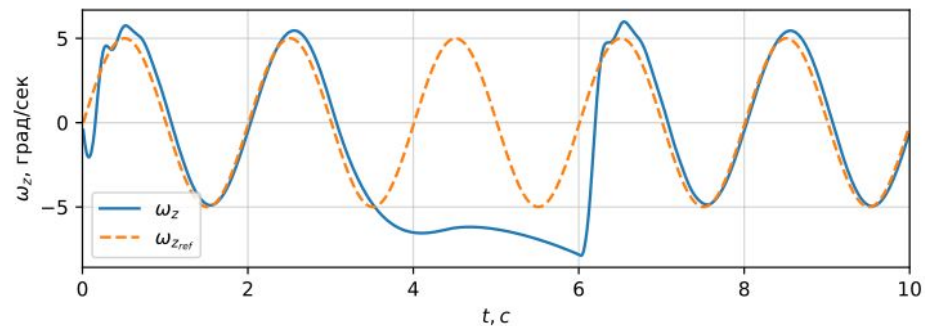


(b) Изменение угловой скорости тангажа ω_z

PPO-контроллер



(a) Управляющие действия



(b) Изменение угловой скорости тангажа ω_z

1. В данной работе был применен метод обучения с подкреплением для формирования РРО-контроллера угловой скорости устойчивого к измерительным шумам и атмосферным турбулентностям.
2. Полученные результаты показывают возможность формирования адаптивного контроллера таким подходом и устойчивости сформированного контроллера в рассмотренных тестовых ситуациях.

Спасибо за внимание!

Реализация
на Python:

