ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2025, № 1, С. 163–180

СИСТЕМЫ УПРАВЛЕНИЯ ДВИЖУЩИМИСЯ ОБЪЕКТАМИ

УДК 629.7.058.4

СОВМЕСТНОЕ ИСПОЛЬЗОВАНИЕ МЕТОДА ДИНАМИЧЕСКОЙ ИНВЕРСИИ И ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ ОПТИМАЛЬНОГО АДАПТИВНОГО УПРАВЛЕНИЯ ДВИЖЕНИЕМ СВЕРХЗВУКОВОГО ПАССАЖИРСКОГО САМОЛЕТА[1]

© 2025 г. Г. Дхиман[a], *, Ю. В. Тюменцев[a], **, Р. А. Цхай[a], ***

[a] *Московский авиационный институт (Национальный исследовательский университет), Москва, Россия*

**e-mail: gd9617@mail.ru*

***e-mail: yutium@gmail.com*

****e-mail: romantskhai106@yandex.ru*

Поступила в редакцию 19.07.2024 г.
После доработки 22.09.2024 г.
Принята к публикации 13.01.2025 г.

Рассматривается задача управления движением летательных аппаратов в условиях неопределенностей, обусловленных неполным и неточным знанием их характеристик, а также нештатными ситуациями в полете, оказывающими влияние на свойства летательного аппарата как объекта управления. Одним из эффективных инструментов решения задач подобного рода, обеспечивающих корректировку алгоритмов управления летательного аппарата с учетом его изменившейся динамики, является обучение с подкреплением в варианте приближенного динамического программирования в сочетании с искусственными нейронными сетями. Применительно к задачам управления поведением сложных динамических систем в последнее десятилетие в рамках приближенного динамического программирования активно развивается семейство методов, известное под наименованием “метод адаптивного критика”. Рассматривается применение одного из вариантов этого подхода и развитие его за счет совместного использования с методом динамической инверсии. Данный подход позволяет формировать оптимальный адаптивный закон управления движением летательного аппарата. Его эффективность демонстрируется на примере управления продольным движением сверхзвукового пассажирского самолета.

Ключевые слова: сверхзвуковой пассажирский самолет, управление движением, динамическая инверсия, приближенное динамическое программирование, метод адаптивного критика, SNAC-подход, оптимальное адаптивное управление

DOI: 10.31857/S0002338825010133 **EDN:** AIJWTM

COMBINED USE OF DYNAMIC INVERSION AND REINFORCEMENT LEARNING FOR OPTIMAL ADAPTIVE CONTROL OF SUPERSONIC TRANSPORT AIRPLANE MOTION

G. Dhiman[a], *, Yu. V. Tiumentsev[a], **, R. A. Tskhay[a], ***

[a] *Moscov Aviation Institute (National Research University), Moscow, Russia*

**e-mail: gd9617@mail.ru*

***e-mail: yutium@gmail.com*

****e-mail: romantskhai106@yandex.ru*

We consider the problem of aircraft motion control under uncertainties caused by incomplete and inaccurate knowledge of the aircraft characteristics, as well as by abnormal situations in flight that affect the properties of the aircraft as a control object. One of the effective tools for solving problems of this kind, providing the adjustment of aircraft control algorithms taking into account its changed dynamics, is reinforcement learning in the variant of Approximate Dynamic Programming (ADP), in combination with artificial neural networks.

[1] Работа подготовлена в рамках Программы развития научно-исследовательского центра мирового класса «Сверхзвук» на 2020-2025 годы, финансируемой Министерством науки и высшего образования Российской Федерации (соглашение от 20 апреля 2022 года, № 075-15-2022-309).

In the last decade, a family of methods known as Adaptive Critic Design (ACD) has been actively developed within the ADP approach to control the behavior of complex dynamical systems. The paper discusses the application of one variant of the ACD approach, namely SNAC (Single Network Adaptive Critic) and its development through combined use with the dynamic inversion method. This approach makes it possible to form an optimal adaptive control law for aircraft motion. Its effectiveness is demonstrated on the example of longitudinal motion control for supersonic transport airplane (SST).

Keywords: supersonic transport airplane, motion control, dynamic inversion, approximate dynamic programming, adaptive critic method, SNAC approach, optimal adaptive control

Введение. Динамическое программирование является средством, потенциально пригодным для синтеза законов управления с обратной связью. Однако данное средство до относительно недавнего времени было малопригодно для решения реальных прикладных задач вследствие характерного для него “проклятия размерности” [1]. Из-за этого обстоятельства требования к вычислительным ресурсам, необходимым для решения задач реального мира, превышали, как правило, разумные пределы. К такого рода задачам, недоступным для решения методами традиционного динамического программирования, относится и синтез законов управления движением летательных аппаратов (ЛА).

Для преодоления этого затруднения П. Вербос предложил подход, известный как приближенное динамическое программирование (approximate dynamic programming (ADP)) [2–8]. Аббревиатуру ADP в ряде случаев расшифровывают как адаптивное динамическое программирование (adaptive dynamic programming). Это обусловлено параметризацией получаемого решения задачи динамического программирования, которая обеспечивает возможность подстройки такого решения в режиме онлайн, согласно изменяющейся ситуации. Другими словами, такая параметризация наделяет рассматриваемую управляемую систему свойством адаптивности.

Подход, предложенный П. Вербосом, позволил совершить прорыв в обучении с подкреплением (reinforcement learning (RL)), для которого динамическое программирование, по крайней мере для задач управления динамическими системами, является математической базой. Как один из вариантов ADP, П. Вербос предложил подход, именуемый методом адаптивного критика (adaptive critic design (ACD)), который объединяет принципы обучения с подкреплением и приближенного динамического программирования. Следует отметить, что ACD-подход существенно основывается на привлечении нейросетевых технологий, а именно нейронных сетей прямого распространения [9]. Именно нейронные сети в составе ACD-алгоритмов обеспечивают их настраиваемую параметризацию, в том числе и в режиме онлайн.

Один из важнейших элементов ACD-алгоритма – так называемый критик, аппроксимирующий некоторую нелинейную функцию, которая представляет собой оценку эффективности формируемого алгоритма в текущей ситуации и при текущих значениях его настраиваемых параметров [10–15]. Помимо критика, в большинстве разновидностей ACD присутствует элемент, именуемый актором. Данный элемент вырабатывает текущее значение управляющего сигнала, т.е. он в терминах динамических систем представляет собой закон управления такой системой.

Оба упомянутых выше элемента (критик и актор), присутствующие в большинстве схем ACD, реализуются чаще всего в виде нейронных сетей прямого распространения. Существует, однако, ACD-схема, в которой имеется только критик – схема SNAC (single network adaptive critic) [16–21]. Формирование управляющего сигнала в схеме SNAC вместо актора осуществляется с помощью оптимизационного алгоритма, основанного на соотношениях для линейно-квадратичного регулятора.

SNAC-подход позволяет формировать законы управления как для линейных, так и для нелинейных систем. Обучение нейронных сетей, входящих в состав алгоритма SNAC, для случая нелинейных систем представляет собой непростую задачу. Работа с линейными системами значительно проще. Однако традиционный подход к линеаризации исходной нелинейной системы основан на использовании разложения в ряд Тейлора. Вследствие этого закон управления будет адекватен лишь в небольшой окрестности режима функционирования системы, для которого выполнялась линеаризация. Пример формирования алгоритма SNAC для этого случая был рассмотрен в нашей предыдущей статье [22]. Альтернативным подходом является

применение нелинейной динамической инверсии (nonlinear dynamic inversion (NDI)) [23–26], позволяющий получить линеаризованную модель, адекватную для всей области режимов функционирования представленной системы. Частным случаем NDI выступает динамическая инверсия (dynamic inversion (DI)), ориентированная исходно на работу с линейными системами. Использование динамической инверсии дает возможность решать задачу корректировки динамических свойств объекта управления путем применения ее во внутреннем контуре, что упрощает синтез закона управления во внешнем контуре.

Динамическая инверсия представляет собой эффективный инструмент, облегчающий решение задачи синтеза законов управления. Однако DI-подход обладает также и существенным недостатком, который состоит в чувствительности данного метода к изменению динамики объекта управления. Устранить этот недостаток можно путем онлайн-корректировки DI, используя линейные нейронные сети.

В следующих разделах дается краткое описание сути динамической инверсии и SNAC-подхода. На этой основе формируется процесс адаптации, обеспечивающий возможность адекватного управления динамической системой в условиях неполного и неточного знания свойств объекта управления и среды, в которой он работает. Применение рассматриваемых методов демонстрируется на примере реальной задачи управления продольным движением сверхзвукового пассажирского самолета второго поколения.

1. Машинное обучение как инструмент синтеза системы управления. *1.1. Искусственные нейронные сети прямого распространения и их значение для обучения с подкреплением при синтезе законов управления.* Развитие теории нейронных сетей обеспечило возможность решать многие задачи, с которыми неудовлетворительно справляются методы традиционной вычислительной математики. Одна из таких задач – аппроксимация нелинейной функции многих переменных. Методы решения этой задачи являются важным инструментом, применяемым при синтезе законов управления системами, основанном на ADP-подходе. В частности, в таком варианте ADP-подхода, как метод адаптивного критика, который активно используется для синтеза законов управления, необходимо решать задачи формирования критика и актора. Эти два элемента синтезируемой системы управления представляют собой параметризованные функциональные зависимости многих переменных. Наиболее удобный вариант представления этих зависимостей состоит в применении нейронных сетей прямого распространения. Как известно, сети такого вида, в составе которых имеются нелинейные элементы, позволяют аппроксимировать функции многих переменных с любой наперед заданной точностью [9, 27–29].

Следует подчеркнуть, что именно сочетание возможностей обучения с подкреплением с возможностями нейронных сетей прямого распространения обусловило активное развитие и использование ADP-подхода для решения проблем оптимального и адаптивного управления динамическими системами различного вида и назначения. Расширение диапазона применимости такого подхода достигается за счет объединения его с методом нелинейной динамической инверсии. Эффективная реализация NDI также требует привлечения нейронных сетей прямого распространения, используемых для представления параметризованного нелинейного преобразования в обратной связи.

Кроме того, во многих случаях для синтеза закона управления требуется модель рассматриваемой динамической системы. Применительно к ЛА, модель движения для них – система обыкновенных дифференциальных уравнений [30, 31]. Важнейшим элементом, входящим в состав такой модели движения, являются соотношения для безразмерных коэффициентов аэродинамических сил и моментов, действующих на ЛА. Эти соотношения представляют собой нелинейные функции нескольких переменных, для которых также может быть реализовано их представление в виде нейронной сети прямого распространения. Особенно эффективным будет такое представление в случае, когда исходные данные для получения функций имеют вид многомерных таблиц, найденных в ходе экспериментов в аэродинамических трубах и в летных испытаниях.

1.2. Обучение с подкреплением и метод адаптивного критика. В своем типовом варианте RL-система может быть представлена как совокупность из следующих четырех компонент:

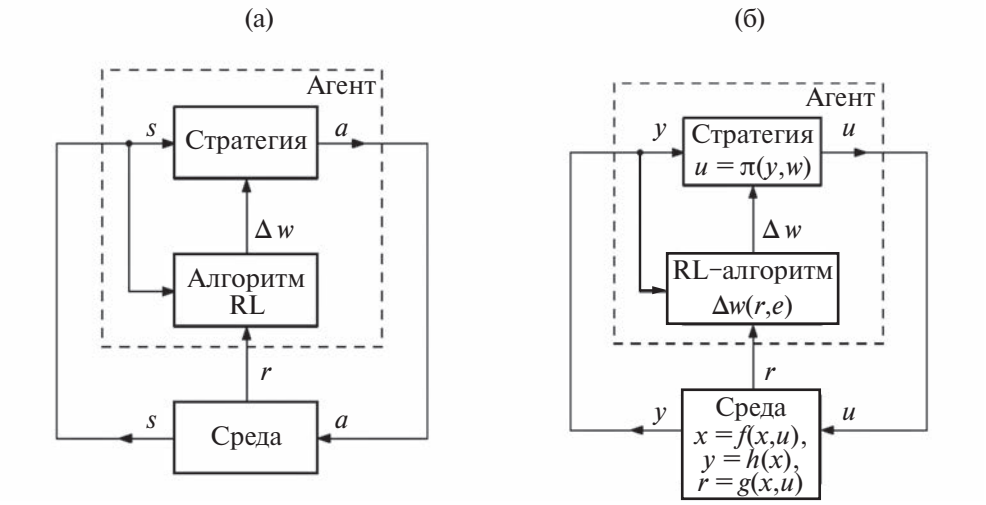

Рис. 1. Обобщенная схема обучения с подкреплением.

$$\mathbb{S}_{\mathrm{RL}} = \{S, A, R, F\}, \tag{1.1}$$

где S – множество состояний системы (1.1); A – множество возможных действий данной системы; R – множество сигналов вознаграждения (подкрепления); F – функция, определяющая переход системы $\mathbb{S}_{\mathrm{RL}}$ из одного состояния в другое, т.е. $F: S \times A \to S$.

Общая схема обучения с подкреплением для некоторой системы $\mathbb{S}_{\mathrm{RL}}$ показана на рис. 1а [32–34]. Стратегия (policy) π для этой системы определяется как отображение $\pi: S \to A$. Система $\mathbb{S}_{\mathrm{RL}}$ в терминологии RL-подхода носит наименование “агент” (agent). В некоторый момент времени t агент $\mathbb{S}_{\mathrm{RL}}$ находится в состоянии $s_t \in S$. Он воспринимает сигнал вознаграждения r_t и предпринимает действие $a_t \in A$, определяемое стратегией π, т.е. $a_t = \pi(s_t)$. В результате $\mathbb{S}_{\mathrm{RL}}$ переходит в некоторое следующее состояние $s_{t+1} = F(s_t, a_t)$, получая при этом сигнал вознаграждения $r_{t+1} = r(s_t, a_t, s_{t+1}) \in R$. На рис. 1б дается вариант общей RL-схемы применительно к задаче управления динамическими системами. Стратегия π в системе $\mathbb{S}_{\mathrm{RL}}$ является параметризованной. Как правило, в ADP-подходе стратегия реализуется в виде нейронной сети прямого распространения. В таком случае настраиваемые параметры w этой сети (синаптические веса и смещения) корректируются в процессе ее обучения, в том числе и при необходимости подстройки закона управления под изменившуюся динамику объекта. Эти корректировки Δw выполняются соответствующим RL-алгоритмом.

Цель обучения с подкреплением состоит в том, чтобы сформировать стратегию π, которая максимизирует суммарное (совокупное) вознаграждение, получаемое системой $\mathbb{S}_{\mathrm{RL}}$, исходя из ее начального состояния s_0 при $t = 0$, т.е. в терминологии задач управления полетом – за выполнение летной операции в целом.

Как уже отмечалось выше, одним из практически важных вариантов ADP-подхода является класс ACD-методов [11–15], который успешно применяется для формирования адаптивных оптимальных законов управления для динамических систем различных видов. При этом оптимальность получаемых законов управления обусловлена использованием средств динамического программирования как основы ACD-подхода, а адаптивность – параметризацией актора и критика в форме нейронных сетей прямого распространения с возможностью их онлайн-корректировки. Общая схема системы, реализующей ACD-подход, показана на рис. 2.

На рис. 2 приняты следующие обозначения. Объект управления описывается нелинейным дифференциальным уравнением:

$$\dot{x} = f\big(x(t)\big) + g\big(x(t)\big)u(t), \tag{1.2}$$

где $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ – вектор состояния и $u = (u_1, \ldots, u_n)^{\mathrm{T}}$ – вектор управления рассматриваемой динамической системы, начальные условия для нее имеют вид $x(t_0) = x_0$.

Критерий эффективности закона управления задается как функционал следующего вида:

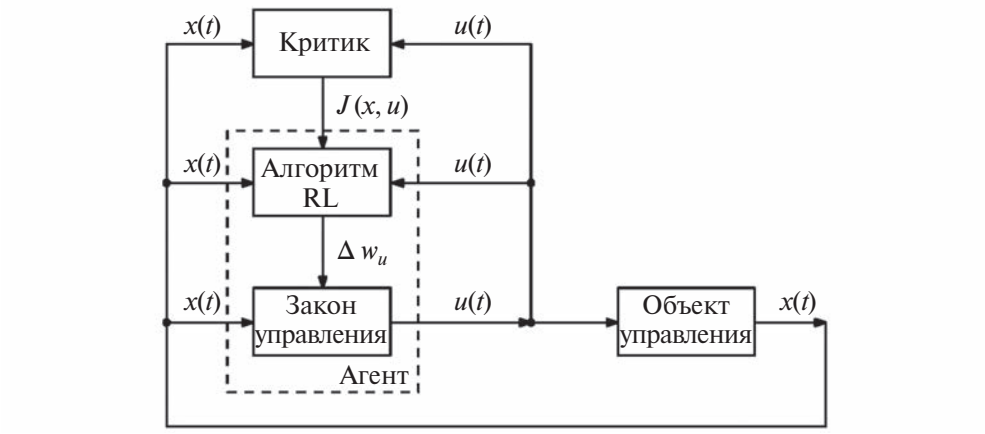

Рис. 2. Общая структура ACD-алгоритма адаптивного управления динамической системой.

$$J\left(x(t),u(t)\right) = \int_{t}^{\infty} F\left(x(\tau),u(\tau)\right)d\tau. \tag{1.2}$$

Данный критерий вытекает из принципа оптимальности Беллмана, который формулируется следующим образом [1]: “Отрезок оптимального процесса от любого текущего момента времени t до конца процесса сам является оптимальным процессом с началом в этот текущий момент времени”. Так как в момент времени t неизвестно, как изменится в дальнейшем динамика объекта, то традиционный прием преодоления данного затруднения – обратная рекурсия, с помощью которой организуются вычисления в направлении от завершающего момента времени к стартовому, что обеспечивает возможность получить в промежуточный момент времени t искомое управление, удовлетворяющее принципу оптимальности Беллмана.

В критерии (1.3) функция $V(x,u)$ применительно к задачам управления движением в общем случае обычно определяется как

$$V\left(x(t),u(t)\right) = P\left(x(t)\right) + u^{\mathrm{T}}(t)Ru(t). \tag{1.4}$$

Соотношение (1.4) в задачах управления возмущенным движением самолета, рассматриваемых в данной статье, можно интерпретировать как штраф за уклонение от заданной опорной траектории (первое слагаемое) и штраф за расход управляющего ресурса на подавление этого уклонения (второе слагаемое).

С учетом (1.3) и (1.4) цель управления, реализуемого ACD-алгоритмом, состоит в том, чтобы получить оптимальный адаптивный закон управления $u^* \in U$ с обратной связью, минимизирующий критерий $J(x,u)$, т.е.

$$J^*\left(x\right) = \min_{u\in U} \int_{t}^{\infty} F\left(x(\tau),u(\tau)\right)d\tau. \tag{1.5}$$

Составные элементы ACD-схемы, показанные на рис. 2, реализуют следующие функции: критик для момента времени t дает оценку $J(x,u)$ эффективности текущего варианта закона управления; агент реализует текущий закон управления и корректирует его, согласно оценке значения критерия $J(x,u)$, полученной от критика; объект управления – рассматриваемая динамическая система с учетом воздействий на нее внешней среды. При этом в большинстве случаев критик, закон управления и модель объекта управления реализуются как многослойные нейронные сети прямого распространения.

2. Реализация метода динамической инверсии для управления движением ЛА. *2.1. Формирование закона управления с помощью динамической инверсии*. Основная область применения метода динамической инверсии, как отмечалось выше, связана с нелинейными системами. Однако и частный случай этого метода, ориентированный на работу с линейным объектом управления [31], может представлять самостоятельный интерес. Это имеет место, например, в случае,

когда на борту ЛА решается задача онлайн-идентификации линеаризованной модели объекта управления. Наличие такой оперативно корректируемой модели дает возможность эффективно использовать динамическую инверсию для корректировки закона управления применительно к текущей динамике самолета [35].

Еще один пример, показывающий возможность применения динамической инверсии в линейном варианте, рассматривается в следующих разделах. Он связан с реализацией регулятора с использованием SNAC-подхода и показывает возможность существенно повысить эффективность данного регулятора за счет совместного применения SNAC и динамической инверсии.

Рассмотрим суть метода динамической инверсии в линейном варианте. Пусть объект управления описывается следующими уравнениями в пространстве состояний:

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx, \end{aligned} \tag{2.1}$$

где $x(t) \in R^n$ – вектор состояний, $u(t) \in R^m$ – вектор управлений, $y(t) \in R^s$ – вектор выходов, $x(t_0) = x_0$ – начальные условия. Будем считать, что для измерений доступны все линейные комбинации компонент вектора $x(t)$.

Примем, что размерности управляющего вектора $u(t)$ и вектора выходов $y(t)$ совпадают ($m = s$). Для самолетов это часто встречающаяся ситуация, когда каждой степени свободы отвечает свой орган управления. Если таких органов управления несколько или они секционированы, то необходимо решать тем или иным способом задачу распределения управлений между степенями свободы самолета (control allocation problem), что позволяет свести задачу к предыдущему случаю.

Будем далее рассматривать задачу отслеживания некоторого задающего сигнала $r(t)$, подаваемого на вход системы, размерность которого совпадает с размерностью вектора выходов. В данной задаче цель управления состоит в том, чтобы максимально точно воспроизводить задающий сигнал на выходе системы. Ошибку отслеживания, которую требуется минимизировать, определим следующим образом:

$$e(t) = r(t) - y(t). \tag{2.2}$$

В методе динамической инверсии используется следующий прием для того, чтобы получить требуемый закон управления. Выражение для выхода $y(t)$ надо продифференцировать столько раз, сколько потребуется, чтобы в выражении для производной появилось управление $u(t)$. В общем случае нелинейной динамической системы данный прием называется “линеаризация обратной связью по входам-выходам” (input-output feedback linearization). Если же система уже является линейной, то дифференцирование выражения для выхода $y(t)$ системы (2.1) приводит к получению следующего уравнения:

$$\dot{y} = C\dot{x} = CAx + CBu, \tag{2.3}$$

где $u(t)$ будет искомым управлением, если матрица CB ненулевая. В этом случае, в силу принятого выше условия о равенстве размерностей векторов $u(t)$ и $y(t)$, решение найдено. Если же $CB = 0$, следует продолжить процесс дифференцирования, пока не будет выполнено условие $CB \neq 0$, например:

$$\ddot{y} = C\ddot{x} = CA\dot{x} + CB\dot{u} = CA^2x + CABu. \tag{2.4}$$

Определим вспомогательный входной сигнал $v(t)$:

$$v = CBu + CAx - \dot{r}, \tag{2.5}$$

откуда

$$u = (CB)^{-1}(\dot{r} - CAx + v). \tag{2.6}$$

Подставляя это выражение для $u(t)$ в выражение (2.3), получим:

$$\begin{aligned}\dot{y} &= CAx + CB\left[\left(CB\right)^{-1}\left(\dot{r} - CAx + v\right)\right] = \\ &= CAx + \dot{r} - CAx + v\end{aligned} \tag{2.7}$$

или

$$\dot{e} = -v. \tag{2.8}$$

Уравнение (2.7) задает динамику ошибки для исходной замкнутой системы в рассматриваемом случае. При этом функция $v(t)$ может быть задана следующим образом:

$$v = Ke, \tag{2.9}$$

тогда уравнение динамики ошибки замкнутой системы примет вид:

$$\dot{e} = -Ke, \tag{2.10}$$

при этом динамика будет устойчива, если матрица K является положительно определенной. На практике часто эта матрица задается диагональной, чтобы исключить перекрестные связи между каналами управления.

С учетом выполненных преобразований закон управления, сформированный с использованием метода динамической инверсии, запишем как

$$u = (CB)^{-1}\left(\dot{r} + Ke - CAx\right). \tag{2.11}$$

2.2. Онлайн-корректировка закона управления. Как уже было сказано ранее, динамическая инверсия требует точной модели объекта управления. Если по каким-либо причинам свойства объекта изменились и эти изменения не представлены в его модели, метод динамической инверсии становится неэффективным. Для преодоления этого препятствия предлагается представить алгоритм динамической инверсии в виде линейной нейронной сети.

Пусть требуется построить регулятор для управления выходным сигналом y_i, который совпадает с x_j, с помощью управления u_k. Рассмотрим дифференциальное уравнение для этого состояния без учета других управляющих сигналов:

$$\dot{x}_j = a_{j1}x_1 + a_{j2}x_2 + \ldots + a_{jn}x_n + b_{jk}u_k,\ j = \overline{1,n}; k = \overline{1,m}. \tag{2.12}$$

Если выбрать управление u_k в виде

$$u_k = -\frac{a_{j1}x_1 + a_{j2}x_2 + \ldots + a_{jn}x_n}{b_{jk}} + r\left(t\right), \tag{2.13}$$

где $r(t)$ – некоторый задающий сигнал, который система должна воспроизвести как можно более точно, то получаем следующее уравнение:

$$\dot{y}_i = \dot{x}_j = r\left(t\right). \tag{2.14}$$

Результатом выполненного преобразования является измененная динамика объекта управления, которая упрощает работу с замкнутой системой.

Рассмотрим уравнение (2.13) без учета задающего сигнала и представим его немного в другом виде:

$$u_k = -\frac{a_{j1}}{b_{jk}}x_1 - \frac{a_{j2}}{b_{jk}}x_2 - \ldots - \frac{a_{jn}}{b_{jk}}x_n. \tag{2.15}$$

Данное выражение перепишем в виде, который можно интерпретировать как линейную нейронную сеть:

$$u_k = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n = wx, \tag{2.16}$$

где w – вектор весовых коэффициентов сети ($w_1 = -a_{j1}/(b_{jk})$, $w_2 = -a_{j2}/(b_{jk})$ и т.д.), x – вектор сигналов, идущих на вход сети, а u_k – выходной сигнал сети. Фактически весовые коэффициенты сети напрямую зависят от динамики объекта, и для их расчета необходимо производить идентификацию динамической системы.

Представим первое уравнение из системы (2.1) в дискретном времени. Получить такое представление можно, например, при помощи схемы Эйлера с шагом дискретизации Δt:

$$x_{p+1} = x_p + \Delta t\left(Ax_p + Bu_p\right), \tag{2.17}$$

где индекс $p = 1, 2, \ldots$ указывает на момент времени t_p, для которого рассматривается значение соответствующей переменной.

Из соотношения (2.17) получим необходимый нам вариант уравнения (2.12) в дискретной форме:

$$x_j(p+1) = x_j(p) + \Delta t\left(a_{j1}x_1(p) + a_{j2}x_2(p) + \ldots + a_{jn}x_n(p) + b_{jk}u_k(p)\right). \tag{2.18}$$

Если переписать его в виде

$$\frac{x_j(p+1) - x_j(p)}{\Delta t} = a_{j1}x_1(p) + a_{j2}x_2(p) + \ldots + a_{jn}x_n(p) + b_{jk}u_k(p), \tag{2.19}$$

то можно построить для момента времени $p+1$ уравнение относительно значений $a_{j1}, a_{j2}, \ldots, a_{jn}, b_{jk}$. Для решения этого уравнения необходимо получить данные о состояниях и управлениях на n временных шагах, начиная с момента времени p. Эти данные позволяют записать систему из $n+1$ линейных неоднородных уравнений, решение которой дает необходимые значения параметров a_{ij} и b_{jk} для соотношений (2.15) и (2.16). Вариант, когда параметр b_{jk} равен нулю, не рассматривается, так как он относится к случаю неуправляемых динамических систем, не являющихся предметом нашего исследования.

3. Реализация SNAC-подхода в задаче управления нелинейной и линейной системой. ADP-подход потенциально пригоден для работы с нелинейными системами. Рассмотрим следующую модель в пространстве состояний:

$$\begin{aligned} \dot{x} &= f(x) + g(x)u, \\ y &= h(x), \end{aligned} \tag{3.1}$$

где $x \in R^n$, $u \in R^m$ и $y \in R^s$ являются векторами состояний, управления и выхода системы соответственно. Необходимо найти такое управление в виде обратной связи, которое минимизировало бы функционал:

$$J = \int_t^T \left(y^{\mathrm{T}} Q y + u^{\mathrm{T}} R u\right) dt = \int_t^T \left(h^{\mathrm{T}}(x) Q h(x) + u^{\mathrm{T}} R u\right) dt. \tag{3.2}$$

Здесь $Q \geqslant 0$ – положительно-полуопределенная матрица весовых коэффициентов состояний системы, $R > 0$ – положительно-определенная матрица весовых коэффициентов управлений системы.

Как уже отмечалось выше, схема SNAC отличается от других разновидностей ACD-схем тем, что в ней закон управления реализуется не как актор в виде нейронной сети прямого распространения, а формируется с помощью оптимизационного алгоритма [16–21].

Для получения соотношения, определяющего оптимальное управление в SNAC, необходима модель объекта управления. При работе с различными вариантами ADP, включая SNAC, обычно используется модель с дискретным временем. Найти ее можно, как и в случае с уравнениями (2.17), с помощью разностной схемы Эйлера с шагом дискретизации Δ_t:

$$\begin{aligned} x_{p+1} &= x_p + \Ät\left[f\left(x_p\right) + g\left(x_p\right)u_p \right] = \\ &= F\left(x_p, u_p\right), \end{aligned} \tag{3.3}$$

$$y_p = h\left(x_p\right),$$

где индекс $p = 1, 2, ...$, как и в случае выражения (2.17), указывает на значение соответствующей переменной в момент времени t_p.

Первое из уравнений в (3.3), описывающее изменение состояния объекта управления, можно переписать как

$$x_{p+1}^{(i)} = F_i\left(x_p, u_p\right),\ i = \overline{1,n}, \tag{3.4}$$

где

$$F_i\left(x_p, u_p\right) = F_i\left(x_p^{(1)}, \ldots, x_p^{(n)}, u_p^{(1)}, \ldots, u_p^{(m)}\right),\ i = \overline{1,n}. \tag{3.5}$$

С помощью процедуры, аналогичной использованной для нахождения (3.3), можно получить дискретную форму представления для критерия оптимальности (3.2), основываясь на методе прямоугольников для вычисления значения определенного интеграла:

$$J = \sum_{p=1}^{T}\left(p_p^{\mathrm{T}} Q y_p + u_p^{\mathrm{T}} R u_p \right)\Delta t \tag{3.6}$$

или с учетом соотношения $y_p = h(x_p)$, из (3.2):

$$J = \sum_{p=1}^{T}\left(h^{\mathrm{T}}(x_p) Q h(x_p) + u_p^{\mathrm{T}} R u_p \right)\Delta t. \tag{3.7}$$

Целью решения задачи управления системой (3.3) является формирование такой последовательности воздействий u_p^*, $p = \overline{1, T}$, что критерий оптимальности (3.7) примет минимальное значение. Решение данной задачи получим путем сочетания принципа оптимальности Беллмана и метода множителей Лагранжа [36, 37].

В терминологии, предложенной П. Вербосом в работе [2], слагаемые в критерии (3.6), (3.7) именуются функцией потерь, которая в момент времени t_p принимает следующий вид:

$$\psi_p = \left(y_p^{\mathrm{T}} Q y_p + u_p^{\mathrm{T}} Q u_p \right) = \left(h^{\mathrm{T}}(x_p) Q h(x_p) + u_p^{\mathrm{T}} R u_p \right). \tag{3.8}$$

Для момента времени $1 \le p \le T$ критерий (3.7) с учетом (3.8) можно переписать как

$$J_p = \sum_{i=p}^{T} \psi\left(x_i, u_i\right) = \psi\left(x_p, u_p\right) + J_{p+1} = \psi\left(x_p, u_p\right) + \sum_{i=p+1}^{T} \psi\left(x_i, u_i\right).$$

Согласно принципу оптимальности Беллмана [1, 4, 37], оптимальная стоимость перехода системы (3.3) по критерию (3.9) из текущего момента времени p в завершающий момент времени T задается соотношением

$$J_p^* = \min_{u_p}\left[\psi\left(x_p, u_p\right) + J_{p+1}^* \right]. \tag{3.10}$$

Тогда оптимальным для момента времени p будет управление, определяемое соотношением

$$u_p^* = \operatorname*{argmin}_{u_p}\left[\psi\left(x_p, u_p\right) + J_{p+1}^* \right]. \tag{3.11}$$

Задачу (3.10)–(3.11) можно переформулировать следующим образом: требуется найти значение вектора $u_p \in R^m$, минимизирующее функцию

$$L_p = L\left(x_p, u_p\right) = \psi\left(x_p, u_p\right) + J^*_{p+1}, \tag{3.12}$$

при условии, что связь между переменными x и u описывается соотношениями (3.3).

Для решения задачи (3.11) с учетом условий (3.4) введем вспомогательную функцию $H_p = H(x_p, u_p, \lambda_p)$ следующего вида:

$$H_p = H\left(x_p, u_p, \lambda_p\right) = L\left(x_p, u_p\right) + \lambda^{\mathrm{T}}_{p+1} F\left(x_p, u_p\right), \tag{3.13}$$

где $F_p = F(x_{\mathrm{p}}, u_p)$ определяется соотношением (3.4), а $l_p \in R^n$ – значение вектора сопряженных переменных (множителей Лагранжа) для момента времени p.

С использованием (3.13) уравнение для состояний объекта управления из (3.3) можно переписать как

$$x_{p+1} = \frac{\partial H_p}{\partial \lambda_{p+1}} = F\left(x_p, u_p\right), \tag{3.14}$$

а соответствующее ему сопряженное уравнение, определяющее переменную λ_p, принимает вид:

$$\lambda_p = \left(\frac{\partial H_p}{\partial x_p}\right)^{\mathrm{T}} \lambda_{p+1} + \frac{\partial \psi_p}{\partial x_p}. \tag{3.15}$$

Необходимое условие оптимальности в рассматриваемом случае выражается следующим образом:

$$\frac{\partial H_p}{\partial u_p} = \left(\frac{\partial F_p}{\partial u_p}\right)^{\mathrm{T}} \lambda_{p+1} + \frac{\partial \psi_p}{\partial u_p} = 0. \tag{3.16}$$

Подставляя $F_p = F(x_p, u_p)$ из (3.3) и ψ_p из (3.8), это уравнение можно упростить:

$$Ru_p + \left[g\left(x_p\right)\right]^{\mathrm{T}} \lambda_{p+1} = 0. \tag{3.17}$$

Имея в виду, что матрица R является положительно-определенной (т.е. R^{-1} существует), можно выразить u_p из (3.17):

$$u_p = -R^{-1}\left[g\left(x_p\right)\right]^{\mathrm{T}} \lambda_{p+1}. \tag{3.18}$$

Из выражения (3.18) следует, что для вычисления управления u_p в момент времени t_p требуется значение сопряженной переменной λ_{p+1} для момента времени t_{p+1}, которое вычисляется методом обратной рекурсии, начиная с завершающего момента времени. В рассматриваемой ADP-схеме сопряженное уравнение имеет следующий рекурсивный вид:

$$\lambda_p = \left(\frac{\partial \psi_p}{\partial x_p}\right) + \left(\frac{\partial F}{\partial x_p}\right)^{\mathrm{T}} \lambda_{p+1}. \tag{3.19}$$

Подставляя в (3.19) значение $F_p = F(x_p, u_p)$ из (3.3) и y_p из (3.8), получим более простую форму данного соотношения:

$$\lambda_p = \Delta t\left[\left(\frac{\partial h\left(x_p\right)}{\partial x_p}\right)^{\mathrm{T}} Qh\left(x_l\right)\right] + \left[\frac{\partial F}{\partial x_p}\right]^{\mathrm{T}} \lambda_{p+1}. \tag{3.20}$$

При синтезе управления на основе SNAC-подхода взаимосвязь между состоянием динамической системы x_p и сопряженным состоянием λ_{p+1} воспроизводится с помощью нейронной

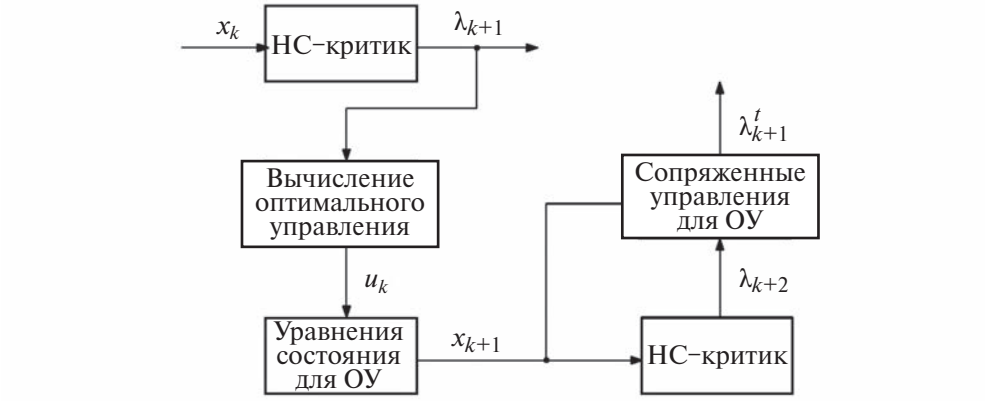

Рис. 3. Схема обучения сети НС-критика при SNAC-подходе к управлению движением.

сети (НС) прямого распространения (НС-критик в терминологии ACD-подхода). Структурная схема, показанная на рис. 3, иллюстрирует процедуру получения обучающих данных для обучения НС-критика. Частично обученный НС-критик выдает вектор сопряженного состояния λ_{p+1} в качестве выхода при векторе состояния x_p в качестве входа. Затем найденный вектор λ_{p+1} подставляется в соотношение для вычисления оптимального управления (3.18) для получения оптимального вектора управления u_p. Вектор состояния x_p и вектор управления u_p затем подставляются в уравнения состояния и сопряженного состояния, заданные соотношениями (3.3) и (3.20) соответственно, чтобы найти целевые значения для вектора сопряженного состояния λ^t_{p+1}. Затем НС-критик обучается, используя данные о состояниях x_p, сопряженных состояниях λ_{p+1} и решениях u_p, полученных путем решения задачи оптимального управления. Обученная сеть предсказывает оптимальное значение λ_{p+1} для заданного x_p. Это значение λ_{p+1} служит основой для вычисления текущего управления u_p для момента времени t_p.

Решение с помощью SNAC-подхода задачи управления для нелинейной динамической системы является достаточно трудоемким процессом. Следует отметить, что в этом случае все равно приходится прибегать к линеаризации исходной нелинейной системы для того, чтобы провести предварительное обучение сети в условиях, близких к оптимальным. Использование линеаризованной системы значительно упрощает задачу предобучения SNAC-регулятора. Такой регулятор обладает достаточно высокой эффективностью до тех пор, пока свойства объекта управления и условия его функционирования соответствуют условиям, при которых производился синтез закона управления.

Для линейного варианта объекта управления (2.1) с использованием методов теории оптимального управления [36, 37] зависимость между значением состояний и сопряженной переменной в момент времени $p+1$ принимает следующий вид:

$$\lambda_{p+1} = Sx_p, \tag{3.21}$$

где S – решение матричного уравнения Рикатти. Соответственно соотношение для оптимального управления для случая линейной системы и квадратичного критерия выражается следующим образом:

$$u_p = -R^{-1}B^{\mathrm{T}}Sx_p \tag{3.22}$$

или с учетом (3.21):

$$u_p = -R^{-1}B^{\mathrm{T}}\lambda_{p+1}. \tag{3.23}$$

Отсюда видно, что для получения оптимального управления, как и в случае с задачей построения линейно-квадратичного регулятора, достаточно решить матричное уравнение Рикатти.

4. Адаптивное управление сверхзвуковым пассажирским самолетом. *4.1. Динамика объекта управления.* В качестве объекта управления используется прототип сверхзвукового пассажирского самолета (СПС) второго поколения [35, 38]. Для данного самолета формируется модель продольного движения, используемая при решении задачи синтеза закона управления.

Как и другие СПС [39], рассматриваемый самолет имеет треугольное крыло малого удлинения с большой стреловидностью по передней кромке [40]. Такое крыло позволяет уменьшить рост аэродинамического сопротивления на трансзвуковых скоростях, но при этом существенно уменьшается и значение производной нормальной силы по углу атаки по сравнению с крылом дозвуковых самолетов. Вследствие этого заход на посадку СПС должен осуществлять на больших углах атаки, чтобы обеспечить необходимую подъемную силу. Поскольку такие значения угла атаки близки к критическим, возникает опасность сваливания самолета при больших возмущениях по углу атаки, вызванных, например, порывом ветра. В связи с этим требуется оперативно подавлять возникающие возмущения и возвращать самолет к сбалансированному состоянию. Исходя из этой особенности СПС, будут рассматриваться именно посадочные режимы полета для него. Один из таких режимов определяется условиями, приведенными в табл. 1. В этой таблице обозначено: m – масса СПС, h – высота полета, V – воздушная скорость СПС, α – угол атаки, Θ – угол наклона траектории, $\delta_в$ – угол отклонения руля высоты.

Таблица 1. Параметры полета СПС в сбалансированном режиме

m, кг	h, м	V, км/ч	α, град	Θ, град	$\delta_в$, град
75000	400	305.7	10.12	0	–3.6

Для улучшения динамических свойств СПС как объекта управления может быть введена стабилизирующая обратная связь по углу атаки, однако даже после этого самолет все равно остается неустойчивым в долгосрочной перспективе.

Исходная модель движения СПС является нелинейной. Для оценки устойчивости его движения будем использовать линеаризованную модель (2.1) продольного движения СПС для интересующего нас режима полета.

Конкретизация модели (2.1), согласно [30, 31], применительно к случаю продольного движения СПС позволяет получить следующую систему линейных дифференциальных уравнений:

$$\begin{bmatrix} \dot V_x \\ \dot V_y \\ \dot\omega_z \\ \dot\vartheta \end{bmatrix} = \begin{bmatrix} x_{V_x} & x_{V_y} & x_{\omega_z} & x_\vartheta \\ y_{V_x} & y_{V_y} & y_{\omega_z} & y_\vartheta \\ m_{V_x} & m_{V_y} & m_{\omega_z} & m_\vartheta \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_x \\ V_y \\ \omega_z \\ \vartheta \end{bmatrix} + \begin{bmatrix} x_{\delta_в} \\ y_{\delta_в} \\ m_{\delta_в} \\ 0 \end{bmatrix} [\delta_в], \tag{4.1}$$

матрицу C в (2.1) для рассматриваемого примера примем единичной.

В уравнении (4.1) V_x и V_y – проекции воздушной скорости на оси Ox и Oy связанной системы координат, ϑ – угол тангажа, ω_z – угловая скорость тангажа. Коэффициенты x_{V_x}, x_{V_y}, x_{ω_z}, x_ϑ, y_{V_x}, y_{V_y}, y_{ω_z}, y_ϑ, m_{V_x}, m_{V_y}, m_{ω_z}, m_ϑ представляют собой сокращенные обозначения для полных производных продольной и нормальной сил, а также момента тангажа по переменным V_x, V_y, w_z и ϑ. Аналогично коэффициенты $x_{\delta_в}$, $y_{\delta_в}$ и $m_{\delta_в}$ есть полные производные продольной силы, нормальной силы и момента тангажа по переменной $\delta_в$.

Для режима полета СПС, удовлетворяющего условиям, представленным в табл. 1, уравнение (4.2) принимает вид:

$$\begin{bmatrix} \dot V_x \\ \dot V_y \\ \dot\omega_z \\ \dot\vartheta \end{bmatrix} = \begin{bmatrix} 0.057 & 0.2421 & -0.0068 & -0.4779 \\ -0.1609 & -1.041 & 0.0866 & 1.3496 \\ 0.1528 & 1.0897 & -0.7309 & -1.2818 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_x \\ V_y \\ \omega_z \\ \vartheta \end{bmatrix} + \begin{bmatrix} -0.0581 \\ 0.1481 \\ -1.0246 \\ 0 \end{bmatrix} [\delta_в]. \tag{4.2}$$

В рассматриваемой задаче учитываются динамические свойства и ограничения привода органа управления, которым в нашем случае является руль высоты. С этой целью привод интерпретируется как нелинейная система первого порядка с постоянной времени 0.05 с, с ограничениями на реализуемый угол отклонения руля высоты (±25°) и на скорость его отклонения (±30 град/с).

Уравнению в пространстве состояний (4.3) можно поставить в соответствие передаточную функцию, получить для нее характеристическое уравнение и оценить устойчивость рассматриваемого объекта управления. Это характеристическое уравнение имеет следующий вид:

$$\left(s - 0.07387\right)\left(s + 0.0000031\right)\left(s^2 + 1.789s + 2.019\right) = 0. \tag{4.3}$$

Уравнение (4.3) имеет два действительных и пару комплексно-сопряженных корней. Оценив значения корней характеристического уравнения, можно сделать вывод, что из-за наличия положительных корней самолет является неустойчивым в длиннопериодическом движении. Отсюда следует, что при синтезе регулятора нам необходимо добиться с его помощью устойчивости движения самолета, кроме того, данный регулятор должен обладать способностью работать в условиях неточного знания динамики объекта управления.

4.2. Сочетание динамической инверсии и SNAC-подхода. Динамическая инверсия в нашем примере используется в продольном канале управления полетом самолета для обеспечения решения задачи стабилизации угловой скорости тангажа. Вначале необходимо сформировать параметры динамической инверсии для объекта управления в его номинальном, т.е. исходном, состоянии. Для этого, согласно выражениям (2.15) и (2.16), получим значения весовых коэффициентов линейной сети (4.2) для рассматриваемого объекта:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} = \begin{bmatrix} -\dfrac{a_{31}}{b_{31}} & -\dfrac{a_{32}}{b_{31}} & -\dfrac{a_{33}}{b_{31}} & -\dfrac{a_{34}}{b_{31}} \end{bmatrix}^T = \begin{bmatrix} 0.149 \\ 1.064 \\ -0.713 \\ 1.251 \end{bmatrix}. \tag{4.4}$$

Алгоритм SNAC используется для формирования управляющего сигнала во внешнем контуре (рис. 4).

С помощью алгоритма SNAC были обучены четыре сети критика со структурой $4-12-1$, т.е. с четырьмя входами, одним выходом и 12 нейронами в единственном скрытом слое. На вход каждой сети подаются все компоненты вектора состояния системы x в момент времени p, выходом же служит значение одной из компонент сопряженного вектора λ_i, $i=\overline{1,4}$, в момент времени $p+1$.

Функцией активации нейронов скрытого слоя во всех сетях является гиперболический тангенс. Выходной слой содержит единственный нейрон с линейной функцией активации. Сети критика выполняют адаптацию системы управления путем корректировки параметров этих сетей непосредственно в полете. При решении задачи в ходе проводившихся экспериментов были выбраны следующие весовые коэффициенты в критерии (3.7) с учетом принятого допущения о равенстве выходного вектора вектору состояний:

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 100 & 0 \end{bmatrix}, R = 2. \tag{4.5}$$

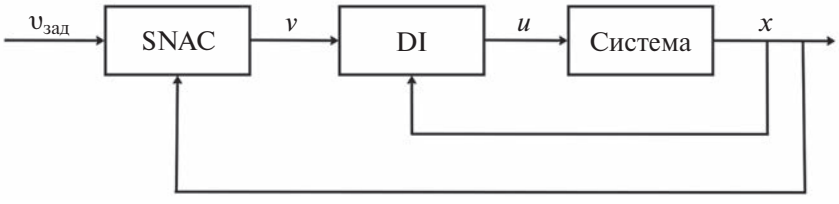

Рис. 4. Схема совместной работы SNAC и DI ($\vartheta_{зад}$ – заданное значение угла тангажа).

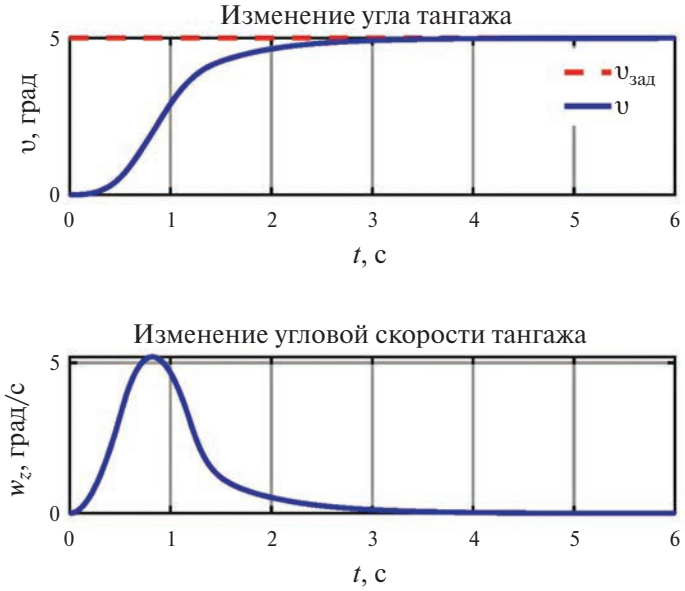

Рис. 5. Отработка заданного угла тангажа, равного 5°, при совместном использовании SNAC и DI.

Такой выбор весовых коэффициентов обусловлен характером реализуемых переходных процессов. А именно, при малых значениях параметра R увеличиваются потребные значения коэффициентов регулятора, что приводит к быстрой раскачке самолета при больших скоростных напорах. С другой стороны, чем больше значения параметра R, тем более затянутым получается переходный процесс. Используемое в рассматриваемом примере значение $R=2$ является компромиссным, подобранным экспериментально. Весовые коэффициенты в матрице Q определяют относительную важность подавления возмущений по соответствующим переменным состояния $x_j, j=\overline{1,4}$. Так как ставится задача стабилизации угловой скорости тангажа ($x_3=\omega_z$), соответствующий вес, являющийся элементом q_{33} матрицы Q, выбран значительно превосходящим значения остальных элементов этой матрицы.

После синтеза системы управления было проведено несколько экспериментов для оценки ее работоспособности. Решалась задача отслеживания задающего сигнала по тангажу, результаты этих экспериментов представлены на рис. 5 и 6. Решалась также важная задача стабилизации угла атаки на посадочном режиме, результаты приведены на рис. 7. Начальные возмущенные значения угла атаки, отличающиеся от балансировочного значения, в экспериментах задавались так, как показано в табл. 2. Помимо этого, оценивалась возможность подобной схемы парировать отказы за счет адаптации сетей, используемых в динамической инверсии и алгоритме SNAC. На рис. 8 демонстрируются результаты моделирования системы при имитации отказа, который привел к уменьшению на 60% эффективности органа управления в продольном канале.

Полученные результаты позволяют сделать следующие выводы. Совместное применение динамической инверсии и SNAC-технологии позволило успешно решить поставленную задачу синтеза отказоустойчивого регулятора угловой скорости тангажа для СПС. При этом использование динамической инверсии упрощает процесс настройки SNAC. Что касается воздействия комбинации DI+SNAC на значения переменных состояния в проводившихся экспериментах, то здесь можно отметить следующее. При отработке задающего сигнала по углу тангажа (переход из состояния $\vartheta=0$ в состояние $\vartheta=5°$) выход на требуемое значение происходит менее чем за 3 с, перерегулирование отсутствует. Для исходных данных, приведенных в табл. 1, была решена также задача перехода СПС к балансировочному значению угла атаки, равному $\alpha=10.12°$. Регулятор, сформированный по схеме DI + SNAC, успешно справляется и с этой задачей. Колебания по углу атаки подавляются, время выхода на балансировочное значение не превышает 3 с, хотя начальные отклонения от этого значения в ряде экспериментов были значительными, как это видно из табл. 1. Лишь в эксперименте с самым большим отклонением от балансировочного значения появляется небольшая колебательность

Таблица 2. Значения возмущенного угла атаки в проводимых экспериментах

Угол атаки	Варианты				
	1	2	3	4	5
α, град	13	15	7	4	1

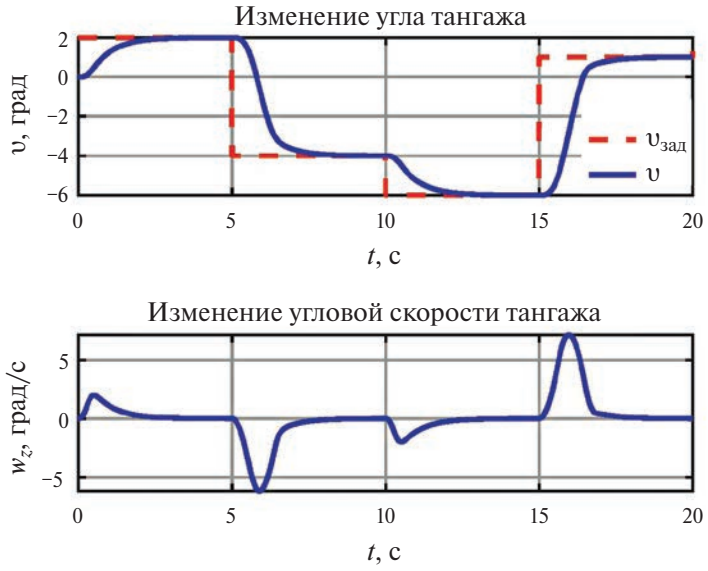

Рис. 6. Отработка многоступенчатого задающего сигнала по углу тангажа при совместном использовании SNAC и DI.

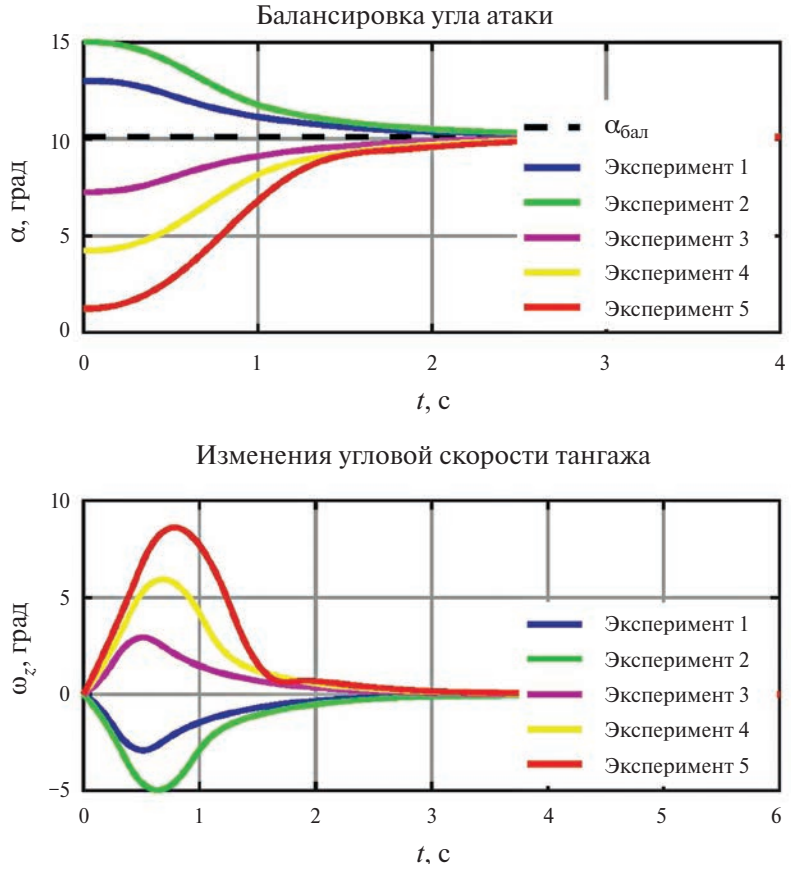

Рис. 7. Стабилизация балансировочного угла атаки при совместном использовании SNAC и DI ($\alpha_{бал}$ – балансировочное значение угла атаки).

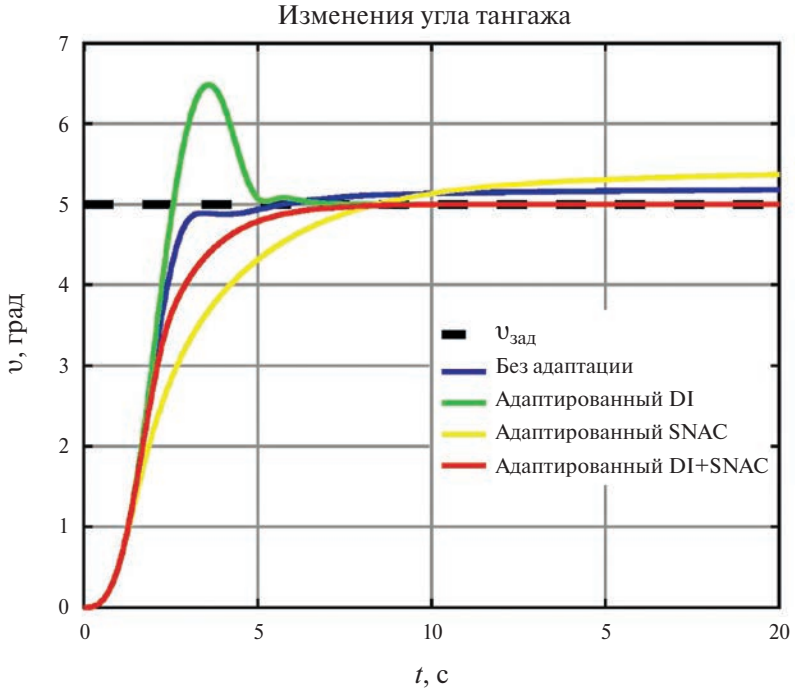

Рис. 8. Сравнение различных вариантов схемы DI + SNAC при моделировании отказа в системе (v – вспомогательный входной сигнал согласно сооотношению (2.5)).

по угловой скорости тангажа. В целом полученные результаты являются вполне приемлемыми для ЛА рассматриваемого вида.

Основное внимание стоит обратить на эксперимент, где производилось сравнение различных вариантов схемы DI + SNAC, с адаптацией и без нее. Как уже было сказано ранее, адаптация динамической инверсии выполнялась с помощью идентификации объекта управления в ходе полета. Адаптация SNAC производилась путем пересчета значений коэффициентов оптимального закона управления. Эта операция была необходима в случае, когда производилась подстройка параметров динамической инверсии. Полученные результаты показывают, что для точного отслеживания задающего сигнала по тангажу такая подстройка необходима. Без нее расхождение между требуемым и реализуемым значениями угла тангажа может доходить до 10% за минуту переходного процесса. Если же подстройка выполняется только для динамической инверсии без адаптации SNAC, то возникает значительное перерегулирование по тангажу, доходящее до 30%. Происходит это из-за того, что для новой замкнутой системы закон управления, реализуемый SNAC, перестает быть оптимальным.

Заключение. Современные ЛА должны обладать способностью эффективно решать поставленные задачи в широком диапазоне условий. С этой целью разрабатываются законы управления, основанные на современной теории адаптивного и оптимального управления. Наличие у таких систем свойства адаптивности позволяет им работать в условиях неопределенности, порожденных неполным и неточным знанием свойств объекта управления, а также изменением динамических свойств этого объекта вследствие отказов оборудования и повреждения конструкции. Инструментарий приближенного динамического программирования и основанного на нем обучения с подкреплением в сочетании с нейронными сетями, позволяет эффективно решать данную проблему, т.е. формировать адаптивные и одновременно оптимальные законы управления. Одно из применений данного подхода связано с управлением движением ЛА. Мы рассматривали уже эту тему в [22, 33, 34], в данной статье расширяем состав используемого инструментария за счет привлечения метода динамической инверсии.

Полученные результаты позволяют сделать вывод о перспективности предлагаемого подхода. Одновременно они демонстрируют необходимость вовлечения средств нелинейной динамической инверсии для обеспечения работы с нелинейными системами. Так как NDI-подход

требует наличия точной модели объекта управления и полной наблюдаемости компонент его вектора состояния, актуальным является решение проблемы корректировки NDI в ситуации, когда динамика объекта управления перестала соответствовать его модели. Решение этой проблемы открывает возможность решения гораздо более сложных задач управления, чем это обеспечивается средствами, существующими в настоящий момент.

СПИСОК ЛИТЕРАТУРЫ

1. *Powell W.B.* Approximate Dynamic Programming: Solving the Curse of Dimensionality. 2nd Ed. Wiley, 2011. 658 p.
2. *Werbos P.J.* Approximate Dynamic Programming for Real-time Control and Neural Modeling // Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches, Van Nostrand Reinhold / Eds D.A. White, D.A. Sofge. N.Y. USA, 1992. P. 493–525.
3. *Lewis F.L., Vrabie D.* Reinforcement Learning and Adaptive Dynamic Programming for Feedback Control // IEEE Circuits and Systems Magazine. 2009. V. 9. № 3. P. 32–50.
4. Reinforcement Learning and Approximate Dynamic Programming for Feedback Control / Eds F.L. Lewis, D. Liu. Wiley, 2013. 634 p.
5. *Liu D., Xue S., Zhao B., Luo B., Wei Q.* Adaptive Dynamic Programming for Control: A Survey and Recent Advances // IEEE Trans. on Systems, Man, and Cybernetics. 2021. V. 51. № 1. P. 142–160.
6. *Wei Q., Song R., Li B., Lin X.* Self-learning Optimal Control of Nonlinear Systems: Adaptive Dynamic Programming Approach. Springer, 2018. 240 p.
7. *Song R., Wei Q., Li Q.* Adaptive Dynamic Programming: Single and Multiple Controllers. Springer, 2019. 278 p.
8. *Liu D., Wei Q., Wang D., Yang X., Li H.* Adaptive Dynamic Programming with Applications in Optimal Control. Springer, 2017. 609 p.
9. *Хайкин С.* Нейронные сети: полный курс. 2-е изд. М.: Вильямс, 2006. 1106 с.
10. *Werbos P.J.* A Menu of Designs for Reinforcement Learning over Time // Neural Networks for Control / Eds W.T.Miller, R.S.Sutton, P.J.Werbos. Cambridge, MA: MIT Press, 1990. P. 67–95.
11. *Ferrari S., Stengel R.F.* Online Adaptive Critic Flight Control // J. Guidance, Control, and Dynamics. 2004. V. 27. № 5. P. 777–786.
12. *Vamvoudakis K.G., Lewis F.L.* Online Actor-critic Algorithm to Solve the Continuous-Time Infinite Horizon Optimal Control Problem // Automatica. 2010. V. 46. P. 878–888.
13. *Wang D., He H., Liu D.* Adaptive Critic Nonlinear Robust Control: A Survey // IEEE Trans. Cybern. 2017. V. 47. № 10. P. 1–22.
14. *Wang D., Mu C.* Adaptive Critic Control with Robust Stabilization for Uncertain Nonlinear Systems. Springer Nature, 2019. 317 p.
15. *Wang D., Ha M, Zhao M.* Advanced Optimal Control and Applications Involving Critic Intelligence. Springer Nature, 2023. 283 p.
16. *Padhi R., Unikrishnan N., Wang X., Balakrishnan S.N.* A Single Network Adaptive Critic (SNAC) Architecture for Optimal Control Synthesis for a Class of Nonlinear Systems // Neural Networks. 2006. V. 19. P. 1648–1660.
17. *Steck J.E., Lakshmikanth G.S., Watkins J.M.* Adaptive Critic Optimization of Dynamic Inverse Control // AIAA Infotech and Aerospace Conf. Garden Grove, California, USA. Preprint 2012–2408. 21 p.
18. *Lakshmikanth G.S., Padhi R., Watkins J.M., Steck J.E.* Single Network Adaptive Critic Aided Dynamic Inversion for Optimal Regulation and Command Tracking with Online Adaptation for Enhanced Robustness // Optimal Control Applications and Methods. 2014. V. 35. P. 479–500.
19. *Lakshmikanth G.S., Padhi R., Watkins J.M., Steck J.E.* Adaptive Flight-Control Design Using Neural-Network-Aided Optimal Nonlinear Dynamic Inversion // J. Aerospace Information Systems. 2014. V. 11. № 11. P. 785–806.
20. *Heyer S.* Reinforcement Learning for Flight Control: Learning to Fly the PH-LAB. MS Thesis. Deft, Netherlands: Delft University of Technology, 2019. 126 p.
21. *Teirlinck C.* Reinforcement Learning for Flight Control: Hybrid Offline-Online Learning for Robust and Adaptive Fault-Tolerance. MS Thesis. Deft, Netherlands: Delft University of Technology, 2022. 153 p.
22. *Tiumentsev Yu.V., Tshay R.A.* SNAC Approach to Aircraft Motion Control // Studies in Computational Intelligence. 2023. V. 1120. P. 420–434.
23. *Enns D., Bugajski D., Hendrick R., Stein G.* Dynamic Inversion: An Evolving Methodology for Flight Control Design // Intern. J. Control. 1994. V. 59. № 1. P. 71–91.
24. *Looye G.* Design of Robust Autopilot Control Laws with Nonlinear Dynamic Inversion // Automatisierungstechnik. 2001. V. 49. № 12. P. 523–531.
25. *Lombaerts T.J.J., Huisman H.O., Chu Q.P., Mulder J.A., Joosten D.A.* Nonlinear Reconfiguring Flight Control Based on Online Physical Model Identification // J. of Guidance, Control, and Dynamics, 2009. V. 32. № 3. P. 727–748.
26. *Lombaerts T.J.J., Looye G.H.N.* Design and Flight Testing of Nonlinear Auto Flight Control Laws // AIAA Guidance, Navigation and Control Conf. Minneapolis, Minnesota, USA. Preprint 2012–4982. 24 p.
27. *Горбань А.Н.* Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей // Сиб. журн. вычисл. математики. 1998. Т. 1. № 1. С.11–24.

28. *Горбань А.Н., Дунин-Барковский, Кирдин А.Н. и др.* Нейроинформатика. Новосибирск: Наука, 1998. 296 с.
29. *Шибзухов З.М.* Некоторые вопросы теоретической нейроинформатики // XIII Всеросс. науч.-техн. конф. “Нейроинформатика-2011”, Школа-семинар “Соврем. проблемы нейроинформатики”. М.: Изд-во МИФИ, 2011. С. 1–30.
30. *Cook M.V.* Flight Dynamics Principles. 2nd Ed. Elsevier, 2007. 496 p.
31. *Stevens B.L., Lewis F.L., Johnson E.N.* Aircraft Control and Simulation: Dynamics, Controls Design and Autonomous Systems. 3rd Ed. Wiley, 2016. 764 p.
32. *Sutton R.S., Barto A.G.* Reinforcement Learning: An Introduction. 2nd Ed. Cambridge, Massachusetts, USA: The MIT Press, 2018. 548 p.
33. *Chulin M.I., Tiumentsev Yu.V., Zarubin R.A.* LQR Approach to Aircraft Control Based on the Adaptive Critic Design // Studies in Computational Intelligence. 2023. V. 1120. P. 406–419.
34. *Tiumentsev Yu.V., Zarubin R.A.* Lateral Motion Control of a Maneuverable Aircraft Using Reinforcement Learning // Optical Memory and Neural Networks. 2024. V. 33. № 1. P. 1–12.
35. *Prodanik V.A., Efremov A.V.* Synthesis of a Controller Based on the Principle of Inverse Dynamics and the Online Identification of a Lateral Motion Model in a Next-Generation Supersonic Transport // Recent Developments in High-Speed Transport / Eds D.Y.Strelets, O.N. Korsun. Springer, 2023. P. 41–49.
36. *Lewis F.L., Vrabie D.L., Syrmos V.L.* Optimal Control. 3rd Ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2012. 550 p.
37. *Bryson A.E., Ho Y.-C.* Applied Optimal Control: Optimization, Estimation and Control. N.Y: Taylor & Francis Group, 1975. 496 p.
38. *Grishina A.Y., Efremov A.V.* Development of a Controller Law for a Supersonic Transport Using Alternative Means of Automation in the Landing Phase // Recent Developments in High-Speed Transport / Eds D.Y. Strelets, O.N. Korsun. Springer, 2023. P. 41–49.
39. *Webb B.D., Takahashi T.T.* Emerging Federal Regulatory Framework for Future Supersonic Transport Aircraft // AIAA SCITECH Forum, San Diego, California, USA. Preprint 2022–0366. 23 p.
40. *Ericsson L., Reding J.* Unsteady Aerodynamics of Slender Delta Wings at Large Angles of Attack // J. Aircraft. 1975. V. 12. № 9. P. 721–729.