РОССИЙСКАЯ АКАДЕМИЯ НАУК
Институт проблем управления
им. В.А. Трапезникова

Д.А. Новиков, И.М. Смирнов,

Т.Е. Шохина

МЕХАНИЗМЫ УПРАВЛЕНИЯ ДИНАМИЧЕСКИМИ АКТИВНЫМИ СИСТЕМАМИ

Москва – 2002

УДК 007
ББК 32.81
Н 73

Новиков Д.А., Смирнов И.М., Шохина Т.Е. **Механизмы управления динамическими активными системами**. М.: ИПУ РАН, 2002. – 124 с.

Настоящая работа содержит результаты исследований теоретико-игровых моделей динамических активных систем (ДАС). Приводится обзор известных результатов, вводится система классификаций моделей ДАС, формулируются и решаются задачи управления ДАС. Значительное внимание уделяется анализу сравнительной эффективности различных режимов управления, а также – влиянию дальновидности и обязательств на эффективность управления.

Работа рассчитана на специалистов (теоретиков и практиков) по управлению организационными системами.

Рецензент: д.т.н., проф. А.В. Щепкин

Утверждено к печати Редакционным советом Института

Текст воспроизводится в виде, утвержденном Редакционным советом Института

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Формальные (теоретико-игровые) модели организационных систем (*активных систем* – АС) исследуются в таких разделах теории управления социально-экономическими системами как: теория активных систем (ТАС) [4, 12-23, 50-60], теория иерархических игр (ТИИ) [30, 32, 41], теория контрактов (ТК) [15, 58, 125] и др. *Модель АС* определяется заданием следующих параметров [23]: *состав системы* (совокупность *участников системы* – управляющих органов (*центров*) и управляемых субъектов (*активных элементов* (АЭ)), различающихся правами принятия решений; *структура системы* – совокупность связей между участниками; *множества допустимых* стратегий участников (выбираемых ими в соответствии с собственными интересами[1] состояний, управлений и т.д.); *целевые функции*, зависящие в общем случае от стратегий всех участников и моделирующие их взаимодействие; *информированность* – та информация, которой обладают участники на момент принятия решений; *порядок функционирования* – последовательность получения участниками АС информации и выбора ими стратегий.

Задача управления формулируется следующим образом. Центр, обладающий правом первого хода, сообщает выбранное им управление активным элементам, которые при известном управлении центра выбирают собственные стратегии с целью максимизации своих целевых функций. Цель центра заключается в том, чтобы, зная реакцию управляемых субъектов на те или иные управления, выбрать такое управление, которое привело бы систему в наиболее предпочтительное с его точки зрения состояние.

Частным случаем задачи управления является *задача стимулирования* [35, 42, 58, 81, 86, 130], в которой центр осуществляет

---

[1] *Характерной особенностью теоретико-игровых моделей является учет активности участников АС – способности к целенаправленному поведению (моделируемому возможностью самостоятельного выбора ими стратегий) в соответствии с собственными предпочтениями и интересами (отражаемыми функциями выигрыша участников, которые они стремятся максимизировать).*

побочные платежи[1] управляемым субъектам, зависящие от выборов (действий) последних.

Простейшей (*базовой*) моделью АС является *двухуровневая система* [18, 59], состоящая из одного центра (*АС с унитарным контролем* [60]) и одного АЭ (*одноэлементная АС*), принимающих решения однократно (*статическая АС* [52]) и в условиях полной информированности (*детерминированная АС* [58]). Расширениями базовой модели являются многоуровневые АС [53], АС с распределенным контролем [60], многоэлементные АС [59], динамические АС [52], АС с неопределенностью [58].

Предметом исследования в настоящей работе являются механизмы управления *динамическими активными системами* (ДАС), то есть АС, в которых последовательность выбора стратегий, характерная для статических АС, повторяется как минимум несколько раз[2].

Интуитивно понятно, что при таком естественном обобщении простейшей базовой (статической) модели, как рассмотрение нескольких несвязанных периодов функционирования, задачу управления удается декомпозировать на набор базовых задач. Трудности появляются при исследовании систем со связанными периодами функционирования. Методы и алгоритмы решения задачи синтеза оптимального механизма управления в этом случае характеризуются высокой структурной и вычислительной сложностью. Как правило, универсального подхода к аналитическому решению этого класса задач найти не удается. Однако, преодоление трудностей анализа оправданно, так как в динамических АС присутствуют новые качественные свойства, отсутствующие в базовой модели (не говоря уже о том, что большинство реальных организационных

---

[1] *В терминах теории иерархических игр задаче стимулирования соответствует игра Γ_2 с побочными платежами, в которой целевые функции обоих игроков зависят только от стратегии второго игрока* [30, 60].

[2] *Отметим, что в настоящей работе не рассматриваются пассивные динамические системы* [28, 48, 54] *и дифференциальные игры* [1, 27, 38, 39, 63, 66]. *Но, в том числе, упомянутым образом может описываться процесс «схождения» АЭ к равновесию в процессе их игры при фиксированном управлении* [8, 24, 50, 51, 61, 74] – *см.* [5, 25, 29, 45, 51]. *Кроме этого, следует упомянуть широкую распространенность динамических моделей в математической экономике* [7, 44, 47, 49, 69, 71, 116, 117, 125].

систем функционируют в течении продолжительного времени и характеризуются относительной повторяемостью условий и самих фактов принятия решений). ДАС, функционирующие в течение длительного времени, существенно отличаются от статических: возможность долговременного сотрудничества, адаптации, пересмотра стратегий – все эти эффекты проявляются при переходе от статических моделей к динамическим.

Изучению задач управления динамическими АС посвящено значительное число исследований (в Приложении к настоящей работе приведен обзор основных результатов исследования задач управления ДАС).

В настоящей работе акцент сделан на задачи стимулирования, являющиеся, как отмечалось выше, частным случаем задачи управления. В качестве обоснования выбора предмета исследования следует подчеркнуть, что задачи стимулирования представляют собой самостоятельный, достаточно обширный и разнообразный класс задач управления, имеющих как хорошие содержательные интерпретации, так и отражающих потребность в практическом использовании теоретических результатов. Кроме того, многие расширения базовой задачи стимулирования исследованы относительно подробно, что позволяет адаптированно переносить ряд известных результатов на относительно малоизученные ДАС.

В частности, в настоящей работе широко используются следующие подходы и результаты. Известный из анализа базовой задачи стимулирования [42, 56, 57] метод анализа множеств реализуемых действий и минимальных затрат на стимулирование оказывается эффективным и в динамических моделях, так как формулируемый на его основе *принцип компенсации затрат* является эффективным инструментом решения задач стимулирования, в частности, позволяющим не акцентировать внимание на исследовании согласованности стимулирования. В многоэлементных АС (в том числе – динамических) применения одного принципа компенсации затрат оказывается недостаточно, так как имеет место игра управляемых активных элементов. В этом случае целесообразно использование *принципа декомпозиции игры* АЭ, в соответствии с которым может быть построено управление со стороны центра, декомпозирующее взаимодействие управляемых субъектов и позволяющее рассматривать задачи согласованного управления каждым АЭ независимо, перенося учет их взаимодействия на этап согласованного планирования [18].

Применение принципа компенсации затрат и принципа декомпозиции игры АЭ [59] позволяет получать аналитические решения задач определения согласованного (побуждающего АЭ выбирать действия, совпадающие с назначаемыми центром планами) управления. Это управление параметрически зависит от *планов* (предпочтительных с точки зрения центра состояний АЭ), следовательно, необходимо формулировать и решать задачу согласованного планирования – определения оптимальных значений планов. На этом этапе эффективным оказывается применение методов динамического программирования, оптимального управления и др. [9-11, 26, 68]. Кроме того, в зависимости от специфики рассматриваемой задачи, ниже широко используются известные результаты исследования многоуровневых АС [53], АС с распределенным контролем [60] и АС с неопределенностью [58].

Как отмечалось выше, характерной чертой ДАС является адаптивность, проявляющаяся, в первую очередь, в возможности участников системы накапливать информацию и корректировать свое поведение с учетом повышения информированности за счет наблюдаемой истории их взаимодействия между собой и с окружающей средой. Роль *неопределенности* в ДАС заслуживает отдельного обсуждения.

В [23, 58] предложена система классификаций, в соответствии с которой выделялась внешняя (относительно параметров окружающей среды) и внутренняя (относительно параметров, характеризующих участников рассматриваемой АС) неопределенность, понимаемая как неполная информированность участников АС о существенных (для процессов принятия ими решений) параметрах. В зависимости от той информации, которой обладает субъект, можно выделять интервальную, вероятностную и нечеткую неопределенность. В соответствии с принципом детерминизма [23, 56, 58], принятие решения [65, 70] осуществляется в условиях полной информированности, поэтому окончательный критерий, которым руководствуется субъект, не должен содержать неопределенных параметров, причем устранение неопределенности должно производиться с учетом всей имеющейся на рассматриваемый момент информации.

Опыт исследования теоретико-игровых моделей механизмов управления АС с неопределенностью [57, 58] свидетельствует, что эффективность управления не возрастает с ростом неопределенности и, соответственно, не убывает с ее уменьшением (точнее – с ростом информированности управляющего органа).

При рассмотрении математических моделей динамических активных систем необходимо различать неопределенности следующих типов[1] (основание классификации – моменты времени, относительно которых у лица, принимающего решение (ЛПР), имеется недостаточная информация):

- *текущая неопределенность*;
- *неопределенность будущего*,

каждая из которых может подразделяться (основание классификации – объекты, относительно которых имеется недостаточная информация) на *объективную неопределенность* (неполная информированность относительно внешних и/или внутренних параметров ЛПР или других субъектов) и *субъективную неопределенность* (неполную информированность ЛПР о поведении других субъектов, входящих в рассматриваемую систему). Последний тип неопределенности иногда называют *игровой неопределенностью*[2].

Традиционно под «неопределенностью» понимают текущую объективную неопределенность и большинство исследований АС с

---

[1] *В литературе описаны несколько классификаций неопределенностей* [58, 62]. *Например, в* [58] *предлагалось выделять «неопределенности природы»* (*факторы, которые неизвестны лицу, принимающему решение, и/или исследователю операций*), *«неопределенности противника»* (*отражающие невозможность полного учета и предсказания действий других активных участников системы*) *и «неопределенности целей»* (*отражающие многокритериальность задач принятия решений*).

[2] *Субъективная* (*игровая*) *неопределенность, как правило, устраняется введением тех или иных предположений о принципах поведения участников системы, позволяющих однозначно доопределить выбираемые ими из множества решений игры стратегии* (*то есть устранение субъективной неопределенности производится в два этапа – на первом этапе определяется концепция равновесия* (*максиминное равновесие, равновесие Нэша, Байеса и т.д.* [107, 128]), *на втором этапе определяется принцип выбора игроками конкретных равновесных стратегий в случае, если последних несколько* (*гипотеза благожелательности, принцип гарантированного результата и т.д.* [56, 58]).

неопределенностью учитывает именно ее [58]. В то же время, для ДАС характерна не только текущая объективная неопределенность, но и неопределенность будущего, которая заключается в том, что, принимая решение, ЛПР, с одной стороны, «влияет» на будущее (это влияние может проявляться в изменении множеств его будущих допустимых действий, выигрышей и т.д. – см. модели ниже), а, с другой стороны, возможности его анализа этого влияния ограничены незнанием будущих значений[1] существенных параметров. Многочисленные примеры проявлений неопределенности будущего приведены ниже в настоящей работе.

Большинство рассматриваемых ниже моделей является «детерминированными» в оговоренном выше традиционном понимании – в них в основном отсутствует неполная информированность участников ДАС друг о друге, об окружающей среде и т.д. В то же время, неопределенность будущего присутствует, естественно, в полной мере.

Изложение материала настоящей работы имеет следующую структуру. В первом разделе вводится система классификаций задач управления динамическими активными системами. Во втором разделе обсуждаются и классифицируются возможности участников АС по учету будущего и взаимосвязь этих возможностей с режимами управления. В третьей части исследуются задачи стимулирования в ДАС. Их описание ведется индуктивно – от простейшей одноэлементной ДАС с несвязанными периодами функционирования к многоэлементной ДАС со связанными периодами функционирования. Далее рассматриваются двух и трехпериодные ДАС (раздел 4), а также эффекты накопления в ДАС (раздел 5). Заключение содержит краткое перечисление основных результатов. В Приложение помещен обзор основных результатов теории активных систем, теории иерархических игр и теории контрактов по управлению ДАС.

---

[1] *Интересно отметить, что обычно (при описании текущей неопределенности) предполагается, что субъект всегда информирован о собственных параметрах (множествах допустимых действий, предпочтениях и т.д.) лучше, чем другие субъекты (будь то другие участники рассматриваемой системы или исследователь операций). В случае неопределенности будущего субъект может иметь неполную информацию о своих собственных параметрах.*

1. КЛАССИФИКАЦИЯ ЗАДАЧ УПРАВЛЕНИЯ ДИНАМИЧЕСКИМИ АКТИВНЫМИ СИСТЕМАМИ

Перечисленные во введении параметры АС являются основой для системы классификаций ДАС[1]. Предлагаются следующие основания системы классификаций ДАС.

1. Наличие или отсутствие неопределенности относительно существенных параметров функционирования АС. Если участники АС принимают решения в условиях полной информированности, то такая АС называется детерминированной. Если хотя бы один из участников[2] не обладает всей значимой в рамках модели информацией, то соответствующая АС называется АС с неопределенностью.

АС с неопределенностью в свою очередь подразделяются на классы в зависимости от типа (1.1 – внутренняя неопределенность – относительно параметров самой АС, 1.2 – внешняя неопределенность – относительно параметров окружающей среды) и вида (1.1.1 (1.2.1) – интервальная неопределенность – известно только множество возможных значений неопределенного параметра; 1.1.2 (1.2.2) вероятностная – дополнительно (помимо допустимого множества) известно вероятностное распределение; 1.1.3 (1.2.3) – нечеткая – когда дополнительно известна функция принадлежности) неопределенности. Может иметь место также смешанная неопределенность – одновременно нескольких типов или видов.

---

[1] *В* [23, 59, 60] *вводились соответственно системы классификаций общих задач управления АС, задач управления многоэлементными АС, АС с распределенным контролем и т.д. Предлагаемая в настоящем разделе система классификаций пересекается с ними по ряду общих для всех АС оснований, но в основном отражает специфику именно динамических АС.*

[2] *Если информированность всех участников одинаковая, то говорят, что имеет место симметричная информированность. Если участники АС обладают разной информацией, то считается, что имеет место асимметричная информированность* [58]. *Понятно, что, если информированность асимметричная, то АС является АС с неопределенностью* (*но не наоборот, так как все участники могут обладать одинаковой неполной информацией*).

2. Параметры модели активной системы, зависящие в каждом периоде от параметров предыдущих периодов[1]. Так как задача стимулирования в статической АС задается перечислением множеств допустимых стратегий центра (множество допустимых функций стимулирования) и АЭ (множество допустимых действий), а также их целевых функций (функция дохода центра и функция затрат АЭ, зависящие от действий последнего), то возможны следующие значения признаков классификации по данному основанию: 2.1 – связанное стимулирование, 2.2 – связанные затраты, 2.3 – связанный доход, 2.4 – связанные допустимые множества, а также все их возможные комбинации.

3. Распределение дальновидностей. Специфической характеристикой ДАС является возможность учета ее участниками будущих последствий принимаемых сегодня решений (*свойство дальновидности*)[2]. В первом приближении можно выделить ДАС с: 3.1 – дальновидными участниками, 3.2 – недальновидными участниками. Естественно, дальновидные участники могут по-разному учитывать будущие периоды (характеристика, отражающая способ учета будущего называется *распределением дальновидностей* [78-80]). В теории игр (в основном при рассмотрении повторяющихся игр – см. Приложение) и в экономике распределение дальновидностей описывается, как правило, дисконтирующими множителями. Остановимся на обсуждении их роли более подробно.

Пусть[3] w^t – выигрыш игрока в момент времени t, t Î $\{1, 2, ..., T\}$. В качестве целевой функции, определяемой для текущего момента времени t £ T, зависящей от выигрышей в текущем и во всех будущих периодах, принимается либо взвешенная сумма выигрышей по периодам[4]: $W(t) = \sum_{t=t}^{T} d^t(t)\, w^t$, либо средний

---

[1] *Ниже для обозначения такой зависимости употребляется термин «связанные параметры».*

[2] *Участники АС, ориентирующиеся при принятии решений только на текущее значение своего выигрыша (полезности, целевой функции и т.д.), называются недальновидными.*

[3] *Условимся, что верхние индексы обозначают временные характеристики (моменты времени, их диапазон и т.д.).*

[4] *Для простоты ограничимся моделями с дискретным временем:* $t = \overline{1,T}$.

по всем (T – $t + 1$) периодам взвешенный выигрыш: $W_T(t) = \dfrac{1}{T - t + 1} \sum_{t=t}^{T} d^{\,t}(t)\, w^t$.

Под распределением дальновидностей игрока будем понимать набор векторов $d^{\,t,T}(t) = (d^{\,t}(t), d^{\,t+1}(t), ..., d^{\,T}(t))$, $t = \overline{1,T}$.

В качестве отступления приведем содержательные интерпретации распределения дальновидностей.

Пусть в первый момент времени имеется актив, обладающий стоимостью W_0. Тогда, если присутствует инфляция, например, b процентов в единицу времени, то стоимость актива в момент времени t ³ 1 составит[1] $W(t) = W_0\,(d)^{\,t-1}$, где $d = 1-b$. Если инфляция отсутствует и имеется возможность, например, приобрести ценные бумаги, приносящие доход b процентов в месяц[2], то $W(t) = W_0\,(d)^{\,t-1}$, где $d = 1+b$. Если рассмотреть «обратную» задачу – определить текущий эквивалент W_0 актива $W(t)$, полученного в периоде t, то получим, что $W_0 = d\,W(t)\,(d)^{\,-t}$. Содержательно, в экономике распределение дальновидностей, отражаемое *дисконтирующим множителем* (коэффициентом дисконтирования) d Î $(0; 1]$, характеризует изменение предпочтений (в большинстве случаев – стоимости активов) во времени[3] – чем больше отсрочка в получении некоторого блага, тем меньше его полезность.

В повторяющихся играх существует другая содержательная интерпретация дисконтирующих множителей. Предположим, что игрокам предлагается повторяющаяся игра, в которой перед каждым периодом (то есть перед каждым повторением) определяется – продолжать игру дальше или нет. Пусть W^t – выигрыш некоторого

---

[1] *Для того, чтобы различать значение* $d^{\,t}$ *распределения дальновидностей* $d^{\,1,T}$ *от степени коэффициента дисконтирования* $(d)^t$, *в последнем случае будем использовать скобки.*

[2] *Имеются в виду сложные проценты. Если начисляются простые проценты, то* $W(t) = W_0\,(1 + b\,(t - 1))$.

[3] *Очевидно, что чем выше значение коэффициента дисконтирования (чем ближе он к единице), тем с большим весом учитываются будущие периоды – в пределе (при* $d = 1$*) все периоды учитываются одинаково, и наоборот – при стремлении* d *к нулю степень учета будущего уменьшается (у недальновидного игрока коэффициент дисконтирования равен нулю).*

игрока в периоде t, если игра в этом периоде состоится, а p^t – вероятность того, что состоится розыгрыш в периоде t (если розыгрыш в периоде t не состоялся, то игра заканчивается и последующие периоды не рассматриваются). Определим $d^t = \prod_{j=1}^{t} p^j$, t ³ 1. Тогда ожидаемый выигрыш $EW(T)$ рассматриваемого игрока за T периодов равен $EW(T) = \sum_{t=1}^{T} d^t W^t$. Если вероятности розыгрыша не зависят от номера периода (то есть одинаковы и равны p), то получаем «классическую» дисконтированную полезность $EW(T) = \sum_{t=1}^{T} (p)^t W^t$. Содержательная вероятностная интерпретация распределения дальновидности такова: чем менее вероятно будущее, тем меньше оцениваемая «сегодня» полезность полученных в этом будущем благ.

4. Режим принятия решений (управления и выбора действий). Тесно связанным с распределением дальновидностей основанием классификации является основание, отражающее последовательность выработки и сообщения управляющих воздействий. Если центр недальновиден и/или в каждом периоде вырабатывает и сообщает АЭ управление, касающееся только данного периода, то такой режим управления называется *текущим* (4.1). Если центр до начала первого периода вырабатывает и сообщает АЭ управления на все будущие периоды, то такой режим управления называется *программным* (4.2). Более гибкой конструкцией является *скользящий режим управления* (4.3), при котором центр в каждом периоде вырабатывает (с учетом вновь поступившей информации) и сообщает АЭ управления на некоторое число будущих периодов.

Перечисленные основания системы классификаций задач стимулирования в ДАС и значения признаков позволяют представить место настоящего исследования (приводимых ниже результатов исследования задач управления ДАС) – см. рисунок 1[1].

---

[1] *Детализация возможных комбинаций режимов управления и дальновидностей производится в следующем разделе* (*см. таблицу* 1). *Кроме того, следует отметить, что допустимы не все комбинации значений призна-*

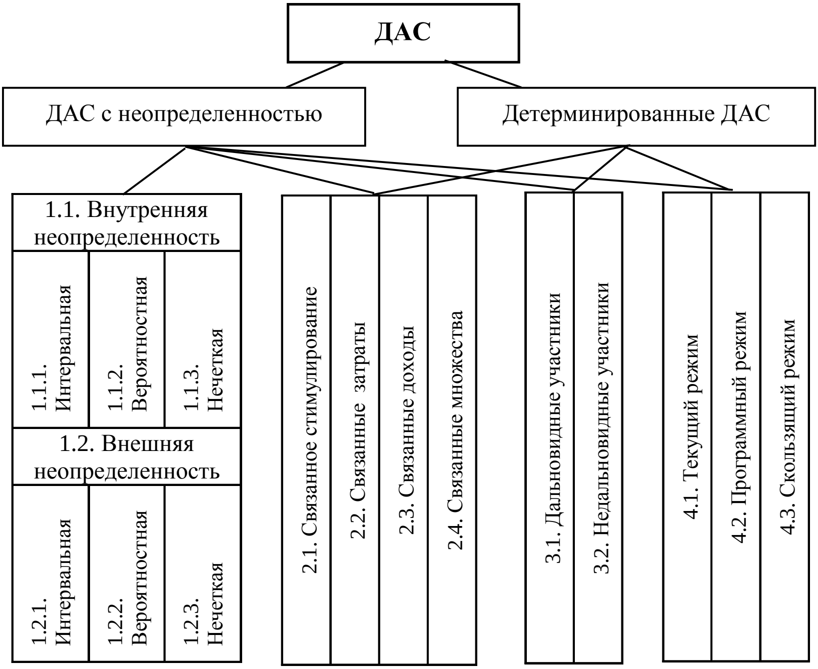

Рис. 1. Система классификаций задач управления в ДАС

2. РАСПРЕДЕЛЕНИЕ ДАЛЬНОВИДНОСТЕЙ И РЕЖИМЫ ПРИНЯТИЯ РЕШЕНИЙ

Пусть $d^t(t)$, $t = \overline{t, T}$, $t = \overline{1, T}$ – распределение дальновидностей (РД) некоторого игрока.

В общем случае, который обозначим РД$_0$, для РД $d^t(t)$ не накладывается никаких ограничений на возможные значения в различные моменты времени, то есть при оценке периода t в момент времени t «привязка» оценки может осуществляться как к оцени-

ков – например, содержательно возможные сочетания распределений дальновидности и режимов управления обсуждаются ниже.

ваемому периоду, так и к тому моменту времени, в котором эта оценка делается. Рассмотрим три частных случая.

1. Оценка периода t не зависит от периода, в котором он оценивается, то есть[1]

(1) $d^t(t) = D^t,\ t = \overline{1,T}$.

Содержательно этот случай, который обозначим РД$_1$, соответствует, например, тому, что внешние условия деятельности в каждом периоде оказывают гораздо большее влияние на значимость благ, получаемых в этом периоде, чем удаленность рассматриваемого периода от момента времени, в котором производится оценка.

2. Оценка периода t зависит от момента времени t, в котором делается оценка, и «удаленности» оцениваемого периода, то есть от разности $(t - t)$ («условия деятельности» в этом периоде не столь важны):

(2) $d^t(t) = D(t - t,\ t),\ t = \overline{t\ ,T}\ ,\ t = \overline{1,T}$,

где $D(\cdot)$ – некоторая функция.

Примером может служить набор дисконтирующих множителей (каждый для «своего» момента t): $D(t - t,\ t) = (D_t)^{t-t}$, где $\{D_t\}$ – некоторые числа. Этот случай обозначим РД$_2$.

3. Наиболее распространенным в прикладных моделях является случай (обозначим его РД$_3$), в котором оценка периода t зависит только от “удаленности” этого периода от момента времени t, в котором делается оценка (периоды функционирования считаются «однородными»), то есть

(3) $d^t(t) = D(t - t),\ t = \overline{t\ ,T}\ ,\ t = \overline{1,T}$,

где $D(\cdot)$ – некоторая (как правило, убывающая) функция.

Примером являются «обычные» дисконтирующие множители: $D(t - t) = (D)^{t-t}$, где D – константа.

Очевидно, для введенных классов распределений дальновидности выполнено следующее вложение: РД$_1 \subseteq$ РД$_0$, РД$_3 \subseteq$ РД$_2 \subseteq$ РД$_0$.

<u>Пример 1.</u> Пусть $T = 3$ и игрок имеет РД, описываемое следующей матрицей (строки соответствуют моментам времени, в которых делаются оценки будущих периодов, столбцы – оценивае-

---

[1] *В настоящей работе принята независимая внутри подразделов нумерация формул.*

мым периодам): $\left\| \begin{matrix} d^1(1) & d^2(1) & d^3(1) \\ \varnothing & d^2(2) & d^3(2) \\ \varnothing & \varnothing & d^3(3) \end{matrix} \right\|$. Тогда в случае $РД_1$ должно быть выполнено: $d^2(1) = d^2(2)$, $d^3(1) = d^3(2) = d^3(3)$, в случае $РД_3$: $d^2(1) = d^3(2)$, $d^1(1) = d^2(2) = d^3(3)$. •[1]

Введем такую (производную по отношению к РД) характеристику игрока как *степень дальновидности* (СД), отражающую число будущих периодов, учитываемых им в текущем периоде:

(4) $x(t) = \max \{t ³ t \mid d^t(t) > 0\} – t$, $t = \overline{1,T}$.

Если $" t = \overline{1,T}$ $x(t) = 0$, то *игрок недальновиден*. Если $" t = \overline{1,T}$ $x(t) = T – t$, то *игрок полностью дальновиден*. Если $\$ x = Const$: $x(t) = min \{x; T – t\}$, то будем говорить, что такой игрок обладает постоянной СД, равной x.

Обсудим соотношение между РД и режимами принятия решений (ПР) об управлении (со стороны центра) и о действиях (со стороны АЭ). Обозначим[2] $L_0(t)$ – число будущих периодов (включая текущий период t), относительно которых центр сообщает свои управления, $t = \overline{1,T}$. Очевидно, $L_0(t)$ £ $T – t +1$. Если $L_0(t) = 1$, $t = \overline{1,T}$, то имеет место *текущий режим управления*, если $L_0(1) = T$, $L_0(t) = 0$, t ³ 2 (или $L_0(t) = T – t + 1$), то реализуется *программное управление*, если 0 £ $L_0(t) < T – t +1$, то реализуется *скользящий режим*[3]. Величину $L(\cdot)$ назовем *горизонтом принятия решений* (для центра горизонт принятия решений (ГПР) иногда называется *горизонтом планирования*).

---

[1] *Символ «·» здесь и далее обозначает окончание примера, доказательства и т.д.*

[2] *Относительно характеристик дальновидности, горизонтов принятия решений и др. будем придерживаться следующих обозначений: нижний индекс «0» обозначает характеристику центра; отсутствие нижнего индекса в одноэлементной АС обозначает характеристику АЭ; нижний индекс обозначает в многоэлементной АС номер АЭ.*

[3] *Понятно, что, если рассматриваются T периодов, то центр должен сообщать некоторые управления для каждого из них («неуправляемые» периоды могут быть исключены из рассмотрения).*

Условие отсутствия «неуправляемых» периодов имеет вид:

(5) $\max\limits_{t=0,t-1} \{L_0(t - t) - t\} \geq 1,\ t = \overline{1,T}$.

Обозначим $L(t)$ – число периодов, на которые принимает решения АЭ в периоде t. Помимо условия

(6) $\max\limits_{t=0,t-1} \{L(t - t) - t\} \geq 1,\ t = \overline{1,T}$,

отражающего тот факт, что на каждый из периодов АЭ должен выбрать стратегию в данном периоде или раньше (ср. с (5)), будем считать, что АЭ может выбирать свои стратегии только на те периоды (текущие и будущие), на которые уже выбрал (и сообщил элементу) свои стратегии центр[1]. Последнее условие запишем в виде

(7) $L(t) \le \max\limits_{t=0,t-1} \{L_0(t - t) - t\},\ t = \overline{1,T}$.

Кроме того, участники АС “не могут” выбирать стратегии на периоды, превышающие их степень дальновидности, то есть

(8) $L_0(t) \le 1 + x_0(t),\ t = \overline{1,T}$,

(9) $L(t) \le 1 + x(t),\ t = \overline{1,T}$.

Условия (5)-(9) накладывают ограничения на допустимые комбинации распределений дальновидности и управлений. Рассмотрим их более подробно.

Введем следующее предположение, которого будем придерживаться в ходе всего последующего изложения (подробное обсуждение роли этого предположения проводится в разделе 3.5).

А.0. Дальновидность и горизонт принятия решений АЭ не превышают соответственно дальновидности и горизонта принятия решений центра.

Предположение А.0 означает, что выполнено следующее условие:

(10) $x(t) \le x_0(t),\ L(t) \le L_0(t),\ t = \overline{1,T}$.

Содержательно предположение А.0 (ГПР АЭ не может превышать ГПР центра – см. условия (5)-(9)) исключает необходимость прогнозирования АЭ будущих управлений со стороны центра.

---

[1] *Отказ от этого предположения, приводящий к эффекту обмена ролями, обсуждается в разделе 3.5.*

В таблице 1 перечислены возможные комбинации РД и ГПР центра и АЭ. Режимы принятия решений обозначены: «Т» – текущий, «С» – скользящий, «П» – программный.

Случаи, исключаемые условиями (8)-(9), заштрихованы.

Случаи, нарушающие предположение А.0, затенены.

На пересечении соответствующих строк и столбцов[1] указаны условные обозначения моделей (ДАС1 – ДАС4), исследуемых ниже (см., в частности, раздел 3.5).

Таким образом, в качестве *базовых моделей ДАС* выделены:

- ДАС1, характеризуемая текущим режимом управления;
- ДАС2, характеризуемая скользящим режимом управления без обязательств;
- ДАС3, характеризуемая скользящим режимом управления с обязательствами;
- ДАС4, характеризуемая программным режимом управления.

В рамках введенной системы классификаций любая модель детерминированной ДАС описывается указанием РД и ГПР центра и АЭ. Например, обозначение ДС-ПТ означает, что рассматривается ДАС с дальновидным центром, использующим скользящий режим ПР, и полностью дальновидным АЭ, использующим текущий режим ПР, и т.д.

[1] *Затененная половина ячейки соответствует случаям, в которых дальновидность АЭ превышает дальновидность центра* (*см. теорему* 4).

Табл. 1. РД и ГПР центра и АЭ

АЭ / Центр		Недальновиден			Дальновиден			Полностью дальновиден		
		Т	С	П	Т	С	П	Т	С	П
Недальновиден	Т	**ДАС1**								
	С									
	П									
Дальновиден	Т	**ДАС2**			**ДАС2**	**ДАС2**				
	С	**ДАС3**			**ДАС3**	**ДАС3**				
	П									
Полностью дальновиден	Т	**ДАС4**			**ДАС4**	**ДАС4**		**ДАС4**	**ДАС4**	**ДАС4**
	С	**ДАС4**			**ДАС4**	**ДАС4**		**ДАС4**	**ДАС4**	**ДАС4**
	П	**ДАС4**			**ДАС4**	**ДАС4**		**ДАС4**	**ДАС4**	**ДАС4**

Введя систему классификаций и рассмотрев возможные взаимоотношения между распределениями дальновидности и горизонтами принятия решений, отражающими степень учета игроками будущего, перейдем к решению задач синтеза оптимальных управлений в динамических активных системах.

3. ЗАДАЧИ СТИМУЛИРОВАНИЯ В ДИНАМИЧЕСКИХ АКТИВНЫХ СИСТЕМАХ

3.1. Задача стимулирования в статической активной системе

Рассмотрим многоэлементную детерминированную статическую двухуровневую *активную систему* (АС), состоящую из *центра* и n *активных элементов* (АЭ). Стратегией АЭ является выбор действий, стратегией центра – выбор *функции стимулирования*, то есть зависимости вознаграждения каждого АЭ от его действий и, быть может, действий других АЭ или других показателей их деятельности.

Обозначим: $y_i \in A_i$ – действие i-го АЭ, $i \in I = \{1, 2, \ldots, n\}$ – множество АЭ, $y = (y_1, y_2, \ldots, y_n) \in A' = \prod_{i=1}^{n} A_i$ – вектор действий АЭ, $y_{-i} = (y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n) \in A_{-i} = \prod_{j \neq i} A_j$ – обстановку игры для i-го АЭ.

Интересы и предпочтения *участников АС* – центра и АЭ – выражены их целевыми функциями. Целевая функция центра является функционалом $F(\sigma, y)$ и представляет собой разность между его доходом $H(y)$ и суммарным вознаграждением $u(y)$, выплачиваемым АЭ: $u(y) = \sum_{i=1}^{n} \sigma_i(y)$, где $\sigma_i(y)$ – стимулирование i-го АЭ, $\sigma(y) = (\sigma_1(y), \sigma_2(y), \ldots, \sigma_n(y))$, то есть

(1) $F(\sigma(\cdot), y) = H(y) - \sum_{i=1}^{n} \sigma_i(y)$.

Целевая функция i-го АЭ является функционалом $f_i(\sigma_i, y)$ и представляет собой разность между стимулированием, получаемым им от центра, и затратами $c_i(y)$, то есть:

(2) $f_i(\sigma_i(\cdot), y) = \sigma_i(y) - c_i(y),\ i \in I.$

Отметим, что индивидуальное вознаграждение и индивидуальные затраты i-го АЭ по выбору действия y_i в общем случае явным или неявным образом зависят от действий всех АЭ (случай сильно связанных АЭ с несепарабельными затратами [59]).

Примем следующий *порядок функционирования АС*. Центру и АЭ на момент принятия решения о выбираемых стратегиях (соответственно – функциях стимулирования и действиях) известны целевые функции и допустимые множества всех участников АС. Центр, обладая правом первого хода, выбирает функции стимулирования и сообщает их АЭ, после чего АЭ при известных функциях стимулирования выбирают действия, максимизирующие их целевые функции.

Относительно параметров АС введем следующие предположения, которые, если не оговорено особо, будем считать выполненными в ходе всего последующего изложения:

А.1. $\forall i \in I$ A_i – отрезок $\Re^1_+$ с левым концом в нуле.

А.2. $\forall i \in I$ 1) функция $c_i(\cdot)$ непрерывна по всем переменным; 2) $\forall y_i \in A_i$ $c_i(y)$ не убывает по y_i, $i \in I$; 3) $\forall y \in A'$, $c_i(y) \geq 0$; 4) $\forall y_{-i} \in A_{-i}$, $c_i(0, y_{-i}) = 0$.

А.3. Функции стимулирования кусочно-непрерывны и принимают неотрицательные значения.

А.4. Функция дохода центра непрерывна и достигает максимума при ненулевом векторе действий АЭ.

Обозначим $P(\sigma)$ – множество равновесных по Нэшу при системе стимулирования σ действий АЭ – *множество реализуемых действий* (то есть будем считать, что АЭ выбирают свои стратегии одновременно и независимо друг от друга, не имея возможности обмениваться дополнительной информацией и полезностью). Минимальными затратами центра на стимулирование по реализации вектора действий АЭ $y' \in A'$ будем называть минимальное значение суммарных выплат элементам, при которых данный вектор действий является равновесием Нэша в игре АЭ, то есть решение следующей задачи: $\sum_{i \in I} \sigma_i(y') \to \min_{\sigma(\cdot) \in \Xi(y')}$, где $\Xi(y') = \{\sigma(\cdot) \mid y' \in P(\sigma)\}$. Как и в одноэлементной АС [56, 58], гарантированной эффективностью (далее просто "эффективностью") стимулирования является минимальное значение целевой функции центра на соответствующем множестве решений игры (всюду, где встречаются минимумы и максимумы, будем предполагать, что они достигаются):

(3) $K(s(\cdot)) = \min\limits_{y \in P(s(\cdot))} F(s(\cdot), y)$.

Задача синтеза оптимальной функции стимулирования заключается в поиске допустимой системы стимулирования s^*, имеющей максимальную эффективность:

(4) $s^* = arg \max\limits_{s(\cdot)} K(s(\cdot))$.

В [59] доказано, что оптимальной (точнее – d-оптимальной, где $d = \sum\limits_{i=1}^{n} d_i$) является квазикомпенсаторная система стимулирования s_K:

(5) $s_{iK} = \begin{cases} c_i(y_i^*, y_{-i}) + d_i, & y_i = y_i^* \\ 0, & y_i \neq y_i^* \end{cases}, i \in I,$

где d_i – сколь угодно малые строго положительные константы, а оптимальное действие y^*, реализуемое системой стимулирования (5) как единственное равновесие в доминантных стратегиях [56], является решением следующей *задачи оптимального согласованного планирования* [18, 58]:

(6) $y^* = arg \max\limits_{y \in A'} \{H(y) - \sum\limits_{i=1}^{n} c_i(y_i)\}$.

Приведенный результат об оптимальности[1] компенсаторных систем стимулирования (5)-(6) получил название *принципа компенсации затрат*. Значение этого результата трудно переоценить, так как он позволяет сразу определить минимальные затраты центра на стимулирование по реализации вектора действий $y \in A'$:

---

[1] *В ходе дальнейшего изложения будем считать, что выполнена гипотеза благожелательности (ГБ), в рамках которой АЭ выбирают из множества $P(s)$ действие, наиболее благоприятное для центра, что позволяет положить $d = 0$. Кроме того, известно, что принцип компенсации затрат (с соответствующими незначительными модификациями) имеет место и в случае, когда необходимо гарантировать АЭ некоторый положительный уровень полезности* [30, 59], *и в задаче стимулирования первого рода* [58], *и в АС, в которых целевая функция АЭ представлена в виде «доход минус штрафы»* [58].

(7) $u(y) = \sum_{i=1}^{n} c_i(y_i)$,

и сконцентрировать все внимание на решении задач выбора оптимальных для центра реализуемых действий (в простейшем случае эта задача имеет вид (6)); в качестве примеров можно привести АС с неопределенностью [58], многоуровневые АС [53], АС с распределенным контролем [60] и др. Как будет видно из последующего изложения, использование принципа компенсации затрат эффективно и при исследовании динамических активных систем.

Кроме того, результаты анализа статической модели позволяют сделать еще один вывод. Система стимулирования (5) побуждает АЭ выбирать соответствующие действия как доминантные стратегии, то есть осуществляет декомпозицию игры АЭ. Возможность добиться подобной декомпозиции в [59] получила название *принципа декомпозиции игры АЭ*. Значимость этого принципа заключается в том, что он позволяет не рассматривать взаимодействие агентов, а решать задачи их стимулирования «независимо». Принцип декомпозиции игры АЭ будет, также как и принцип компенсации затрат, широко использоваться в настоящей работе при исследовании динамических АС, поэтому можно ограничиться рассмотрением задач управления одним АЭ, так как переход к аналогичным АС с несколькими взаимодействующими АЭ приводит лишь к количественному росту сложности оптимизационных задач, не привнося при этом никаких качественных эффектов (многоэлементные ДАС рассмотрены в разделе 3.4).

Перейдем к описанию задач стимулирования в динамических АС.

3.2. Динамические активные системы с несвязанными периодами функционирования

Рассмотрим простейшую модель одноэлементной ДАС с несвязанными периодами функционирования. Взаимодействие участников в данной модели является совокупностью T повторений их взаимодействия в одноэлементной статической модели, то есть центр в каждом периоде $t = \overline{1,T}$ сообщает АЭ управление $s^t(\cdot)$ на этот период, после чего АЭ выбирает действие y^t, причем ни один из параметров модели АС текущего периода не зависит ни от одного из параметров прошлых периодов.

Пусть $y^t \in A^t$ – стратегия АЭ в периоде t, $s^t(\cdot): A^t \to \Re_1^+$ – используемая центром в этом периоде система стимулирования, $t = \overline{1,T}$. Относительно параметров ДАС будем предполагать, что они удовлетворяют предположениям А.1-А.4.

Выигрыш АЭ в периоде t равен

(1) $f^t(s^t, y^t) = s^t(y^t) - c^t(y^t)$, $t = \overline{1,T}$,

где $c^t(\cdot)$ – функция затрат АЭ в этом периоде.

Выигрыш центра в периоде t равен

(2) $F^t(s^t, y^t) = H^t(y^t) - s^t(y^t)$,

где $H^t(\cdot)$ – функция дохода центра в этом периоде, $t = \overline{1,T}$.

Если не оговорено особо, будем считать, что центр должен выбирать такие управления, чтобы в каждом периоде значение целевой функции АЭ было неотрицательно, то есть $f^t(s^t, y^t) \geq 0$, $t = \overline{1,T}$ (условие участия или условие индивидуальной рациональности – Individual Rationality).

Если в каждом периоде целевые функции и допустимые множества удовлетворяют предположениям А.1-А.4, то в соответствии с принципом компенсации затрат[1] задача центра заключается в последовательном определении и реализации *плановой траектории* $x^{1,T} = (x^1, x^2, ..., x^T)$ как результата решения следующей совокупности независимых задач оптимального согласованного планирования:

(3) $x^t = arg \max\limits_{y^t \in A^t} \{H^t(y^t) - c^t(y^t)\}$, $t = \overline{1,T}$.

---

[1] *При ссылке на принцип компенсации затрат здесь и ниже, если не оговорено особо, по умолчанию предполагается, что при использовании центром компенсаторной системы стимулирования или ее модификаций АЭ выбирают действия, равные планам (все рассуждения по обоснованию этого факта повторяют приведенные в* [59, 60] *и опускаются – см. также доказательство теоремы 1 ниже), что требует от центра затрат на стимулирование равных затратам АЭ.*

Если целевая функция центра определяется суммой (по всем периодам) значений его выигрышей (2), то задача оптимального согласованного планирования имеет вид:

$$(4)\ x^{1,T} = arg \max_{y^{1,T} \in A^{1,T}} \sum_{t=1}^{T} \{H^t(y^t) - c^t(y^t)\},$$

где $A^{1,t} = \{y^{1,t} \mid y^t \in A^t, t = \overline{1,t}\} = \prod_{t=1}^{t} A^t$, $t = \overline{1,T}$.

Очевидно, что при несвязанных периодах функционирования решение задачи (3) разбивается на решение T несвязанных однопериодных задач оптимального согласованного планирования, а решения задач (4) и (3) совпадают, что объясняется независимостью периодов.

Если периоды *слабо связаны* (то есть существует единственное ограничение, связывающее действия, или множества допустимых действий, или затраты, или доходы, или вознаграждения и т.д. – см. аналогии в задачах стимулирования в многоэлементных АС со слабо связанными АЭ [59]), то задача (4) превращается в задачу условной оптимизации (изменяется множество действий, по которому ищется максимум).

Основная идея решения задачи стимулирования в этом классе моделей заключается в том, чтобы «перенести» все ограничения на множество допустимых траекторий, а затем решать задачу выбора оптимальной (по критерию суммарного выигрыша центра) допустимой (с учетом всех ограничений) траектории в *расширенном пространстве состояний*[1] (см. также [78-80]). Например, если наложено ограничение R на суммарные выплаты АЭ, то, вводя множество $P(R)$ реализуемых при данном ограничении действий АЭ: $P(R) = \{y^{1,T} \in A^{1,T} \mid \sum_{t=1}^{T} c^t(y^t) \le R\}$, получаем, что оптимальной

---

[1] *Под «расширенным пространством состояний» понимают множество* $A^{1,T}$ *всех допустимых траекторий. При определенных условиях* [78-80] *последовательность задач выбора стратегий участниками ДАС можно рассматривать как задачу однократного выбора стратегии в статической модели с расширенным пространством состояний.*

будет плановая траектория, являющаяся решение следующей задачи:

(5) $x^{1,T} = arg \max\limits_{y^{1,T} \in P(R)} \sum\limits_{t=1}^{T} \{H^t(y^t) - c^t(y^t)\}$.

При этом, очевидно, решение задачи (5) в общем случае не будет являться совокупностью T решений задач (3).

Пример 2. Пусть $T = 2$, $H^t(y^t) = y^t$, $c^t(y^t) = (y^t)^2/2r^t$, $A^t = \Re^1_+$, $t = 1, 2$. Решение задачи (3) имеет вид: $x^t = r^t$, $t = 1, 2$.

Обозначим $R^T = \sum\limits_{t=1}^{T} r^t$. Тогда решение задачи (5) имеет вид:

$$x^t = \begin{cases} r^t, & R \geq R^T / 2 \\ r^t \sqrt{2R / R^T}, & R \leq R^T / 2 \end{cases}, \; t = \overline{1,T}.$$

Выигрыш центра при этом равен min $\{R^T/2; \sqrt{2RR^T} - R\}$. Если предположить, что затраты АЭ одинаковы во всех периодах, то есть $c^t(y^t) = (y^t)^2/2r$, то выигрыш центра монотонен по T. Отсюда следует качественный вывод: если суммарный ресурс ограничен и игроки не учитывают будущее, то центру выгодно «растягивать» процесс взаимодействия с АЭ до бесконечности, побуждая его выбирать в каждом периоде как можно меньшее действие. Отметим, что аналогичный результат имел место при решении задачи определения оптимального числа однородных АЭ, включаемых в состав АС (см. примеры в [53, 59]). •

3.3. Динамические активные системы со связанными периодами функционирования

Рассмотрев ДАС с несвязанными периодами, перейдем к последовательному анализу систем, отличающихся наличием одного и только одного из присущих именно динамическим АС параметров. В соответствии с введенной выше системой классификаций такими параметрами являются: стимулирование, затраты АЭ, доход центра и множества допустимых действий АЭ.

Относительно распределения дальновидностей и режимов управления, если не оговорено особо, будем в настоящем разделе предполагать, что центр полностью дальновиден и использует программный режим управления, а АЭ либо недальновиден, либо полностью дальновиден и выбирает свои действия в каждом периоде.

ДАС с зависимым стимулированием

Если стимулирование АЭ в каждом периоде зависит как от его действия в этом периоде, так и от его действий во всех предыдущих периодах, то есть $s^t = s^t(y^{1,t})$, то в соответствии с принципом компенсации затрат оптимальной будет система стимулирования

(1) $$s^t_K(x^{1,T}, y^{1,t}) = \begin{cases} c^t(x^t), \text{если } y^i = x^i,\ i = \overline{1,t} \\ 0, \quad \text{в остальных случаях} \end{cases},\ t = \overline{1,T},$$

где оптимальная плановая траектория определяется как и в ДАС с несвязанными периодами (см. выше):

(2) $$x^{1,T} = arg \max_{y^{1,T} \in A^{1,T}} \sum_{t=1}^{T} \{H^t(y^t) - c^t(y^t)\}.$$

ДАС с зависимыми затратами

Как отмечалось выше, под связанными (зависимыми) затратами в ДАС будем понимать такие функции затрат АЭ, которые в каждом периоде зависят не только от действия АЭ в этом периоде, но и от его действий во всех предыдущих периодах, то есть $c^t = c^t(y^{1,t})$.

Введем следующее предположение относительно свойств функции затрат АЭ.

А.2'. $\forall\ t = \overline{1,T}$ 1) функция $c^t(\cdot)$ непрерывна по всем переменным; 2) $\forall\ y^{1,t} \in A^{1,t}$ $c^t(y^{1,t})$ не убывает по y^t, $t = \overline{1,T}$; 3) $\forall\ y^{1,t} \in A^{1,t}$ $c^t(y^{1,t}) \geq 0$, $t = \overline{1,T}$; 4) $\forall\ y^{1,t-1} \in A^{1,t-1}$ $c^t(y^{1,t-1}, 0) = 0$, $t = \overline{2,T}$.

Если центр сообщает недальновидному АЭ управление в каждом периоде (текущий режим управления – см. выше), то в соответствии с принципом компенсации затрат оптимальной будет система стимулирования

(3) $s^t_K(x^{1,T}, y^t) = \begin{cases} c^t(x^{1,t}), & \text{если } y^t = x^t \\ 0, \text{ в остальных случаях} & \end{cases}, \ t = \overline{1,T}$,

где оптимальная плановая траектория определяется как решение следующей задачи:

(4) $x^{1,T} = arg \max\limits_{y^{1,T} \in A^{1,T}} \sum\limits_{t=1}^{T} \{H^t(y^t) - c^t(y^{1,t})\}$.

Если АЭ полностью дальновиден, а центр использует программный режим управления, то есть сообщает АЭ до начала первого периода управление сразу на все T периодов, то оптимальной будет система стимулирования

$$s^t_K(x^{1,T}, y^t) = \begin{cases} c^t(x^t, y^{1,t-1}), & \text{если } y^t = x^t \\ 0, \text{ в остальных случаях} & \end{cases}, \ t = \overline{1,T}.$$

Пример 3. Пусть $T = 2$, $H^t(y^t) = y^t$, $c^t(y^t) = (y^t - g\, y^{t-1})^2/2r^t$, $A^t = \Re^1_+$, $t = 1, 2$. Решение задачи (4) имеет вид: $x^1 = (1 + g)\, r^1$, $x^2 = r^2 + g\,(1 + g)\, r^1$ и при $g = 0$ переходит в решение, оптимальное в соответствующей ДАС с несвязанными периодами (см. пример 2). •

ДАС со связанным доходом

Если доход центра в каждом периоде зависит от действий АЭ, выбранных в данном и всех предыдущих периодах, то оптимальной будет следующая плановая траектория:

(5) $x^{1,T} = arg \max\limits_{y^{1,T} \in A^{1,T}} \sum\limits_{t=1}^{T} \{H^t(y^{1,t}) - c^t(y^t)\}$

при компенсаторной функции стимулирования

(6) $s^t_K(x^t, y^t) = \begin{cases} c^t(x^t), \text{ если } y^t = x^t \\ 0, \quad\quad \text{если } y^t \neq x^t \end{cases}, \ t = \overline{1,T}$.

ДАС со связанными ограничениями

Пусть множество допустимых действий АЭ в периоде t зависит от его действий в предыдущих периодах, то есть $A^t = A^t(y^{1,t-1})$, $t = \overline{2,T}$, множество A^1 считается фиксированным. Тогда, используя систему стимулирования (6), центр определяет оптимальную плановую траекторию как решение следующей задачи оптимального согласованного планирования

$$(7)\ x^{1,T} = arg \max_{y^{1,T} \in A_0^{1,T}} \sum_{t=1}^{T} \{H^t(y^t) - c^t(y^t)\},$$

где

$$(8)\ A_0^{1,t} = \{y^{1,t} \in A^{1,t} \mid y^t \in A^t(y^{1,t-1}),\ t = \overline{1,t}\},\ t = \overline{1,T}.$$

<u>Пример 4.</u> Пусть $T = 2$, $H^t(y^t) = y^t$, $c^t(y^t) = (y^t)^2/2r^t$, $t = 1, 2$, $A^1 = \Re^1_+$, $A^2(y^1) = [y^1; +\infty]$. Фиксируем $y^1 \geq 0$, тогда оптимальный план $x^2(y^1)$ на второй период равен: $x^2(y^1) = \begin{cases} r^2, & y^1 \leq r^2 \\ y^1, & y^1 \geq r^2 \end{cases}$. Следовательно, решение задачи (7) имеет вид:

$$x^{1,2} = \begin{cases} (r^1, r^2), & r^1 \leq r^2 \\ (2r^1r^2/(r^1+r^2),\ 2r^1r^2/(r^1+r^2)), & r^1 \geq r^2 \end{cases}. \bullet$$

Общая модель детерминированной ДАС

Итак, из рассмотрения четырех описанных выше частных моделей детерминированных ДАС со связанными периодами можно сделать качественный вывод, что для решения соответствующих задач стимулирования, наряду с принципом компенсации затрат, приходится использовать обобщения[1] *принципа оптимальности*

---

[1] *В случае полностью зависимых периодов непосредственное использование принципа Беллмана неэффективно, так как условно оптимальные управления на последнем шаге (в последнем периоде) в общем случае должно параметрически зависеть от управлений во всех предыдущих периодах (так как присутствует «последействие»), начиная с первого.*

Беллмана (см. примеры 3 и 4), что качественно отличает их от модели ДАС с несвязанными или со слабо связанными периодами, в которых применение принципа компенсации затрат сводило задачу управления к стандартной задаче условной оптимизации.

Выше мы рассмотрели четыре модели ДАС, отличающиеся от ДАС с несвязанными периодами «связанностью» периодов по одной из компонент (стимулирование, затраты, доход, допустимые множества). Можно было бы последовательно продолжать усложнение моделей, рассматривая попарные комбинации «связанных» компонент (таких моделей было бы 6), затем комбинации из трех «связанных» компонент (таких моделей было бы 4) – по аналогии с тем как это делалось для АС с распределенным контролем в [60]. Но анализ первых четырех моделей свидетельствует, что решение задачи стимулирования в них имеет простой вид, поэтому сформулируем сразу задачу стимулирования в детерминированной ДАС со связанными стимулированием, затратами, доходом и допустимыми множествами.

Пусть $s^t = s^t(y^{1,t})$, $c^t = c^t(y^{1,t})$, $H^t = H^t(y^{1,t})$, $A^t = A^t(y^{1,t})$, $t = \overline{1,T}$, а центр и АЭ полностью дальновидны и центр использует программный режим управления.

Принцип Беллмана в явном виде эффективен, если параметры каждого периода зависят только от параметров предыдущего периода, то есть, например, $A^t = A^t(y^{t-1})$, $c^t = c^t(y^t, y^{t-1})$, $H^t = H^t(y^t, y^{t-1})$ *и т.д. Такая «неполная» зависимость параметров во многих случаях достаточно хорошо отражает специфику ДАС. Однако, в настоящей работе мы не будем обращать внимания на «вычислительные» трудности, считая задачу управления решенной если она сведена к известной оптимизационной (пусть даже достаточно сложной и требующей дополнительного исследования методов ее решения) задаче. Принцип компенсации затрат разделяет исходную задачу на две составляющих – задачу согласованного стимулирования, решением которой является (9), и задачу согласованного планирования (10). Если методы решения первого класса задач (задач согласованного планирования) для детерминированных АС хорошо известны* [6, 18, 19], *то в ДАС основную сложность представляет решение именно задач согласованного планирования. При этом желательно не только свести ту или иную задачу планирования к известной оптимизационной задаче, но и проанализировать зависимость свойств ее решения от параметров модели ДАС (см. теоремы 5, 6, 7).*

Теорема 1. Если выполнены предположения А.0, А.1, А.2’, А.3 и А.4, то при использовании центром системы стимулирования[1]

$$(9)\quad s^t{}_K(x^{1,T}, y^{1,t}) = \begin{cases} c^t(x^{1,t}), \text{ если } y^i = x^i,\ i = \overline{1,t} \\ 0, \quad \text{в остальных случаях} \end{cases},\ t = \overline{1,T},$$

где оптимальная плановая траектория определяется:

$$(10)\quad x^{1,T} = arg \max_{y^{1,T} \in A_0^{1,T}} \sum_{t=1}^{T} \{H^t(y^{1,t}) - c^t(y^{1,t})\},$$

а $A_0^{1,t}$ – выражением (8). Действия АЭ при этом совпадут с планами и эффективность стимулирования K_0 будет максимально возможной, где

$$K_0 = \max_{y^{1,T} \in \mathrm{A}_0^{1,\mathrm{T}}} \sum_{t=1}^{T} \{H^t(y^{1,t}) - c^t(y^{1,t})\}.$$

Доказательство[2]. Задача стимулирования заключается в выборе центром системы стимулирования $s^{1,T}(\cdot)$, которая максимизировала бы его целевую функцию, учитывающую в силу полной дальновидности центра все T периодов.

$$(11)\quad F(s^{1,T}(\cdot),\ y_*^{1,T}) = \sum_{t=1}^{T} \{H^t(y_*^{1,t}) - s^t(y_*^{1,t})\}$$

---

[1] *Если вознаграждение АЭ в любом периоде может зависеть от его действий, выбранных только в этом периоде, то есть* (9) *заменяется на* (3), *то результат теоремы 1 уже не имеет места – см. теорему 2.*

[2] *Доказательство теоремы 1 следует общей схеме доказательства всех результатов об оптимальности тех или иных систем стимулирования: сначала показывается, что достаточно поощрять агента за выбор только одного действия* (*вектора действий, траектории и т.д.*), *далее доказывается, что для этого необходимо как минимум компенсировать его затраты, после чего доказательство состоит из двух шагов – на первом шаге проверяется, что при некотором* (*произвольном*) *плане и использовании соответствующей компенсаторной системы стимулирования АЭ будет выбирать действия, совпадающие с планами* (*этап проверки согласованности системы стимулирования*), *затем на втором шаге* (*этап согласованного планирования*) *ищутся оптимальные реализуемые планы* (*см. также* [15, 58]).

при условии, что действия АЭ $y_*^{1,T}$, выбираемые им при известной системе стимулирования, максимизируют его целевую функцию $f(\sigma^{1,T}(\cdot), y^{1,T})$, , учитывающую в силу полной дальновидности АЭ все T периодов, то есть

$$(12)\ y_*^{1,T} \in Arg \max_{y^{1,T} \in A^{1,T}} \sum_{t=1}^{T} \{\sigma^t(y_*^{1,t}) - c^t(y_*^{1,t})\}.$$

Фиксируем произвольную плановую траекторию $z^{1,T} \in A^{1,T}$. Пусть некоторая система стимулирования $\sigma^{1,T}(\cdot)$ реализует эту плановую траекторию, то есть

$$(13)\ \sum_{t=1}^{T} \{\sigma^t(z^{1,t}) - c^t(z^{1,t})\} \geq \sum_{t=1}^{T} \{\sigma^t(y^{1,t}) - c^t(y^{1,t})\},\ \forall\, y^{1,T} \in A^{1,T}.$$

Перейдем от системы стимулирования $\sigma^{1,T}(\cdot)$ к соответствующей квази-системе стимулирования [42] $q\sigma^{1,T}(\cdot)$ следующим образом:

$$(14)\ q\sigma^t(y^{1,t}) = \begin{cases} \sigma^t(z^{1,t}), & y^{1,t} = z^{1,t} \\ 0, & y^{1,t} \neq z^{1,t} \end{cases},\ t = \overline{1,T}.$$

Если заменить в выражении (13) $\sigma^{1,T}(\cdot)$ на $q\sigma^{1,T}(\cdot)$, то система неравенств останется в силе, то есть плановая траектория $z^{1,T}$ будет реализовываться и системой стимулирования $q\sigma^{1,T}(\cdot)$, а фактические выплаты активному элементу не изменятся.

Таким образом, мы доказали, что без потери эффективности можно ограничиться классом систем стимулирования типа (14), которому в том числе принадлежит система стимулирования (9).

Фиксируем произвольную плановую траекторию $z^{1,T} \in A^{1,T}$. Из (11) и (13) следует, что при фиксированной плановой траектории центр стремится найти реализующую ее систему стимулирования, которая обладала бы минимальными затратами на стимулирование, то есть центр решает следующую задачу:

$$(15)\ \sum_{t=1}^{T} \sigma^t(z^{1,t}) \to \min$$

$$(16)\ \sum_{t=1}^{T} \{\sigma^t(z^{1,t}) - c^t(z^{1,t})\} \geq -\sum_{t=1}^{T} c^t(y^{1,t}),\ \forall\, y^{1,T} \in A^{1,T}.$$

Из предположения А.2’ следует, что максимум правой части выражения (16) достигается в том числе при нулевых действиях АЭ и равен нулю. Кроме того, выше предполагалось, что центр должен в каждом периоде обеспечить АЭ неотрицательную полезность, то есть каждое из слагаемых в левой части выражения (16) неотрицательно. Следовательно, одно из решений задачи (15)-(16) имеет вид

(17) $s^t(z^{1,t}) = c^t(z^{1,t}),\ t = \overline{1,T}$.

Значит минимальная система стимулирования, реализующая плановую траекторию $z^{1,T}$, удовлетворяет одновременно (14) и (17), что дает выражение (9). При этом значение целевой функции АЭ в каждом периоде неположительно, а при выборе действий, совпадающих с планами, равно нулю.

То, что агент при использовании центром управления (9)-(10) выберет действия, совпадающие с планами, следует из подстановки (9) в (12) – если в любом из периодов АЭ выбирает действия, отличающиеся от планов, то значение его целевой функций не увеличивается (для того, чтобы планы были единственными точками максимума достаточно доплачивать АЭ за их выбор, помимо компенсации затрат, сколь угодно малую, но строго положительную величину – см. выше и [30, 59, 60]).

Суммируя (17) по всем периодам, получим следующую оценку минимальных затрат $u(\cdot)$ на реализацию плановой траектории $z^{1,T}$:

(18) $u(z^{1,T}) = \sum_{t=1}^{T} c^t(z^{1,t})$.

Таким образом, мы показали, что системы стимулирования вида (14), (17) реализуют плановую траекторию[1] $z^{1,t}$ с минимальными затратами центра на стимулирование, определяемыми (18). Вспоминая, что плановая траектория выбиралась произвольной, получаем, что необходимо найти плановую траекторию, которая максими-

---

[1] *Еще раз подчеркнем, что на значения целевой функции АЭ в каждом периоде могут быть наложены дополнительные ограничения* (*гарантированное обеспечение ненулевой резервной полезности, или некоторого значения полезности, зависящего от действий АЭ, и т.д.*). *Однако, как доказано в* [42, 58, 59], *введение подобных ограничений не меняет вида и основных свойств решения задачи стимулирования, поэтому рассматривать подобные искусственные «усложнения» модели мы не будем, стремясь акцентировать внимание на специфике динамики.*

зировала бы разность между $\sum_{t=1}^{T} H^t(z^{1,T})$ и $u(z^{1,T})$ (см. (11)), что и отражено выражением (10). •

Обсудим результат теоремы 1.

Очевидно, что, во-первых, в соответствии с (9) центр может не запоминать какие действия выбирает АЭ в каждом периоде – ему необходимо лишь знать отклонялся ли АЭ в прошлом хотя бы раз от планов или нет.

Во-вторых, в силу полной дальновидности центра результат теоремы 1 справедлив для любого режима управления активным элементом со стороны центра, то есть центр может в рамках предположения А.0 как сообщать АЭ всю информацию (9)-(10) до начала первого периода, так и в каждом периоде сообщать только управление для этого периода и/или на любое число будущих периодов (см. более подробное обсуждение в разделе 3.5).

В третьих, введение различного учета будущего[1] участниками АС не изменяет результата теоремы 1, за исключением того, что оптимальная плановая траектория, независимо от распределения дальновидностей АЭ, будет иметь вид

$$x^{1,T} = arg \max_{y^{1,T} \in A^{1,T}} \sum_{t=1}^{T} d_0^{\,t} \{H^t(y^{1,t}) - c^t(y^{1,t})\},$$

где $d_0^{1,T} = (d_0^{\,1}, d_0^{\,2}, ..., d_0^{\,T})$ – распределение дальновидностей центра типа $РД_1$ или $РД_3$. Справедливость последнего утверждения следует из того, что при использовании центром системы стимулирования (9) выигрыш АЭ в каждом периоде тождественно равен нулю.

Сделав маленькое отступление, отметим, что отказ от предположения о том, что центр должен в каждом периоде обеспечить АЭ неотрицательную полезность, и замена его требованием обеспечения неотрицательной суммарной (по всем T периодам) полезности, приводит к тому, что центр должен решать следующую задачу:

$$(19)\quad \sum_{t=1}^{T} \{H^t(y^{1,t}) - e^t\} \;®\; \max_{y^{1,T} \in A_0^{1,T},\{e^t\}},$$

[1] *Рассматриваемый учет участниками АС будущих периодов может считаться автоматически включенным в зависимость функции дохода центра и функции затрат АЭ от времени.*

где последовательность $\{e^t\}$ неотрицательных чисел и вектор $y^{1,T}$ таковы, что

(20) $\sum_{t=1}^{T} e^t = \sum_{t=1}^{T} c^t(y^{1,t})$.

Условие (20) гарантирует АЭ компенсацию суммарных затрат (то есть обеспечивает неотрицательность суммарной (по всем T периодам) полезности АЭ.

Отметим также, что (10) удовлетворяет условию (20), но в общем случае не является ни одним из решений задачи (19)-(20). Другими словами, при отказе от условий индивидуальной рациональности АЭ в каждом периоде множество допустимых (с точки зрения условий участия (индивидуальной рациональности) и согласованности стимулирования) управлений увеличивается и, следовательно, не снижается эффективность управления.

Если условие *индивидуальной рациональности* АЭ (условие его участия) имеет вид (ср. с (20)) $\sum_{t=1}^{T} b^t f^t$ ³ U, где $b^{1,T} = (b^1, b^2, ..., b^T)$ – распределение дальновидностей АЭ, то центру следует в каждом периоде доплачивать АЭ (помимо компенсации затрат) величину e^t, совокупность которых определится из решения задачи (ср. с (19))

$$\sum_{t=1}^{T} d^t \{H^t(y^{1,t}) - c^t(y^{1,t}) - e^t\} \text{ ® } \max_{y^{1,T} \in A_0^{1,T},\ \{e^t \geq 0\}: \sum_{t=1}^{T} b^t e^t \geq U}.$$

В четвертых, если центр одинаково учитывает будущие полезности (выигрыши всех периодов входят в его целевую функцию с одинаковыми весами), то система стимулирования (9)-(10) оптимальна и при отказе от необходимости обеспечения неотрицательности целевой функции АЭ в каждом периоде (достигается минимум суммарного стимулирования, компенсирующего затраты и реализующего требуемую плановую траекторию).

Если центр по-разному учитывает будущие периоды и не требуется обеспечивать АЭ в каждом периоде неотрицательную полезность, то система стимулирования (9)-(10) в общем случае не оптимальна. Например, если

(21) 1 ³ d_0^1 ³ $... > 0,$

то оптимальной для центра будет следующая система стимулирования $s_T^{1,T}(\cdot)$;

(22) $s_T^{1,t}(y^{1,t}) = 0,\ t = \overline{1, T-1};\ s_T^{1,T}(y^{1,T}) = \begin{cases} \sum_{t=1}^{T} c^t(x^{1,t}),\ y^{1,T} = x^{1,T} \\ 0,\ y^{1,T} \neq x^{1,T} \end{cases}$

при плановой траектории

(23) $x^{1,T} = arg \max_{y^{1,T} \in A^{1,T}} \{\sum_{t=1}^{T} d_0^t\ H^t(y^{1,t}) - d_0^T \sum_{t=1}^{T} c^t(y^{1,t})\}$.

Содержательно, использование центром управлений (22)-(23) в рассматриваемом случае означает, что ему выгодны аккордные системы оплаты деятельности АЭ, при которых расчет осуществляется «полной суммой», но откладывается до последнего момента времени (при этом, правда, нарушается требование обеспечения условия индивидуальной рациональности АЭ в каждом периоде).

Таким образом, так как методика анализа одинакова, то будем считать, что условия $f^t \geq 0$ выполнены для всех периодов $t = \overline{1,T}$.

Завершив обсуждение теоремы 1, рассмотрим случай, когда вознаграждение АЭ в каждом периоде может зависеть только от его действий в этом периоде, то есть $s^t = s^t(y^t)$ и центр использует систему стимулирования (3). Если АЭ недальновиден, или если его затраты не связаны, то в рамках предположения А.0 оптимальна и реализуема плановая траектория (10). Отличие появляется при использовании центром программного управления, то есть сообщения дальновидному АЭ со связанными затратами до начала первого периода сразу всей (или части) плановой траектории и всех (или части) зависимостей вознаграждения от действий. Оказывается, что при связанных затратах и несвязанном стимулировании множество реализуемых траекторий не шире, а эффективность стимулирования не выше, чем при связанном стимулировании (см. описание контрактов с памятью в приложении) – ср. (10) и (24), (25).

Теорема 2. Если выполнены предположения А.0, А.1, А.2’, А.3 и А.4, то при использовании центром системы стимулирования[1] (3) и оптимальной плановой траектории:

---

[1] *Отметим, что в соответствии с* (3) *выплаты АЭ в текущем периоде зависят от его действий в этом периоде и от планов* (*но не действий!*) *в*

$$(24)\ x^{1,T} = arg \max_{y^{1,T} \in X^{1,T}} \sum_{t=1}^{T} \{H^t(y^{1,t}) - c^t(y^{1,t})\},$$

где[1]

$$(25)\ X^{1,T} = \{x^{1,T} \in A_0^{1,T} \mid \forall\, y^{1,T} \in A_0^{1,T} \quad \sum_{\substack{t=1 \\ x^t \neq y^t}}^{T} c^t(y^{1,t}) \geq \sum_{\substack{t=1 \\ x^t = y^t}}^{T} [c^t(x^{1,t}) - c^t(y^{1,t-1}, x^t)]\},$$

действия АЭ совпадут с планами и эффективность стимулирования будет максимально возможной при несвязанном стимулировании.

Доказательство. Отметим, что формулировка теоремы 2 отличается от формулировки теоремы 1 только видом системы стимулирования (ср. (3) и (9)) и тем множеством траекторий, по которому ведется максимизация при определении оптимальной плановой траектории.

Невозможность реализации произвольной плановой траектории системой стимулирования (3) обусловлена тем, что, выбирая в некотором периоде действия, отличные от планов, в случае связанных затрат АЭ может в общем случае изменить свои затраты в будущих периодах, а центр не имеет возможности в текущем периоде наказать АЭ за отклонения в прошлых периодах.

Система неравенств (25) отражает невыгодность отклонения АЭ от плана. Действительно, при отклонениях АЭ несет потери, фигурирующие в левой части (суммирование ведется по тем периодам, в которых планы не выполнялись), в правой части стоит выигрыш от отклонений. Если потери превышают выигрыш, то отклонение невыгодно.

Итак, выражение (25) определяет множество плановых траекторий, реализация которых выгодна для АЭ (точнее – невыгодно отклонение от них). В остальном доказательство теоремы 2 повторяет доказательство теоремы 1 и опускается. •

Содержательно множество $X^{1,T}$, определяемое выражением (25), может интерпретироваться как множество *согласованных планов*.

предыдущих периодах (*ср. для контраста с* (*26*)).

[1] *В частности, при несвязанных затратах выполнено* $X^{1,T} = A^{1,T}$.

Отметим, что, если вместо (3) центр может использовать следующую систему стимулирования, являющуюся более «мягкой», чем (9):

(26) $$s^t_K(x^{1,T}, y^{1,t}) = \begin{cases} c^t(y^{1,t-1}, x^t), \text{если } y^t = x^t \\ 0, \quad \text{в остальных случаях} \end{cases}, t = \overline{1,T},$$

то реализуема любая траектория из $A_0^{1,T}$, но при этом, в соответствии с (26), выплаты АЭ в текущем периоде зависят уже от всей предыстории (в отличие от (3)). Это утверждение сформулируем в виде следствия из теорем 1 и 2:

Следствие 1. Системы стимулирования (9) и (26) характеризуются максимальным множеством реализуемых действий и максимальной эффективностью.

Содержательно, при использовании системы стимулирования (9) центр отслеживает отклонения АЭ от плана в течение всей предыстории (по отношению к рассматриваемому периоду) и выплачивает АЭ ненулевое вознаграждение (компенсирует ему затраты) только если он ни разу не отклонился от плана. В соответствии с (26) центр может не «помнить» отклонения, а компенсировать в каждом периоде затраты АЭ при выполнении им плана в этом периоде с учетом фактически сложившейся истории. Легко видеть, что при этом АЭ не может получить в текущем периоде выигрыша за счет отклонений в предыдущих периодах (ср. с выражением (25)).

3.4. Многоэлементные динамические активные системы

Рассмотрим кратко многоэлементную модель – ДАС с n АЭ, стратегией каждого из которых в каждом периоде является выбор (при известном управлении со стороны центра) некоторого действия $y_i^t \in A_i^t$, i Î $I = \{1, 2, ..., n\}$, $t = \overline{1,T}$ (см. также обозначения и основные результаты исследования многоэлементных статических АС в разделе 3.1). Обозначим $y^t = (y_1^t, y_2^t, ..., y_n^t)$ – вектор стратегий всех игроков в момент времени t, $y^{1,T} = (y^1, y^2, ..., y^T)$ – вектор стратегий всех игроков за периоды с первого периода по период T.

Пусть $\sigma_i^t = \sigma_i^t(y^{1,t})$, $c_i^t = c_i^t(y^{1,t})$, $H^t = H^t(y^{1,t})$, $A_i^t = A_i^t(y^{1,t-1})$, $i \in I$, $t = \overline{1,T}$. Определим $A^t = \prod_{i \in I} A_i^t$, $A_{-i}^t = \prod_{j \neq i} A_j^t$, $A^{1,t} = \prod_{t=1}^{t} A^t$,

(1) $A_0^{1,t} = \{y^{1,t} \in A^{1,t} \mid y^t \in A^t(y^{1,t-1}),\ t = \overline{1,t}\}$, $t = \overline{1,T}$.

Введем дополнительное предположение относительно свойств функций затрат АЭ (отметим, что данное предположение является «объединением» предположений А.2 и А.2', отражающих свойства функций затрат, соответственно, в статической многоэлементной АС и в одноэлементной ДАС).

А.2''. $\forall\ t = \overline{1,T}$, $i \in I$ 1) функция $c_i^t(\cdot)$ непрерывна по всем переменным; 2) $\forall\ y^{1,t} \in A^{1,t}$ $c_i^t(y^{1,t})$ не убывает по y_i^t; 3) $\forall\ y^{1,t} \in A^{1,t}$ $c_i^t(y^{1,t}) \geq 0$; 4) $\forall\ y^{1,t-1} \in A^{1,t-1}$, $\forall\ y_{-i}^t \in A_{-i}^t$ $c^t(y^{1,t-1}, y_{-i}^t, 0) = 0$.

<u>Теорема 3.</u> Если выполнены предположения А.1, А.2'', А.3 и А.4, то при использовании центром системы стимулирования

(2) $\sigma_{iK}^t(x^{1,T}, y^{1,t}) = \begin{cases} c_i^t(x_i^{1,t}, y_{-i}^{1,t}), & \text{если } y_i^k = x_i^k,\ k = \overline{1,t} \\ 0, & \text{в остальных случаях} \end{cases}$, $t = \overline{1,T}$, $i \in I$,

где

(3) $x^{1,T} = \arg \max\limits_{y^{1,T} \in A_0^{1,T}} \sum\limits_{t=1}^{T} \{H^t(y^{1,t}) - \sum\limits_{i \in I} c_i^t(y^{1,t})\}$,

действия АЭ совпадут с планами и эффективность стимулирования будет максимально возможной.

Доказательство. В [59] был введен принцип декомпозиции игры АЭ в задачах стимулирования, заключающийся в том, что при использовании в многоэлементных АС компенсаторных систем стимулирования, в которых АЭ компенсировались затраты в случае выбора им соответствующей плановой компоненты (независимо от действий других АЭ!), выбор действий, совпадающих с планами, является доминантной стратегией каждого АЭ.

Если выполнено (2), то, применяя принцип декомпозиции, получаем возможность независимо рассматривать n задач управления несвязанными между собой активными элементами. Для каждой из этих задач в отдельности применима теорема 1. •

Для многоэлементных ДАС в предположении, что АЭ в каждом периоде выбирают равновесные по Нэшу стратегии, справедливы аналоги теоремы 2 и следствия 1.

3.5. Влияние распределений дальновидности и горизонтов принятия решений на эффективность управления

Результаты теорем 1-3 были получены в предположении, что центр полностью дальновиден и использует программный режим управления, а АЭ либо недальновиден и принимает решения на текущий период, либо полностью дальновиден и принимает решения сразу на все T периодов (что в силу предположения А.0 возможно только при программном управлении со стороны центра).

Приведенная во втором разделе таблица 1 содержит классификацию ДАС по распределениям дальновидности и горизонтам принятия решений. Исследуем сравнительную эффективность различных режимов управления при тех или иных распределениях дальновидностей, считая стимулирование связанным[1]. Будем последовательно рассматривать модели, соответствующие незаштрихованным ячейкам таблицы 1, двигаясь из верхнего левого угла вправо и вниз. При этом, если не оговорено особо, стимулирование будем считать связанным.

Модель НТ-НТ (ДАС1). Данная модель подробно исследована выше. Центр в каждом периоде сообщает АЭ систему стимулирования

$$(1)\ s^t_K(x^{1,t}, y^{1,t}) = \begin{cases} c^t(y^{1,t-1}, x^t), \text{если } y^t = x^t \\ 0, \quad \text{в остальных случаях} \end{cases},\ t = \overline{1,T},$$

где $y^{1,t-1}$ – *траектория реализаций*, сложившаяся к периоду t, и план x^t, а АЭ выбором действия y^t стремится максимизировать свой выигрыш в текущем периоде.

Очевидно, что выбор действия, совпадающего с планом, выгоден для АЭ, поэтому центру достаточно решить задачу выбора плановой траектории исходя из условия, что план каждого периода максимизирует выигрыш центра в этом (и только в этом, в силу недальновидности центра) периоде:

---

[1] *Для случая несвязанного стимулирования, но связанных остальных параметров ДАС, оценки сравнительной эффективности управления получаются по аналогии с теоремой 2 (ср. теоремы 1 и 2) и поэтому опускаются.*

$$(2)\ x^t = \tilde{x}^t(x^{1,t-1}) = arg \max_{y^t \in A^t(x^{1,t-1})} \{H^t(x^{1,t-1}, y^t) - c^t(x^{1,t-1}, y^t)\},\ t = \overline{1,T}\ .$$

Обозначим K_1 – эффективность стимулирования в модели ДАС1: $K_1 = F(s^{1,T}, x^{1,T})$, где $s^{1,T}$ удовлетворяет (1), а $x^{1,T}$ удовлетворяет (2).

Напомним (см. выше), что в рассматриваемой модели центр может добиться той же эффективности, используя систему стимулирования

$$(1а)\ s^t_K(x^{1,T}, y^{1,t}) = \begin{cases} c^t(x^{1,t}), \text{ если } y^i = x^i,\ i = \overline{1,t} \\ 0, \quad \text{в остальных случаях} \end{cases},\ t = \overline{1,T}\ ,$$

с планами (2).

Модель НТ-ДТ. В данной модели центр использует управления (1)-(2), так как его дальновидность и ГПР не изменились по сравнению с моделью ДАС1. Так как АЭ дальновиден, то он должен либо отказаться от своей дальновидности и принимать решения, стремясь максимизировать текущие однопериодные выигрыши (при этом приходим к модели ДАС1), либо прогнозировать будущие управления центра в пределах своего горизонта дальновидности. Прогнозы зависят от тех предположений, которые АЭ делает о поведении центра, то есть от того, каким образом он устраняет существующую игровую неопределенность. Если АЭ рассчитывает на максимальный гарантированный результат, то есть предполагает, что стимулирование со стороны центра в будущие периоды будет тождественно равно нулю, то и его будущие действия должны быть равны нулю. Этот случай практически не интересен. Поэтому будем считать, что при прогнозе АЭ рассчитывает, что центр будет выбирать ненулевое стимулирование в будущих периодах.

Тогда в общем случае будет иметь место ***эффект обмена ролями***[1] (ЭОР), заключающийся в том, что АЭ начнет играть

[1] *Эффект обмена ролями в терминах теории иерархических игр заключается в переходе от игры Γ_2 к игре Γ_2^*, в которой АЭ становится «первым игроком», то есть игроком, делающим первый ход. Обсуждение распределения ролей также проводилось с теоретической точки зрения в* [60], *с точки зрения трудовых контрактов – в* [42], *с точки зрения задач рекрутинга – в* [43].

роль «центра», навязывая «настоящему центру» будущие управления. Поясним последнее утверждение. Задача АЭ в периоде t заключается в следующем: выбрать такое действие y^t, которое максимизировало бы сумму его выигрышей за периоды с t по $(t + x(t))$ при условии что центр в периодах $(t+1, t + x(t))$ использует принцип планирования (2), подставляя в него вместо плановой траектории $x^{1,t-1}$ траекторию реализаций (историю игры) $y^{1,t-1}$, то есть АЭ рассчитывает на назначение центром плана

$$x^t = arg \max_{y^t \in A^t(x^{1,t-1})} \{H^t(y^{1,t-1}, y^t) - c^t(y^{1,t-1}, y^t)\},\ t = \overline{1,T}\ .$$

Обозначая $\tilde{s}^{t,t+x(t)}(\cdot)$ – предположения АЭ об управлении со стороны центра, формально задачу принятия АЭ решений можно записать в виде:

$$(3)\ y^{t,t+x(t)} = arg \max_{z^{t,t+V(t)} \in A_0^{t,t+V(t)}} \sum_{t=t}^{t+V(t)} \{ \tilde{s}^{\,t}(x^{1,t-1},\ \tilde{x}^{t}\ (x^{1,t-1}, z^{t,t}), y^{1,t-1}, z^t) - c^t(x^{1,t-1}, z^{t,t})\}.$$

Итак, АЭ имеет возможность манипулировать центром, влияя на «историю» игры (то есть, выбирая, например, $y^t \neq x^t$, побуждать центр выбрать в периоде $(t+1)$ план $\tilde{x}^{t+1}(x^{1,t-1}, y^t)$). При наличии ЭОР в общем случае действия АЭ, выбираемые им в соответствии с (3), не совпадают с планами, назначаемыми центром в соответствии с (2), то есть АЭ становится неуправляемым и может манипулировать центром[1].

Отметим, что при использовании центром системы стимулирования (1) ЭОР в рассматриваемой модели не возникает, так как выигрыш АЭ в каждом периоде не может быть строго положительным. Действительно, предположим, что АЭ в некотором периоде не выполнил план, тогда его выигрыш в этом периоде неположителен. Даже, если он этим изменил будущую плановую траекторию, то в следующих периодах, независимо от плановой

---

[1] *Отметим, что в рамках введенных предположений центр не может отразить (осознать и принять соответствующие меры) наличие ЭОР, так как это потребовало бы от него дальновидности не меньшей, чем у АЭ.*

траектории и независимо от выполнения или невыполнения плана, он также получит неположительный выигрыш.

Ситуация меняется, если стимулирование несвязанно, или если центр должен обеспечивать АЭ в каждом периоде строго положительную резервную полезность, зависящую от траектории реализаций, то есть обеспечивать выполнение $f^t(y^{1,t}) \geq \overline{U}^t$, $t = \overline{1,T}$, где $\{\overline{U}^t\}$ – набор резервных полезностей. В случае несвязанного стимулирования ЭОР исключается использованием центром планов, согласованных в смысле теоремы 2.

Таким образом, справедливо следующее утверждение.

Теорема 4. При связанном стимулировании или при несвязанном стимулировании с согласованными планами эффекта обмена ролями в ДАС не возникает.

ЭОР может приводить как к снижению эффективности управления, так и к ее увеличению (точнее, к увеличению значения целевой функции центра). Приводимый ниже пример иллюстрирует возможность возникновения ситуации, когда более дальновидный, чем центр, АЭ навязывает центру планы, которые выгодны им обоим с точки зрения суммарных по всем периодам выигрышей.

Пример 5. Пусть $T=2$, $A^1 = A_2 = \{0; 1\}$, а значения затрат, доходов и резервных полезностей равны приведенным таблице 2.

Табл. 2. Значения затрат и доходов в примере 5.

y^1	y^2	c^1	c^2	H^1	H^2	U^1	U^2
0	0	1	1	2	4	0	2
0	1	1	2	2	6	0	2
1	0	3	1	3	12	0	7
1	1	3	2	3	9	0	7

Пусть центр недальновиден. Тогда с его точки зрения оптимальна плановая траектория (0; 1), дающая ему за два периода выигрыш $K_1 = 3$. Если центр полностью дальновиден, то оптимальная плановая траектория есть (1; 0), дающая выигрыш $K_4 = 4$.

Если центр недальновиден, а АЭ полностью дальновиден, то АЭ в рамках ЭОР в первом периоде при плане $x^1 = 0$ выберет действие $y^1 = 1$, что заставит центр во втором периоде назначить план $x^2 = 0$. АЭ во втором периоде выберет действие, совпадающее с планом, что даст центру суммарный выигрыш $K_2 = 7$, что превышает и его выигрыш при отсутствии ЭОР, равный K_1, и его выигрыш в случае полной дальновидности[1]. Заметим, что и АЭ выгодно отклонение от плана, так как выполняя планы, назначенные недальновидным центром, он получает суммарный выигрыш, равный 2, а отклоняясь в первом периоде и выполняя план во втором периоде, он получает суммарный выигрыш, равный 4. Другими словами, АЭ ценой потери в первом периоде трех единиц полезности, навязывает центру стратегию выгодную им обоим, то есть компенсирующую потери АЭ от отклонения и обеспечивающую центру полезность, большую, чем при полной его дальновидности. •

Таким образом, ЭОР может возникать в случаях, когда горизонт дальновидности АЭ больше горизонта дальновидности центра[2]. Подробно исследовать теоретико-игровые модели управления организационными системами, в которых проявляется ЭОР, мы не будем, так как в них управление осуществляет не центр, на позиции которого стоит обычно исследователь операций, а АЭ.

---

[1] *Качественно данный эффект можно объяснить тем, что в первом периоде центр получает доход от выбора АЭ некоторого действия не неся при этом расходов на стимулирование.*

[2] *Если в некоторой организационной систем имеет место ЭОР, то с нормативной точки зрения исследователя операций необходимо изменять состав системы (назначать более дальновидный центр), а с точки зрения центра следует либо ограничить горизонты дальновидности управляемых субъектов (обсуждение соответствующих способов выходит за рамки настоящего исследования), либо изменить состав системы – заменив дальновидные АЭ на менее дальновидные (отметим, что мы не рассматриваем ситуацию жестких штрафов за невыполнение планов, что противоречило бы предположению о неотрицательности стимулирования), либо не обеспечивать резервной полезности, гарантируя АЭ в случая выполнения плана лишь нулевой выигрыш.*

Отказ от рассмотрения ЭОР позволяет исключить из дальнейшего анализа часть моделей (другую часть составляют системы, в которых АЭ использует скользящий режим ПР с обязательствами – см. ниже), для которых соответствующие ячейки в таблице 1 затенены (если у некоторой ячейки затенена половина, то это означает, что исключаются из рассмотрения те комбинации, в которых дальновидность АЭ превышает дальновидность центра).

Модель НТ-ДС. В данной модели у АЭ имеются две возможности: в период t выбрать действия, оптимальные с точки зрения текущего горизонта его дальновидности, и затем, либо следовать этим действиям в периодах $(t+1, t + x(t))$ – соответствующий режим принятия решений будем называть *скользящим режимом с обязательствами*, либо рассматривать эти действия как свой личный прогноз и оставлять за собой право при получении новой информации, например, в периоде $(t+1)$, выбирать другие действия – соответствующий режим принятия решений будем называть *скользящим режимом без обязательств*.

Скользящий режим без обязательств соответствует модели НТ-ДТ, то есть текущему режиму принятия решений АЭ, а скользящий режим с обязательствами может оказаться невыгодным АЭ по той причине, что, взяв обязательства, ставшие известными центру, на периоды $(t+1, t + x(t))$, он может оказаться в ситуации, когда центр установит на эти периоды стимулирование тождественно равное нулю (для центра это выгодно даже в текущем режиме принятия своих решений). Если же центр обязан в каждом периоде обеспечивать АЭ ненулевую полезность, то взятие АЭ обязательств на периоды вне горизонта дальновидности центра обязывает последнего оплачивать ему выбор соответствующих действий, то есть в этом случае имеет место ЭОР (см. также пример выше). Рассматривать подобные ситуации мы не будем по причинам, оговоренным выше.

Модель ДТ-НТ (ДАС2). Центр в каждом периоде сообщает АЭ систему стимулирования

$$(4)\quad s^t_K(y^{1,t}) = \begin{cases} c^t(x^{1,t}), & \text{если } y^{1,t} = x^{1,t} \\ 0, & \text{в остальных случаях} \end{cases}, \; t = \overline{1,T},$$

или систему стимулирования (1), и план x^t, а АЭ выбором действия y^t стремится максимизировать свой выигрыш в текущем периоде.

Очевидно, что выбор действия, совпадающего с планом, выгоден для АЭ, поэтому центру достаточно решить задачу выбора плановой траектории исходя из условия, что план каждого периода максимизирует суммарный выигрыш центра при текущем горизонте дальновидности:

$$(5)\ x^t = Proj^t\ arg \max_{y^{t,t+V_0(t)} \in A_0^{t,t+V_0(t)}} \sum_{t=t}^{t+V_0(t)} \{H^t(x^{1,t-1}, y^{t,t}) - c^t(x^{1,t-1}, y^{t,t})\},$$

$$t = \overline{1,T}\ .$$

Обозначим K_2 – эффективность стимулирования в модели ДАС2: $K_2 = F(s^{1,T}, x^{1,T})$, где $s^{1,T}$ удовлетворяет (4), а $x^{1,T}$ удовлетворяет (5). Можно привести примеры, когда K_2 оказывается как больше, так и меньше K_1 (см. пример 6). Исследование сравнительной эффективности моделей проводится ниже.

Модели ДТ-ДТ, ДТ-ДС. Если горизонты дальновидности и принятия решений у АЭ не превышают соответственно горизонтов дальновидности и принятия решений у центра, и АЭ не использует скользящего режима с обязательствами (см. обсуждение модели НТ-ДС выше), то получаем модель ДАС2. В остальных случаях (затененные половинки ячеек в строке ДТ таблицы 1) получаем «неуправляемую» систему с ЭОР, то есть ситуацию, не рассматриваемую по причинам, оговоренным выше.

Модель ДС-НТ (ДАС3). Центр в каждом периоде сообщает АЭ систему стимулирования (4) и план (5). Если центр использует скользящий режим без обязательств, то получаем модель ДТ-НТ. Поэтому интерес представляет случай, когда центр использует скользящий режим с обязательствами.

Обозначим K_3 – эффективность стимулирования в модели ДАС3 с обязательствами. Можно привести примеры, когда K_3 оказывается как больше, так и меньше K_1 и/или K_2 (см. пример 7). Исследование сравнительной эффективности моделей проводится ниже.

Модели ДС-ДТ, ДС-ДС. Если горизонты дальновидности и принятия решений у АЭ не превышают соответственно горизонтов дальновидности и принятия решений у центра, и АЭ не использует скользящего режима с обязательствами (см. обсуждение модели НТ-ДС выше), то получаем модель ДАС3. В остальных случаях (затененные половинки ячеек в строке ДС таблицы 1) получаем «неуправляемую» систему, то есть ситуации, не рассматриваемые по причинам, оговоренным выше.

Модель ДАС4. Предположим, что центр полностью дальновиден. Тогда оптимальной в соответствии с теоремой 1 является система стимулирования (4) со следующими планами:

(6) $x^{1,T} = arg \max\limits_{y^{1,T} \in A_0^{1,T}} \sum\limits_{t=1}^{T} \{H^t(y^{1,t}) - c^t(y^{1,t})\}$.

Отметим, во-первых, что модель ДАС4 подробно исследована выше (см. теоремы 1-3). Во-вторых, при полностью дальновидном центре[1] в рамках предположения А.0 не важна ни дальновидность АЭ, ни то, какой режим управления центр использует (текущий, скользящий, программный), ни наличие или отсутствие у центра обязательств – во всех случаях эффективность управления одинакова и равна $K_4 = F^{1,T}(x^{1,T}, s^{1,T})$, где $s^{1,T}$ удовлетворяет (4), а $x^{1,T}$ определяется как решение задачи (6). В третьих, очевидно, что в отсутствии ЭОР эффективность управления в модели ДАС4 максимальна, то есть выполнено:

(7) $K_4 \geq K_1, K_4 \geq K_2, K_4 \geq K_3$.

Описав четыре базовые модели ДАС, различающихся распределениями дальновидности и горизонтами принятия решений, перейдем к исследованию их сравнительной эффективности.

В качестве отступления отметим, что рассматриваемая в настоящей работе постановка задачи управления ДАС не является

---

[1] *Случай полной дальновидности центра является «идеальным» с точки зрения эффективности управления – при этом невозможен эффект обмена ролями и т.д. Однако, с точки зрения практики полная дальновидность является искусственным понятием – непонятно что считать максимальным горизонтом дальновидности в реальных системах. По-видимому, по мере удаления будущего от момента принятия решений увеличивается неопределенность (неопределенность будущего), поэтому определение максимального горизонта дальновидности должно зависеть от используемого способа устранения этой неопределенности.*

исчерпывающей, так как в ней не учитывается то, что отношение центра к выигрышу $F^t(y^{1,t})$ может изменяться в зависимости от в периода времени, когда он принимает решение. Формально можно ввести, следуя работам [36, 37, 78-80], понятие распределения дальновидности $\{d_t(t)\}$, такое что в любой момент времени t центр максимизирует $\sum_{t=t}^{M} d_t(t)\Phi^t(y^{1,t})$, где M выбирается в зависимости от дальновидности центра ($M = \min\{T, t+k\}$ для моделей ДАС2 и ДАС3, $M = T$ для ДАС4). Это означает, что при оценке сравнительной эффективности моделей ДАС1 – ДАС4 необходимо искать не только условия на функцию выигрыша центра $F^t(y^{1,t})$, но также и на распределение дальновидности, что, очевидно, существенно усложнит задачу.

В предложенной выше классификации (см. таблицу 1) в одну модель ДАС4 были, фактически, объединены случаи с полностью дальновидным центром вне зависимости от того, какой режим управления он использует: текущий, скользящий или программный. В общем случае такое объединение (без потери общности) не имеет места. В работе [80] был подробно рассмотрен случай влияния изменения распределения дальновидности на эффективность управления при полностью дальновидном центре, и выявлены условия на распределение дальновидности, при которых реализация и прогноз в каждом периоде совпадают, то есть когда эту общую задачу можно свести к нашей классификации. Приведем основные результаты:

1. Если для распределений дальновидностей $\{d_t\}$ и $\{d_t'\}$ задачи $\sum_{t=1}^{T} d_t F^t(y^{1,t}) \to \max_{y^{1,T} \in Y^{1,T}}$ и $\sum_{t=1}^{T} d_t' F^t(y^{1,t}) \to \max_{y^{1,T} \in Y^{1,T}}$ имеют одинаковые решения, тогда для распределения дальновидности $d_t'' = a d_t + b d_t'$, $t = \overline{1,T}$, $a, b \geq 0$ задача $\sum_{t=1}^{T} d_t'' F^t(y^{1,t}) \to \max_{y^{1,T} \in Y^{1,T}}$ имеет такое же решение. Таким образом, можно получить важный результат о том, что множество распределений дальновидно-

сти, которые дают одинаковое решение рассматриваемой задачи, является выпуклым конусом.

2. Если в каждом периоде функционирования центр определяет свою реализацию и прогноз из решения задачи $\sum_{t=t}^{T} d_t(t) F^t(y^{1,t}) \to \max_{y^{t,T} \in Y^{t,T}, y^{1,t-1} = x^{1,t-1}}$, существуют $\{d_t\}$ и $\{d_t'\}$, для которых решения задачи в первом периоде совпадают и распределение дальновидности центра в периоде t может быть представлено в виде $d_t(t) := a_t d_t + b_t d_t'$, $t = \overline{1,T}$, $a_t, b_t \geq 0$, то прогноз, сделанный центром в первом периоде, совпадает с реализацией в каждом из последующих периодов, и, соответственно, прогнозы, сделанные центром в последующих периодах совпадают с прогнозом, сделанным в первом периоде. Это означает, что, если в каждом периоде выбирается вектор распределения дальновидностей из определенного в первом пункте конуса, то реализация совпадает с планом, то есть вне зависимости от режима управления: текущего, скользящего или программного, центр получит одно и то же значение оптимальных планов.

3. Пусть $V = \{d \mid \sum_{t=1}^{T} d_t F^t(y^{1,t}) \to \max$ при $y^{1,T} = x^{1,T}\}$, то есть множество таких распределений дальновидности центра, что наилучшим планом, определенным в первом периоде, является $x^{1,T}$; $d(t) := \{d_t(t), d_{t+1}(t), \mathbf{K}, d_T(t)\}$. Если в первом периоде $d(1) \in V$, то для совпадения реализации и прогноза во всех периодах функционирования, начиная со второго, достаточно, чтобы для каждого распределения дальновидности $d(t)$ существовал бы вектор $\tilde{d}(t) = \{\tilde{d}_1(t), \mathbf{K}, \tilde{d}_{t-1}(t)\}$, $t = \overline{2,T}$, такой, что распределение дальновидности $\{\tilde{d}(t), d(t)\} \in V$.

4. В предыдущих пунктах описаны процедуры поиска наилучшего плана для одного распределения дальновидности $\{d_t\}$. Ответ на вопрос о том, как найти все множество распределений дальновидности (конус *V*), либо хотя бы часть его, для которого наилучший план такой же, дает следующее утверждение. Пусть

$x^{1,T}$ является решением задачи $\sum_{t=1}^{T} d_t F^t(y^{1,t}) \to \max_{y^{1,T} \in Y^{1,T}}$ для распределения дальновидности $\{d_t\}$, тогда, если для распределения дальновидности $\{d'_t\}$ выполняется соотношение: $\frac{d'_{t+1}}{d'_t} > \frac{d_{t+1}}{d_t}$, $t = \overline{1,T}$, и для любого $y^{1,T}$ выполняется условие $Y^t(y^{1,t-1}) \subseteq Y^t(x^{1,t-1})$, тогда $x^{1,T}$ является решением задачи с распределением дальновидности $\{d'_t\}$.

Завершив описание результатов, приведенных в [80], отметим, что при решении многих экономических задач полагают, что распределение дальновидности имеет специфический вид $d_t(t) = d^{t-t}$, где d является некоторой константой (так называемым коэффициентом дисконтирования – см. также выше). Оказывается, что в этом случае задача сводится к исходной с помощью замены $\tilde{\Phi}^t(y^{1,t}) = d^t \Phi^t(y^{1,t})$. Действительно, в каждый момент принятия решения t центр ищет максимум функции $\sum_{t=t}^{M} d^{t-t} F^t(y^{1,t})$, что эквивалентно отысканию максимума функции $\sum_{t=t}^{M} \tilde{F}^t(y^{1,t})$. Таким образом, ограничимся в дальнейшем постановкой задачи без учета зависимости распределения дальновидности от момента принятия решений.

Вернемся к оценке эффективности различных режимов управления и ГПР. Фиксируем некоторое распределение дальновидностей центра $x_0(t)$, $t = \overline{1,T}$, и будем исследовать эффективность режимов управления при этом распределении дальновидностей. Обозначим $L_0^{1,T}$ = ($L_0(1)$, $L_0(2)$, ..., $L_0(T)$) – ГПР центра (как отмечалось выше, $L_0(t)$ £ $x_0(t)$); $t_1 = 1$, $t_2 = t_1 + L_0(t_1)$, $t_3 = t_2 + L_0(t_2)$ и т.д. – моменты принятия решений центром в модели ДАС3 с обязательствами (как отмечалось выше, ДАС3 отличается от ДАС2 наличием обязательств), следовательно

$[t_i;\ t_{i+1}]$ – интервалы времени, на которые центр фиксирует планы в моменты времени t_i, $i = 1, 2, ..., i_{max}(L_0^{1,T}) - 1$, где i_{max}: $t_{i_{\max}} = T$.

Если, с учетом решения задачи согласованного стимулирования (см. теоремы 1-2), целевая функция центра имеет вид $F^t(y^{1,t}) = H^t(y^{1,t}) - c^t(y^{1,t})$, $t = \overline{1,T}$, то оптимальные в моделях ДАС1-ДАС4 плановые траектории $x_1^{1,T}$, $x_2^{1,T}$, $x_3^{1,T}$ и $x_4^{1,T}$, соответственно, определяются следующим образом[1]:

(8) $x_1^t = \tilde{x}_1^t(x_1^{1,t-1}) = arg \max\limits_{y^t \in A^t(x_1^{1,t-1})} F^t(x_1^{1,t-1}, y^t),\ t = \overline{1,T}$;

(9) $x_2^t = \tilde{x}_2^t(x_2^{1,t-1}) =$

$$= Proj^t\ arg \max_{y^{t,t+V_0(t)} \in A_0^{t,t+V_0(t)}} \sum_{t=t}^{t+V_0(t)} F^t(x_2^{1,t-1}, y^{t,t}),\ t = \overline{1,T};$$

(10) $x_3^{t_i,t_{i+1}} = arg \max\limits_{y^{t_i,t_{i+1}} \in A_0^{t_i,t_{i+1}}} F^t(x_3^{1,t_i-1},\ y^{t_i,t}),\ i = \overline{1, i_{\max}-1}$;

(11) $x_4^t = Proj^t\ arg \max\limits_{y^{1,T} \in A_0^{1,T}} \sum\limits_{t=1}^{T} F^t(y^{1,t}),\ t = \overline{1,T}$.

Рисунки 2-5 иллюстрируют последовательность принятия решений центром в моделях ДАС1-ДАС4 (черная точка обозначает горизонт дальновидности, стрелка – горизонт принятия решений с обязательствами).

---

[1] *В принципах планирования* (2), (5) (6), (8)-(11) *планы на текущий и будущий периоды* (*в зависимости от распределения дальновидности и горизонта принятия решений*) *определяются исходя из максимизации целевой функции центра в предположении, что действия АЭ в предыдущих периодах совпадали с планами. Как отмечалось выше, отказ от этого предположения, то есть зависимость будущих планов от наблюдаемой траектории реализаций, является эффективным средством борьбы с эффектом обмена ролями и т.д.*

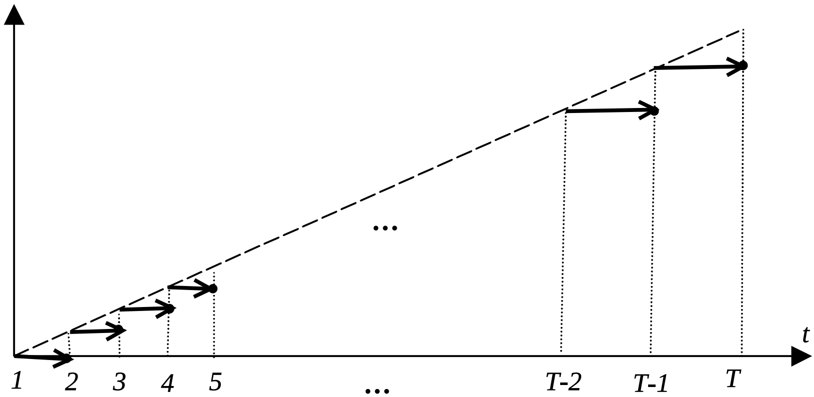

Рис. 2. Последовательность принятия решений центром в модели ДАС1

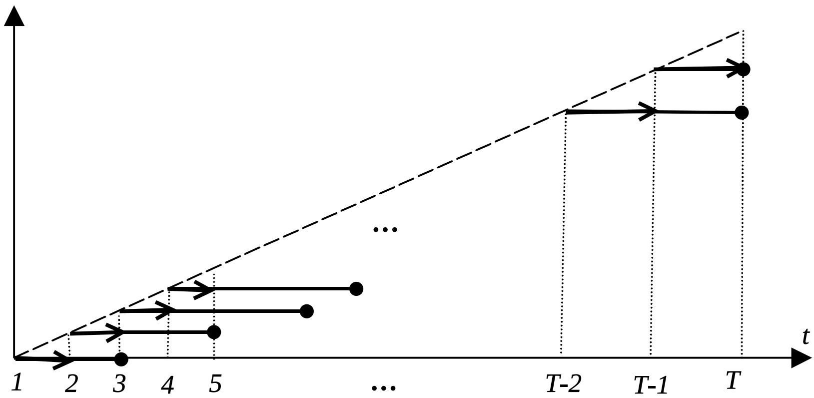

Рис. 3. Последовательность принятия решений центром в модели ДАС2

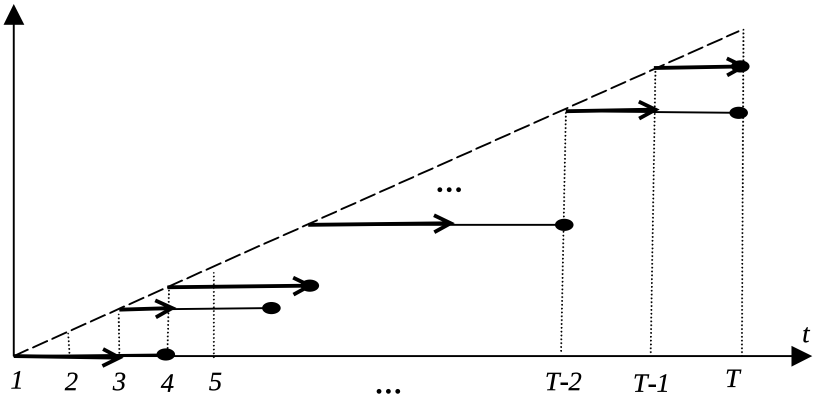

Рис. 4. Последовательность принятия решений центром в модели ДАС3

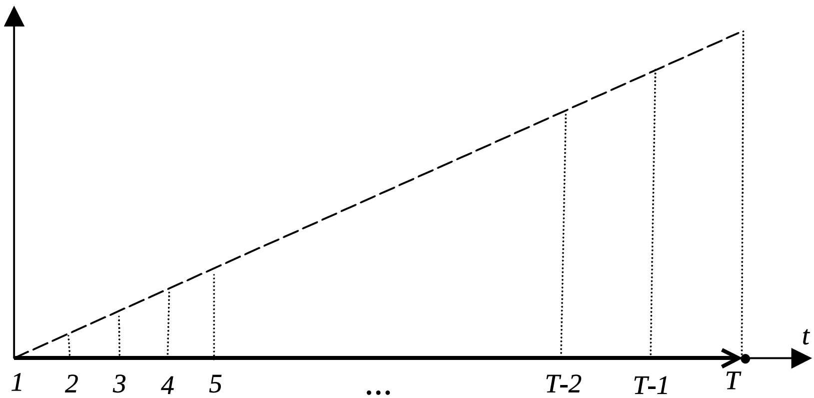

Рис. 5. Последовательность принятия решений центром в модели ДАС4

Обсудим специфику модели ДАС3. Пусть центр обладает фиксированной дальновидностью x_0 ($1 < x_0 < T$), принимает решения через каждые m_0 периодов, и фиксирует свои планы на L_0 периодов вперед. Условием того, что центр распланирует каждый период времени, является *1* £ m_0 £ L_0. Таким образом, предполагая что центр в каждый момент времени может принимать решения только на те периоды, которые лежат в пределах его дальновидности, получаем условие *1* £ m_0 £ L_0 £ x_0 < T (последнее неравенство отличает ДАС 3 от ДАС4).

Лемма 1. Пусть центр обладает фиксированной дальновидностью x_0 ($1 < x_0 < T$), принимает решения через каждые m_0 периодов, и фиксирует свои планы на L_0 периодов вперед (см. рисунки 6а) и 6б)). Такой способ принятия решения центром эквивалентен тому, что в первый период времени центр принимает и фиксирует план на L_0 периодов вперед с дальновидностью x_0, далее центр принимает и фиксирует решение на m_0 периодов вперед в моменты времени $L_0 + 1, L_0 + m_0, L_0 + 2\, m_0, \ldots, L_0 + n\, m_0,$ где $n = \left[\dfrac{T - L_0}{m_0}\right]$, с дальновидностью $x_0 - L_0 + m_0$ (см. рисунок 6б).

Доказательство. В первый момент принятия решений центр находит планы по следующей формуле:

$$x^{1,L_0} = \operatorname*{Proj}_{y^{1,L_0}} \arg \max_{y^{1,x_0} \in A_0^{1,x_0}} \sum_{t=1}^{x_0} \left[\Phi^t (y^{1,t})\right]$$

Опишем поведение центра в следующий момент m_0 принятия решения. Согласно описанной выше модели поведения ДАС3, в этот момент центр должен принять и зафиксировать решения на следующие L_0 периодов. Но так как в прошлый раз (в первый момент времени) он уже фиксировал план на L_0 первых периодов, а рассматривается момент принятия решения m_0, и план на $L_0 - m_0$ периодов вперед уже существует, то центр не имеет права его менять. Таким образом, в момент m_0 центр принимает и фиксирует план на m_0 периодов, начиная с $L_0 + 1$. Оптимальные планы находятся по следующей формуле (здесь и далее до окончания настоящего раздела в целях упрощения обозначений зависимость множеств допустимых действий от истории будет опускаться):

$$x^{L_0+1}, \mathbf{K}, x^{L_0+m_0} = \operatorname*{Proj}_{y^{L_0+1,L_0+m_0}} \arg \max_{y^{L_0+1,m_0+x_0} \in A^{L_0+1} \times \mathbf{L} \times A^{m_0+x_0}} \sum_{t=m_0}^{m_0+x_0} \Phi^t (x^{1,\min(t,L_0)},$$

$$y^{L_0+t}) = \operatorname*{Proj}_{y^{L_0+1,L_0+m_0}} \arg \max_{y^{L_0+1,m_0+x_0} \in A^{L_0+1} \times \mathbf{L} \times A^{m_0+x_0}} \{ \sum_{t=m_0}^{L_0} \Phi^t (x^{1,t}) +$$

$$
+\sum_{t=L_0+1}^{m_0+x_0} \Phi^t\,(x^{1,L_0}, y^{L_0+1,t})\} =
$$

$$
= \operatorname*{Proj}_{y^{L_0+1,L_0+m_0}} \arg \max_{y^{L_0+1,m_0+x_0} \in A^{L_0+1} \times \mathbf{L} \times A^{m_0+x_0}}
$$

$$
\sum_{t=L_0+1}^{L_0+(x_0+m_0-L_0)} \Phi^t\,(x^{1,L_0}, y^{L_0+1,t}).
$$

Последнее соотношение означает, что поведение центра в момент времени m_0 равносильно такому его поведению, при котором он принимает решения в момент $L_0 + 1$ на m_0 периодов вперед с дальновидностью $x_0 + m_0 - L_0$.

Аналогично можно показать, что задача оптимального выбора L_0 (фактически – выбора m_0) планов в периоде $n\,m_0$ (где n-целое и $n \le \dfrac{T - L_0}{m_0}$) для центра с дальновидностью x_0, эквивалентна задаче оптимального выбора m_0 планов в периоде $L_0 + (n - 1)\,m_0 + 1$ с дальновидностью $x_0 + m_0 - L_0$. Действительно:

$$
x^{L_0+(n-1)m_0+1}, \mathbf{K}, x^{L_0+nm_0} =
$$

$$
\operatorname*{Proj}_{y^{L_0+(n-1)m_0+1,L_0+nm_0}} \arg \max_{y^{L_0+(n-1)m_0+1,x_0+nm_0} \in A^{L_0+(n-1)m_0+1,x_0+nm_0}}
$$

$$
\sum_{t=nm_0}^{x_0+nm_0} \Phi^t\,(x^{1,\min(L_0+(n-1)m_0,t)}, y^{L_0+(n-1)m_0+1,t}) =
$$

$$
\operatorname*{Proj}_{y^{L_0+(n-1)m_0+1,L_0+nm_0}} \arg \max_{y^{L_0+(n-1)m_0+1,x_0+nm_0} \in A^{L_0+(n-1)m_0+1,x_0+nm_0}} \sum^{L_0+(n-1)m_0} \Phi^t\,(x^{1,t}) +
$$

$$
\sum_{t=L_0+(n-1)m_0+1}^{x_0+nm_0} \Phi^t\,(x^{1,L_0+(n-1)m_0}, y^{L_0+(n-1)m_0+1,t}) =
$$

$$
\operatorname*{Proj}_{y^{L_0+(n-1)m_0+1,L_0+nm_0}} \arg \max_{y^{L_0+(n-1)m_0+1,x_0+nm_0} \in A^{L_0+(n-1)m_0+1,x_0+nm_0}}
$$

$$
\sum_{t=L_0+(n-1)m_0+1}^{L_0+(n-1)m_0+(x_0+m_0-L_0)} \Phi^t\,(x^{1,L_0+(n-1)m_0}, y^{L_0+(n-1)m_0+1,t}). \bullet
$$

На рисунке 6, иллюстрирующем лемму 1, ромбиком обозначен момент принятия решения, жирной стрелкой – горизонт принятия решения (или на какие периоды принимаются решения в данный момент времени), жирной точкой обозначен горизонт дальновидности.

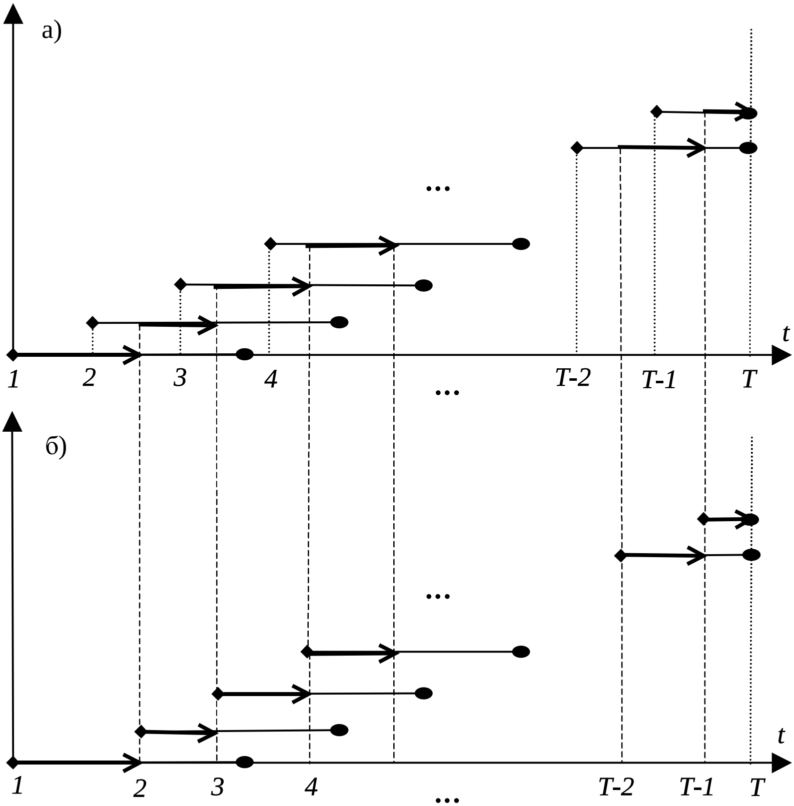

Рис. 6а), 6б). Принятие решений в модели ДАСЗ в соответствии с леммой 1

Завершив обсуждение специфики модели ДАСЗ, введем следующие функции[1]:

---

[1] *Отметим, что при переходе к целевым функциям вида* (*12*) *«автоматически» учитывается требование принадлежности плановой траектории соответствующей допустимой области* (*см. описание метода штрафов в* [59]), *что позволяет в* (*13*)-(*16*), *в отличие от* (*8*)-(*11*), *искать в каждом периоде максимумы по независимым от предыстории*

(12) $F^t(x^{1,t}) = \begin{cases} \sum\limits_{t=1}^{t} \Phi^t(x^{1,t}), & x^{1,t} \in A_0^{1,t} \\ -\infty, & x^{1,t} \notin A_0^{1,t} \end{cases}, \ t = \overline{1,T}$,

тогда плановые траектории (8)-(11) можно определить следующим образом (положим $F^0(\cdot) = 0$):

(13) $x_1^t = \tilde{x}_1^t(x_1^{1,t-1})$ Î $Arg \max\limits_{y^t \in A^t} [F^t(x_1^{1,t-1}, y^t) - F^{t-1}(x_1^{1,t-1})], \ t = \overline{1,T}$;

(14) $x_2^t = \tilde{x}_2^t(x_2^{1,t-1})$ Î

Î $Proj^t \ Arg \max\limits_{y^{t,t+V_0(t)} \in A^t \times A^{t+1} \times \ldots \times A^{t+x_0(t)}} [F^{t+x_0(t)}(x_2^{1,t-1}, y^{t,t+x_0(t)}) -$

$F^t(x_2^{1,t-1})], \ t = \overline{1,T}$;

(15) $x_3^{t_i, t_{i+1}}$ Î $Arg \max\limits_{y^{t_i,t_{i+1}} \in A^{t_i} \times A^{t_i+1} \times \ldots \times A^{t_{i+1}}} [F^{t_{i+1}}(x_3^{1,t_i-1}, y^{t_i,t_{i+1}}) -$

$- F^{t_i}(x_3^{1,t_i-1})], \ i = \overline{1, i_{\max}-1}$;

(16) x_4^t Î $Proj^t \ Arg \max\limits_{y^{1,T} \in A^{1,T}} F^T(y^{1,T}), \ t = \overline{1,T}$.

В соответствии с выражениями (13)-(16), эффективности управления в моделях ДАС1-ДАС4 можно записать в виде:

(17) $K_i = F^T(x_i^{1,T}), \ i = \overline{1,4}$.

Вернемся к сравнению эффективностей различных режимов управления в динамических АС.

Обозначим $J(t)$ – множество периодов, от которых зависит выигрыш в периоде t. В силу принципа причинности и введенных выше предположений " t Î $J(t)$ t £ t, $t = \overline{1,T}$. Положим также, что t Î $J(t)$, $t = \overline{1,T}$.

Обозначим $N(t)$ – множество периодов, выигрыши в которых зависят от стратегий, выбираемых в периоде t. В силу принципа причинности и введенных выше предположений " t Î $N(t)$ t ³ t, t Î $N(t)$, $t = \overline{1,T}$.

Множества $J(t)$ и $N(t)$ взаимозависимы:

допустимым множествам.

(18) $J(t) = \{t \le t \mid t \in N(t)\},\ N(t) = \{t \ge t \mid t \in J(t)\}, t = \overline{1,T}$.

Предположим, что существуют целые числа J и N не меньшие единицы и не большие T, такие, что

(19) $\forall\ t = \overline{1,T}\quad J(t) = \{\max\,(1, t - J); \dots; t\}$,

(20) $\forall\ t = \overline{1,T}\quad N(t) = \{t; \dots; \min\,(t + N, T)\}$.

Очевидно, что, если выполнено (18)-(20), то $J = N$.

Параметр J назовем *памятью АС* (точнее – памятью центра), так как он отражает максимальное число предыдущих периодов (исключая текущий), влияющих на выигрыш в текущем периоде.

Напомним, что выше были введены такие параметры центра как: $x_0(t)$ – его *дальновидность*, отражающая число будущих периодов (исключая текущий период), которые он принимает во внимание при выборе своей стратегии в текущем периоде (периоде t), и горизонт принятия решений $L_0(t)$, который в модели ДАСЗ соответствует числу будущих периодов (включая текущий период), на которые центр берет *обязательства* в текущем периоде.

Обозначим

(21) $x_0 = \min\limits_{t=\overline{1,T}} x_0(t),\ L_0 = \max\limits_{t=\overline{1,T}} L_0(t)$

и рассмотрим соотношение между памятью J, дальновидностью x_0 и обязательствами L_0. Введем следующее условие:

(22) $J + (L_0 - 1) \le x_0$.

Выполнение условия (22) можно назвать *принципом адекватности*[1] *для ДАС* (адекватности возможностей системы управления – центра – условиям функционирования и сложности управляемой системы), так как оно требует, чтобы в любой момент времени дальновидность центра, то есть его возможности по учету будущих последствий принимаемых решений, были не ниже суммы сложности системы (отражаемой ее памятью) и условий функционирования (отражаемых вынужденными обязательствами).

Принцип адекватности позволяет выявить условия, при которых взятие обязательств не изменяет эффективности управления – ниже приводится ряд формальных результатов.

---

[1] *См. аналоги и ссылки в* [51].

Теорема 5а. Если выполнены предположения А.0, А.1, А.2'', А.3, А.4 и условие (22), то в ДАС со связанным стимулированием режимы управления ДАС2 и ДАС3 эквивалентны: $K_2 = K_3$.

Теорема 5а является частным случаем формулируемой и доказываемой ниже теоремы 5б.

Если условия типа (22) не выполняются,[1] то существуют ДАС, в которых реализуются любые соотношения между эффективностями K_2 и K_3 (обоснованием справедливости этого утверждения являются приводимые ниже примеры 6 и 7).

Интуитивно можно было бы предположить, что ДАС1 должна обладать минимальной эффективностью, далее должна была бы следовать ДАС3 (дальновидность увеличилась по сравнению с ДАС1, но имеются обязательства), затем – ДАС2 (отказ от обязательств), и, наконец, ДАС4. То, что ДАС4 обладает максимальной (среди базовых четырех ДАС) эффективностью очевидно. Однако, оказывается, что возможны любые соотношения между эффективностями ДАС1 и ДАС2, а также ДАС2 и ДАС3. Ниже приводятся примеры, иллюстрирующие противоречия "здравому смыслу": в примере 6 рассматривается модель ДАС, в которой эффективность ДАС1 выше, чем ДАС2 (то есть увеличение дальновидности не приводит к увеличению эффективности), а в примере 7 – модель ДАС, в которой эффективность ДАС3 выше, чем ДАС2 (наличие обязательств приводит к повышению эффективности).

Пример 6. (эффективность ДАС1 выше эффективности ДАС2).

Рассмотрим трехпериодную модель, в которой человек (например, чиновник) выбирает свою судьбу – быть ли ему богатым, но брать взятки, или не купаться в роскоши, но быть честным. Чиновник имеет два возможных действия: "Воровать" или работать честно ("Не воровать"). Во все три периода у него для выбора есть эти два действия.

В первом периоде, если он выбирает "Не воровать", то его полезность $\Phi^1(y^1)$ равна 3. Если он выбирает действие "Воро-

---

[1] *Для этого достаточно нарушения принципа адекватности в одном периоде.*

вать", то в этом периоде его полезность равна 1, то есть меньше, чем если бы он не воровал из-за угрызений совести, которые он испытывает:

$$\Phi^1(\text{"Не воровать"}) = 3,$$

$$\Phi^1(\text{"Воровать"}) = 1.$$

Во втором периоде полезность $\Phi^2(y^1, y^2)$ зависит как от действий, выбранных во втором периоде, так и от действий, выбранных в первом периоде и равна:

$$\Phi^2(\text{"Не воровать"}, \text{"Не воровать"}) = 3,$$

$$\Phi^2(\text{"Воровать"}, \text{"Не воровать"}) = 3,$$

$$\Phi^2(\text{"Не воровать"}, \text{"Воровать"}) = 1,$$

$$\Phi^2(\text{"Воровать"}, \text{"Воровать"}) = 15.$$

Таким образом, если чиновник брал взятки и в первом, и во втором периоде, то он получает очень большую полезность по сравнению с тем, если бы он был честным оба периода.

В третьем периоде полезность, вне зависимости от выбранного действия y^3, зависит только от действия, которое чиновник выбрал в первом периоде – $\Phi^3(y^1)$:

$$\Phi^3(\text{"Не воровать"}) = 3,$$

$$\Phi^3(\text{"Воровать"}) = -100.$$

Эта ситуация означает, что вора однозначно выявляют в третьем периоде и, например, сажают в тюрьму. Таким образом, если чиновник выбирает "Воровать" в первом периоде, это означает что в третьем периоде он сядет в тюрьму, то есть понесет ущерб несравнимо больший, чем он бы заработал за первые два периода.

Рассмотрим ДАС1 – случай недальновидного чиновника. В первом периоде он выберет "Не воровать", во втором – "Не воровать", а в третьем – все равно какую стратегию он выберет, всего за три периода он получит полезность равную 3 + 3 + 3 = 9.

В модели ДАС2 – дальновидный чиновник с дальновидностью 2 в первом периоде видит, что если он будет воровать в этом периоде, то во втором он получит за это много большую полезность, чем если бы он был честным. Таким образом, в первом периоде он выбирает "Воровать", во втором периоде он уже "видит" третий период и понимает, что сделал неверный шаг, но уже поздно, и выбирает "Воровать". В третьем периоде все равно что он выбирает, в любом случае он садится в тюрьму, и его суммарная полезность за три периода равна 1 + 15 – 100 = – 84.

Итак в рассматриваемом примере ДАС1 обладает более высокой эффективностью, чем ДАС2. •

Пример 7. (эффективность ДАС3 выше эффективности ДАС2). Рассмотрим ДАС, в которой $T = 4$, а множество допустимых действий в каждом периоде содержит две альтернативы $A^t = \{0; 1\}$, $t = \overline{1,4}$. Следовательно, возможны шестнадцать траекторий – выигрыши центра в каждый момент времени приведены в узлах дерева на рисунке 6в (в квадратных скобках жирным шрифтом для каждой траектории приведены суммарные по всем четырем периодам выигрыши).

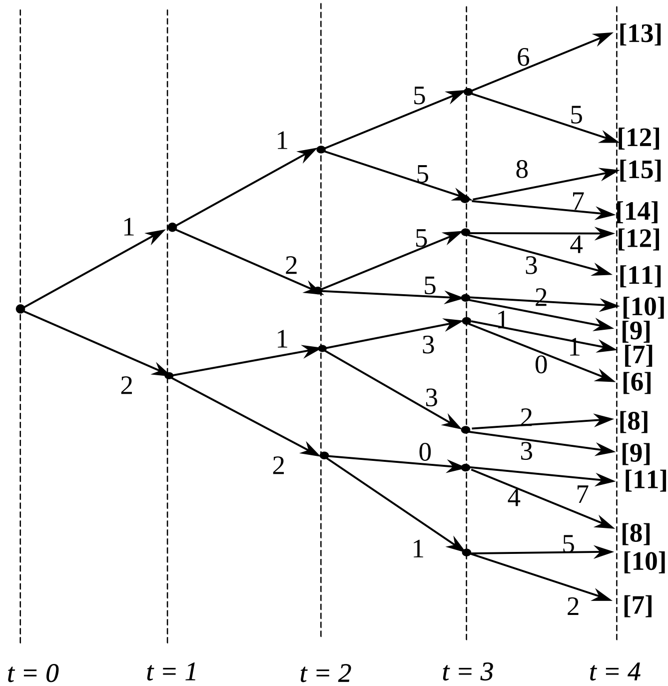

Рис. 6в. Выигрыши центра в примере 7

Пусть центр обладает дальновидностью $x_0 = 2$ и может брать обязательства на один будущий период. Тогда в модели ДАС4 (полная дальновидность – программное управление) оптимален план $x_4^* = (1; 1; 0; 1)$ (будем считать, что ноль соответствует движению вниз, а единица – движению вверх), а эффективность равна $K_4 = 15$. В модели ДАС1 (недальновидность – текущее управление) оптимален план $x_1^* = (0; 0; 0; 1)$, а эффективность равна $K_1 = 10$.

Легко видеть, что наличие обязательств (в модели ДАС3 оптимален план $x_3^* = (0; 0; 1; 1)$, в модели ДАС2 оптимален план $x_2^* = (0; 1; 0; 0)$) выгодно для центра, так как $K_3 = 11 > K_2 = 9$.

Таким образом, в данном примере имеет место следующее соотношение между эффективностями различных режимов управления:

$$K_4 > K_3 > K_1 > K_2. \bullet$$

Примеры 6 и 7 свидетельствуют, что в **общем случае возможны любые соотношения между эффективностями ДАС1, ДАС2 и ДАС3 – единственная априорная оценка:**

$$\mathbf{K_4 \;^{3}\; \max \{K_2, K_3\} \;^{3}\; K_1}.$$

Для упорядочения режимов управления по эффективности необходимо вводить определенные предположения, либо на взаимосвязь между периодами – см. теоремы 5а-5в, либо на монотонное увеличение информированности центра с ростом его дальновидности – см. теорему 6.

Вернемся к обсуждению результата теоремы 5а, который справедлив и для бесконечного T.

Следствие 2. Взятие центром обязательств на max $\{1, x_0 - J + 1\}$ периодов (включая текущий период) не снижает эффективности управления.

Следствие 3. Принимать решения центру следует не реже, чем каждые $\max \{1, x_0 - J + 1\}$ периодов[1].

Выражения (19) и (20), во-первых, означают, что память постоянна (не зависит от номера периода), а, во-вторых, что отсутствуют «разрывы» в прошлом, то есть, если некоторый период оказывает влияние на выигрыш в текущем периоде, то и все последующие (лежащие между ним и текущим) периоды также оказывают влияние на текущий период. Кроме того, в (22) фигурируют гарантированные оценки дальновидности и обязательств (см. условие (21)). Поэтому результат теоремы 5а может быть обобщен (условия (19)-(21) ослаблены) на случай переменных памяти, обязательств и дальновидности следующим образом.

Введем следующие величины:

(23) $J^{-}(t) = \min \{t \;\hat{I}\; J(t)\}, \ t = \overline{1,T}$,

---

[1] *Результаты следствия 2 и леммы 1 позволяет для данной ДАС ввести эквивалентную ДАС с меньшим числом периодов принятия решений, и разрабатывать для последней аналоги метода динамического программирования.*

(24) $N^+(t) = \max \{\tau \in N(t)\},\ t = \overline{1,T}$.

<u>Теорема 5б.</u> Если выполнены предположения А.0, А.1, А.2’’, А.3, А.4 и любое из следующих условий

(25) $\min\limits_{\tau > t + x_0(t)} J^-(\tau) > t + L_0(t) - 1,\ t = \overline{1,T}$,

(26) $\max\limits_{t \le \tau \le t + L_0(t) - 1} N^+(\tau) \le x_0(t),\ t = \overline{1,T}$,

то в ДАС со связанным стимулированием режимы управления ДАС2 и ДАС3 эквивалентны: $K_2 = K_3$.

Доказательство. Во-первых, заметим, что теорема 5а является частным случаем настоящей теоремы, так как, если выполнено (19)-(21), то из (25) с учетом (23) следует (22). Кроме того, из (18), (23) и (24) следует, что условия (25) и (26) эквивалентны, то есть доказательство можно проводить либо пользуясь одним из них, либо обеими условиями независимо.

Запишем определение (14) планов, выбираемых в ДАС2, для периодов от t до $t + L_0(t) - 1$:

(27) $x_2^t \in Proj^t Arg \max\limits_{y^{t,t+x_0(t)} \in A^{t,t+x_0(t)}} \sum\limits_{\tau = t}^{t+x_0(t)} \Phi^\tau (y^{J^-(\tau),\tau})$

. . .

(28) $x_2^{t+L_0(t)-1} \in Proj^{t+L0(t)-1}$

$$Arg \max\limits_{y^{t+L_0(t)-1,t+L_0(t)-1+x_0(t+L_0(t)-1)} \in A^{t+L_0(t)-1,t+L_0(t)-1+x_0(t+L_0(t)-1)}} \sum\limits_{\tau = t+L_0(t)-1}^{t+L_0(t)-1+x_0(t+L_0(t)-1)} \Phi^\tau (y^{J^-(\tau),\tau}).$$

Запишем определение (15) планов, выбираемых в ДАС3, для периодов от t до $t + L_0(t) - 1$:

(29) $x_3^t \in Proj^t Arg \max\limits_{y^{t,t+x_0(t)} \in A^{t,t+x_0(t)}} \sum\limits_{\tau = t}^{t+x_0(t)} \Phi^\tau (y^{J^-(\tau),\tau})$

. . .

(30) $x_3^{t+L_0(t)-1} \in Proj^{t+L0(t)-1} Arg \max\limits_{y^{t,t+x_0(t)} \in A^{t,t+x_0(t)}} \sum\limits_{\tau = t}^{t+x_0(t)} \Phi^\tau (y^{J^-(\tau),\tau})$.

Докажем, что в рамках условий (25) или (26) планы (28) и (30) совпадают (аналогично можно доказать и совпадение других планов из рассматриваемого временного промежутка).

Идея доказательства заключается в следующем: разобьем максимизацию в (28) на вычисление максимумов по множествам от $t + L_0(t) - 1$ до $t + x_0(t)$ и от $t + x_0(t) + 1$ до $t + L_0(t) - 1 + x_0(t + L_0(t) - 1)$, а в (30) – от t до $t + L_0(t) - 2$ и от $t + L_0(t) - 1$ до $t + x_0(t)$, а затем воспользуемся (25)-(26).

Фиксируем произвольный момент времени t. Из (25) следует, что всегда имеет место

$$t + L_0(t) - 1 + x_0(t + L_0(t) - 1) \geq t + x_0(t) \geq t + L_0(t) - 1.$$

Запишем (28) в виде:

$$(31)\quad x_2^{t+L_0(t)-1} \in Proj^{t+L0(t)-1} Arg \max_{y^{t+L_0(t)-1,t+V_0(t)} \in A^{t+L_0(t)-1,t+x_0(t)}} \max_{y^{t+x_0(t)+1,t+L_0(t)-1+x_0(t+L_0(t)-1)} \in A^{t+x_0(t)+1,t+L_0(t)-1+x_0(t+L_0(t)-1)}}$$
$$[\sum_{t=t+L_0(t)-1}^{t+x_0(t)} \Phi^t(y^{J^-(t),t}) + \sum_{t=t+x_0(t)+1}^{t+L_0(t)-1+x_0(t+L_0(t)-1)} \Phi^t(y^{J^-(t),t})] =$$
$$= Proj^{t+L0(t)-1} Arg \max_{y^{t+L_0(t)-1,t+V_0(t)} \in A^{t+L_0(t)-1,t+x_0(t)}} [\sum_{t=t+L_0(t)-1}^{t+x_0(t)} \Phi^t(y^{J^-(t),t}) +$$
$$+ \max_{y^{t+x_0(t)+1,t+L_0(t)-1+x_0(t+L_0(t)-1)} \in A^{t+x_0(t)+1,t+L_0(t)-1+x_0(t+L_0(t)-1)}} \sum_{t=t+x_0(t)+1}^{t+L_0(t)-1+x_0(t+L_0(t)-1)} \Phi^t(y^{J^-(t),t})].$$

Запишем (30) в виде

$$(32)\quad x_3^{t+L_0(t)-1} \in Proj^{t+L0(t)-1} Arg \max_{y^{t,t+L_0(t)-2} \in A^{t,t+L_0(t)-2}} \max_{y^{t+L_0(t)-1,t+x_0(t)} \in A^{t+L_0(t)-1,t+x_0(t)}} [\sum_{t=t}^{t+L_0(t)-2} \Phi^t(y^{J^-(t),t}) +$$
$$+ \sum_{t=t+L_0(t)-1}^{t+x_0(t)} \Phi^t(y^{J^-(t),t})] =$$

$$Proj^{t+L0(t)\text{-}1}\ Arg \max_{y^{t,t+L_0(t)-2} \in A^{t,t+L_0(t)-2}} \Big[\sum_{t=t}^{t+L_0(t)-2} \Phi^t (y^{J^-(t),t}) +$$

$$+ \max_{y^{t+L_0(t)-1,t+x_0(t)} \in A^{t+L_0(t)-1,t+x_0(t)}} \sum_{t=t+L_0(t)-1}^{t+x_0(t)} \Phi^t (y^{J^-(t),t}) \Big].$$

Заметим, что в силу (25)-(26) второе слагаемое в (31) не зависит явным образом от плана, выбираемого центром в периоде $t + L_0(t) - 1$, а первое слагаемое в (31) совпадает со вторым слагаемым в (32). Таким образом, планы (28) и (30) совпадают, следовательно эффективности режимов ДАС2 и ДАС3 одинаковы. •

Рассмотрим случай, в котором множество периодов функционирования ДАС может быть разбито на набор непересекающихся и «невзаимодействующих» подмножеств. Для этого необходимо формализовать понятие «взаимодействия».

Предположим, что существует разбиение множества $\{1, 2, \dots T\}$ на подмножества $[t_i, t_{i+1}]$, где границы подмножеств $\{t_i\}^q_{i=1}$, такие, что 1 £ q £ T, t_i ³ t_{i-1}, t_1 ³ 1, t_q £ T и для этого разбиения выполняется

(33) $x_0(t_i)$ ³ $t_{i+1} - t_i$, $i = \overline{1, q-1}$,

(34) " t Î $[t_i; t_{i+1}]$ $N^+(t)$ £ t_{i+1}.

<u>Теорема 5в.</u> Если выполнены предположения А.0, А.1, А.2’’, А.3, А.4 и условия (33)-(34), то в ДАС со связанным стимулированием все режимы управления[1] эквивалентны: $K_2 = K_3 = K_4$.

Доказательство. Справедливость утверждения теоремы следует из принципа оптимальности Беллмана и того, что в силу (33)-(34) имеет место: " t Î $[t_i; t_{i+1}]$ $J(t)$ Ç $\{1, \dots, t_i - 1\}$ = Æ. •

Содержательно, условия (33)-(34) означают, что в моменты $\{t_i\}^q_{i=1}$ «рвется связь времен», то есть множество всех периодов (от единицы до T) может быть разбито на q подмножеств, таких, что внутри каждого из них центр полностью дальновиден – см. условие (33), и решения, принимаемые внутри периодов, принадлежащих любому подмножеству, не влияют на выигрыши в пе-

[1] *За исключением, естественно, скользящего режима управления (ДАС1), при котором дальновидность центра равна единице (в этом случае ДАС2 и ДАС3 совпадают с ДАС1).*

риодах, принадлежащих другим подмножествам – см. условие (34).

Результат теоремы 5 позволяет сравнивать различные режимы управлений по эффективности и, в частности, дает ответ на вопрос о том – в каких случаях взятие обязательств не снижает эффективности управления. Тем не менее, сам факт того, что наличие обязательств может приводить не только к не снижению, но и к повышению (см. пример 7) эффективности управления, представляется несколько удивительным и противоречащим здравому смыслу. Качественное объяснение этого факта таково – так как в рассматриваемой модели ДАС неопределенность будущего заключается в п о л н о м незнании функций выигрыша вне горизонта дальновидности, то любое принятое решение может оказаться как эффективным, так и неэффективным с точки зрения значений функций выигрыша в некоторых будущих периодах. Для того, чтобы исключить подобные явления необходимо ввести предположения о "монотонности" функций выигрыша, которое исключало бы возможность резких и непредвиденных ее изменений. Приведем формальные определения.

Пусть функция выигрыша в периоде t зависит от истории $y^{1,t-1}$, действия y^t в этом периоде и неопределенного параметра r^t Î W^t, то есть $F^t(y^{1,t}, r^t)$, $t = \overline{1,T}$. Информированность центра (ту информацию, которой он обладает о неопределенном параметре) будем описывать совокупностью множеств $W(t, t)$ Í W^t, t ³ t, отражающих его знание в момент времени t о возможных значениях неопределенного параметра в настоящий ($t = t$) и будущие ($t = \overline{t+1,T}$) моменты времени.

Неопределенность будущего будем отражать следующим условием:

(35) " $t = \overline{1,T}$, " t_1 £ t_2 £ t $W(t_2, t)$ Í $W(t_1, t)$ Í W^t.

Из (35) следует, что

(36) " $t = \overline{1,T}$, " t £ t_1 £ t_2 $W(t, t_1)$ Í $W(t, t_2)$ Í Ω^{t_2}.

Содержательная интерпретация условий (35)-(36), которые будем называть *условиями монотонности*, заключается в том, что по мере удаления (приближения) рассматриваемого момента времени от оцениваемого, и наоборот, неопределенность не уменьшается (не увеличивается).

Введем критерий сравнения неопределенностей. Будем говорить, что, в первой ситуации, которой соответствует информированность $W_1(\cdot)$, центр более информирован (неопределенность меньше), чем во второй ситуации, которой соответствует информированность $W_2(\cdot)$, если выполнено

(37) $'' \ t = \overline{1,T}\ , \ '' \ t = \overline{t,T} \quad W_1(t,\ t) \ Í \ W_2(t,\ t).$

Введем также предположение о том, что центр при принятии решений устраняет неопределенность, ориентируясь на максимальный гарантированный в рамках своей информированности результат. В рассматриваемой модели дальновидность, как таковая, отсутствует[1], а текущий режим управления совпадает со скользящим. В отсутствии обязательств в периоде t центр при известной истории $x^{1,t-1}$ решает задачу определения плана на текущий период:

(38) $x^t = Proj^t \operatorname{Arg} \max\limits_{y^{1,T} \in A_0^{1,T}} \sum\limits_{t=t}^{T} \min\limits_{r^t \in \Omega(t,t)} F^{\,t}(x^{1,t}, r^t),\ t = \overline{1,T}\ .$

В присутствии обязательств в периоде t центр при известной истории $x^{1,t-1}$ решает задачу определения планов на текущий период и на горизонт принятия решений, состоящий из $L_0(t) - 1$ будущих периодов:

(39) $x^{t,t+L0(t)-1} = Proj^{t,t+L0(t)-1} \operatorname{Arg} \max\limits_{y^{t,T} \in A_0^{t,T}} \sum\limits_{t=t}^{T} \min\limits_{r^t \in \Omega(t,t)} F^{\,t}(x^{1,t-1}, r^t),$

где $t = 1,\ 1 + L_0(1),\ \ldots$.

Эффективность управления определяется значением суммарного по всем периодам гарантированного априори выигрыша от траекторий (38) – в отсутствии обязательств, или (39) – в присутствии обязательств, то есть

---

[1] *Следовательно, при фиксированной информированности бессмысленно говорить о режиме ДАС4 (то есть о полной дальновидности), но можно условно считать, что более информированный центр обладает большей дальновидностью.*

(40) $K(y^{1,T}) = \sum_{t=1}^{T} \min_{r^t \in \Omega^t} \Phi^t(y^{1,t}, r^t)$.

Теорема 6. Если выполнены предположения А.0, А.1, А.2’’, А.3, А.4 и условие монотонности, то в ДАС со связанным стимулированием:

а) с ростом неопределенности (в смысле (37)) эффективность управления не увеличивается;

б) взятие обязательств не увеличивает эффективности управления.

Справедливость утверждения теоремы 6 следует из (38)-(40) с учетом (35)-(37).

Таким образом, результаты настоящего раздела дают возможность сравнивать эффективности различных режимов управления ДАС, в том числе – выгодность взятия обязательств. В то же время, во многих моделях реальных АС, например, условия (25) или (26), или условия монотонности и т.д., не выполнены, поэтому с тем, чтобы, учитывая результаты примеров 6-7 и др., разобраться в качественной специфике влияния дальновидности и обязательств на эффективность управления ДАС, рассмотрим частные модели, а именно ДАС, функционирующие в течение двух и трех периодов (аналогом теоремы 5 для которых являются соответственно теоремы 7 и 8), ДАС с накоплением и др.

4. ДВУХПЕРИОДНЫЕ И ТРЕХПЕРИОДНЫЕ ДИНАМИЧЕСКИЕ АКТИВНЫЕ СИСТЕМЫ

В настоящем разделе рассматриваются частные случаи общей модели ДАС, а именно – двухпериодная ДАС и трехпериодная ДАС (см. обоснование необходимости их рассмотрения в конце предыдущего раздела), на примере которых анализируется сравнительная эффективность различных режимов управления.

Рассмотрим *двухпериодную ДАС*, то есть динамическую АС, функционирующую в течение двух периодов ($T = 2$).

В модели ДАС 1 в первом периоде центр решает задачу планирования[1]:

(1) $x_1^1 = arg \max\limits_{y^1 \in A^1} \{H^1(y^1) - c^1(y^1)\}$,

и назначает систему стимулирования

(2) $s_K^1(x_1^1, y^1) = \begin{cases} c^1(x_1^1), & y^1 = x_1^1 \\ 0, & \text{в остальных случаях} \end{cases}$.

Во втором периоде в модели ДАС1 решается задача планирования:

(3) $x_1^2 = arg \max\limits_{y^2 \in A^2(x_1^1)} \{H^2(x_1^1, y^2) - c^2(x_1^1, y^2)\}$,

и назначается система стимулирования

(4) $s_K^2(x_1^2, y^1, y^2) = \begin{cases} c^2(y^1, x_1^2), & y^2 = x_1^2 \\ 0, & \text{в остальных случаях} \end{cases}$.

Как отмечалось выше, использование системы стимулирования (4) исключает ЭОР в моделях НТ-ДТ и др.

Так как рассматривается двухпериодная ДАС, то дальновидность центра эквивалентна его полной дальновидности (то есть модели ДАС 2, ДАС 3 и ДАС 4 в случае двух периодов эквивалентны) и независимо от режима управления (программного или скользящего, с обязательствами или без них) оптимальны планы

(5) $(x_4^1,\ x_4^2) = arg \max\limits_{y^1 \in A^1,\, y^2 \in A^2(y^1)} \{H^1(y^1)+H^2(y^1, y^2)-c^1(y^1)-c^2(y^1, y^2)\}$

и системы стимулирования вида (2), (4) с планами (5). Сравнивая эффективности, получаем, что $K_1 = F^2(x_1^2,\ x_1^2) \le F^2(x_4^1,\ x_4^2) = K_4$.

Таким образом, мы обосновали следующий достаточно очевидный вывод:

Теорема 7. В двухпериодных ДАС имеет место $K_1 \le K_2 = K_3 = K_4$, то есть увеличение дальновидности центра и/или использование обязательств в скользящем режиме управления не снижает эффективности управления.

---

[1] *Напомним, что верхний индекс обозначает номер периода, а нижний – номер модели.*

Качественно, в двухпериодных системах еще не проявляются все эффекты, характерные для ДАС (см. теоремы 5 и 7). Поэтому перейдем к рассмотрению трехпериодной ДАС ($T = 3$).

В модели ДАС 1 в первом периоде центр решает задачу планирования (1) и назначает систему стимулирования (2). Во втором периоде решается задача планирования (3) и назначается система стимулирования (4). В третьем периоде решается задача планирования

(6) $x_1^3 = arg \max\limits_{y^3 \in A^3(x_1^1,\, x_1^2)} \{H^3(x_1^1,\ x_1^2, y^3) - c^3(x_1^1,\ x_1^2, y^3)\}$,

и назначается система стимулирования[1]

(7) $s_K^3(x_1^3, y^1, y^2, y^3) = \begin{cases} c^3(y^1, y^2, x_1^3), & y^3 = x_1^3 \\ 0, & \textit{в остальных случаях} \end{cases}$.

Как отмечалось выше, использование систем стимулирования (4), (7) исключает ЭОР в моделях НТ-ДТ и др.

Модели ДАС 2 и ДАС 3 в трехпериодной ДАС различаются планами, назначаемыми центром во втором и третьем периодах. Это объясняется тем, что дальновидность, но не полная дальновидность, центра означает, что при принятии решений в первом периоде он учитывает свои полезности за первый и второй период, а при принятии решений во втором периоде – за второй и третий периоды. Следовательно, в ДАС 2 во втором и третьем периодах центр назначает планы, которые оптимальны с точки зрения суммы выигрышей по соответствующим двум периодам:

(8) $x_2^1 = Proj^1 \, arg \max\limits_{y^1 \in A^1,\ y^2 \in A^2(y^1)} \{H^1(y^1) + H^2(y^1, y^2) - c^1(y^1) - c^2(y^1, y^2)\}$,

(9) $x_2^2 = Proj^2 \, arg \max\limits_{y^2 \in A^1(x_2^1),\ y^3 \in A^3(x_1^2,\, y^2)} \{H^2(x_2^1, y^2) + H^3(x_2^1, y^2, y^3) - c^2(x_2^1, y^2) - c^3(x_2^1, y^2, y^3)\}$,

(10) $x_2^3 = arg \max\limits_{y^3 \in A^3(x_2^1,\, x_2^2)} \{H^3(x_2^1,\ x_2^2, y^3) - c^3(x_2^1,\ x_2^2, y^3)\}$.

---

[1] *Повторяться и выписывать системы стимулирования, оптимальные в трехпериодных ДАС 2, ДАС 3 и ДАС 4 мы не будем.*

В ДАС 3 с обязательствами[1] (ДАС 3 без обязательств совпадает с ДАС 2) план в первом периоде определяется (8), план на второй период фиксируется в первом периоде:

$$(11)\quad x_3^2 = Proj^2 \, arg \max_{y^1 \in A^1,\, y^2 \in A^2(y^1)} \{H^1(y^1) + H^2(y^1, y^2) - c^1(y^1) - c^2(y^1, y^2)\},$$

а план на третий период фиксируется во втором периоде:

$$(12)\quad x_2^3 = arg \max_{y^3 \in A^3(x_3^1,\, x_3^2)} \{H^3(x_3^1, x_3^2, y^3) - c^3(x_3^1, x_3^2, y^3)\}.$$

В ДАС 4 оптимальны планы

$$(13)\quad (x_4^1, x_4^2, x_4^3) = arg \max_{y^1 \in A^1,\, y^2 \in A^2(y^1),\, y^3 \in A^3(y^1, y^2)} \{H^1(y^1) + H^2(y^1, y^2) + H^3(y^1, y^2, y^3) - c^1(y^1) - c^2(y^1, y^2) - c^3(y^1, y^2, y^3)\}.$$

Объединяя (1), (3), (6), (8)-(13), получим:

$$(14)\quad K_1 = F^3(x_1^1, x_1^2, x_1^3),\ K_2 = F^3(x_2^1, x_2^2, x_2^3),\ K_3 = F^3(x_3^1, x_3^2, x_3^3),\ K_4 = F^3(x_4^1, x_4^2, x_4^3).$$

Сравним эффективности управления ДАС2 и ДАС3 в трехпериодной модели. Для этого введем следующую *гипотезу о консервативности центра*: если при решении задачи оптимального планирования при различных режимах управления центр получит в обоих случаях одинаковые множества оптимальных планов, то из этих множеств он в обоих случаях выберет одинаковые планы.

<u>Лемма 2.</u> Если дальновидность центра $x_0 = T - 1$, и выполнена гипотеза о консервативности центра, то $K_2 \geq K_3$.

Доказательство. Задачи максимизации, которые решает центр, при выборе плана в первом периоде для моделей ДАС2 и ДАС3 совпадают. Это значит что оптимальный план на первый период для обоих этих моделей будет одним и тем же, при условии выполнения гипотезы о консервативности центра. Во втором периоде в модели ДАС2 в силу $x_0 = T - 1$ центру известны функции полезности во всех будущих периодах, поэтому он может выбрать планы которые максимизируют сумму его полезности за

[1] *В трехпериодной ДАС обязательства центра имеют место только относительно второго периода.*

последние $T – 1$ периодов в условиях «полной дальновидности». В модели ДАС3 в силу $x_0 = T – 1$ центру также известны функции полезности во всех будущих периодах, но так как он взял в первом периоде на себя обязательства относительно нескольких следующих периодов, и обязан их выполнять, то за последние $T – 1$ периодов сумма его полезности будет меньше, чем если бы он мог поменять свое решение, то есть действовал бы в рамках модели ДАС2. Учитывая, что в первый момент времени центр в обеих моделях получает одинаковую полезность, получаем, что ДАС2 не менее эффективна, чем ДАС3. •

Теорема 8. В трехпериодной модели режим управления ДАС2 всегда более эффективен, чем ДАС3.

Действительно, в трехпериодной модели ($T = 3$), условие, накладываемое на дальновидность для моделей ДАС2 и ДАС3 ($1 < x_0 < T$), оставляет единственно возможное значение дальновидности $x_0 = 2$. Следовательно, для трехпериодной модели имеем $x_0 = T – 1$, что попадает под условия леммы 2.

Отметим, что результат теоремы 8 не противоречит примеру 7, иллюстрирующему возможность превосходства ДАС3 над ДАС2, так как в этом примере $T = 4$.

5. ЭФФЕКТЫ НАКОПЛЕНИЯ В ДИНАМИЧЕСКИХ АКТИВНЫХ СИСТЕМАХ

Важным случаем общей задачи управления ДАС является модель, в которой текущий доход центра или затраты активного элемента зависят не от каких-то конкретных действий в прошлом, а от суммы всех действий за предыдущие периоды. При этом можно говорить об *эффекте накопления*, который проявляется в том, что на настоящее оказывает влияние сумма предыдущих действий.

Итак, рассмотрим частный случай общей задачи управления ДАС, когда функция дохода центра зависит от действия АЭ в текущем периоде и суммы действий за предыдущие периоды: $H^t(y^{1,t}) = y^t g(\sum_{t=1}^{t-1} y^t)$, а функция затрат активного элемента зависит только от действия в текущем периоде: $c^t = c(y^t)$. Пусть центр обладает не зависящими от времени дальновидностью x_0 и горизонтом обязательств L_0. Предполагается также, что функции $g(\times)$ и $c(\times)$ являются непрерывными и дифференцируемыми, множества допустимых действий активного элемента A^t Í $\Re^1_+$ являются отрезками, содержащими ноль. Задача, как и в разделе 3.5, заключается в сравнении эффективности различных режимов управления и различной дальновидности.

В момент времени t центр находит оптимальный план x^t (если рассматривается ДАСЗ, то $x^t, \mathbf{K}, x^{t+L_0}$) при уже известных для него оптимальных планах на предыдущие периоды, решая задачу максимизации:

(1) $$\max_{y^t \in A^t, \mathbf{K}, y^{t+x_0-1} \in A^{t+x_0-1}} \sum_{t=t}^{t+x_0-1} (y^t g(x_\Sigma^t + y^t + \mathbf{L} + y^{t-1}) - c(y^t)),$$

где $x_\Sigma^t = \sum_{t=1}^{t-1} x^t$

Сначала найдем решение этой задачи во внутренней точке области допустимости $A^{t,t+x_0-1} = A^t \times \mathbf{L} \times A^{t+x_0-1}$, а потом будем исследовать когда решения лежат на границе, и что это за решения.

Для отыскания внутреннего решения продифференцируем выражения, стоящие под знаком максимума в (1), по переменным $y^t, \mathbf{K}, y^{t+x_0-1}$, и приравняем первые производные к нулю. В итоге получим систему из x_0 уравнений с x_0 неизвестными:

$$
(2)\quad \begin{cases}
g(x_\Sigma^t) - c'(y^t) + y^{t+1} g'(x_\Sigma^t + y^t) + \mathbf{L} + \\
+ y^{t+x_0-1} g'(x_\Sigma^t + y^t + \mathbf{L} + y^{t+x_0-2}) = 0 \\
g(x_\Sigma^t + y^t) - c'(y^{t+1}) + y^{t+2} g'(x_\Sigma^t + y^t + y^{t+1}) + \\
+ \mathbf{L} + y^{t+x_0-1} g'(x_\Sigma^t + y^t + \mathbf{L} + y^{t+x_0-2}) = 0 \\
\mathrm{M} \\
g(x_\Sigma^t + y^t + \mathbf{L} + y^{t+x_0-2}) - c'(y^{t+x_0-1}) = 0
\end{cases}
$$

Вычитая попарно последующие уравнения из предыдущих, можно упростить систему (2) к виду:

$$
(3)\quad \begin{cases}
g(x_\Sigma^t + \mathbf{L} + y^{t+t-1}) - c'(y^{t+t}) + y^{t+t+1} g'(x_\Sigma^t + \mathbf{L} + y^{t+t}) = \\
= g(x_\Sigma^t + \mathbf{L} + y^{t+t}) - c'(y^{t+t+1}) \\
g(x_\Sigma^t + \mathbf{L} + y^{t+x_0-2}) - c'(y^{t+x_0-1}) = 0
\end{cases}
$$

Рассмотрим случай с линейной функцией $g(x) = a + bx$, подставляя которую в первое выражение из (3), получаем $by^{t+t+1} + c'(y^{t+t+1}) = by^{t+t} + c'(y^{t+t})$. Таким образом, независимо от вида функции затрат АЭ $c(x)$, при линейной функции $g(x)$ оптимальной является точка, в которой $y^t = y^{t+1} = \mathbf{L} = y^{t+x_0-1}$. Из второго выражения (3) находим что

(4) $a + bx_\Sigma^t + (x_0 - 1)by^t - c'(y^t) = 0$.

Зная функцию затрат АЭ и подставив ее в выражение (4), можно найти $y^t = f(x_\Sigma^t, x_0)$.

Чтобы получить аналитическое решение уравнения (4), ограничимся рассмотрением конкретного вида функции затрат

$c(x) = \frac{x^2}{2}$ и случаем неотрицательных действий. Тогда решение задачи (1) для периода t будет даваться выражением:

(5) $y^{t*} = y^{t+1*} = \mathbf{L} = y^{t+x_0-1*} = \frac{a + bx_{\Sigma}^t}{1-(x_0-1)b}$.

Эффективность управления K определяется следующим выражением:

$$K = \sum_{t=1}^{T} \Phi^t(x^{1,t}) = \sum_{t=1}^{T}(x^t(a + b\sum_{t=1}^{t-1} x^t) - \frac{(x^t)^2}{2}) =$$

(6) $$= a\sum_{1}^{T} x^t + b \sum_{i,j=1\mathbf{K}T, i\neq j} x^i x^j - \tfrac{1}{2}\sum_{1}^{T}(x^t)^2 =$$

$$= \sum_{1}^{T} x^t(a + \frac{b}{2}\sum_{1}^{T} x^t) - \frac{(b+1)}{2}\sum_{1}^{T}(x^t)^2$$

Остается важный вопрос о том является ли найденная точка экстремума искомой точкой максимума. В общем случае ответить на этот вопрос непросто, но в нашем конкретном случае с линейной функцией $g(x)$ и квадратичными затратами $c(x)$, матрица вторых производных максимизируемой функции (1) имеет размерность $x_0 \times x_0$, имеет на диагонали -1, а во всех остальных клетках b. При $a > 0$, $\frac{1}{x_0 - 1} > b > -1$, матрица является отрицательно определенной, а значит найденная точка экстремума является точкой максимума.

Найдем оптимальные планы и вычислим эффективность для моделей ДАС1 – ДАС4.

<u>**ДАС1.**</u> В модели ДАС1 $x_0 = L_0 = 1$, план $x^t = y^{t*}$. Таким образом, из (6) имеем, что оптимальные внутренние планы будут:

(7) $x^t = a(1+b)^{t-1}$, $t = \overline{1,T}$.

Для того чтобы найти эффективность K_1 ДАС1, подставим планы (7) в формулу (6):

(8) $K_1 = \frac{a^2}{2}\frac{(1+b)^{2T} - 1}{b(b+2)}$

Рассмотрим вопрос от том, когда решение будет достигаться внутри области $A^1 \times \mathbf{L} \times A^T = \Re_+^T$, то есть при каких значениях параметров a и b (7) будет действительно решением (1), а когда решение будет достигаться на границе и каким будет решение в этом случае.

При $a < 0$ для любого b решение достигается на границе и равно $x^t = 0$, $t = \overline{1,T}$, эффективность K_1 в этом случае равна нулю.

При $a > 0,\ b \le -1$ решение достигается на границе области $\Re_+^T$, и оптимальными планами будут $x^i = a,\ x^1, \mathbf{K}, x^{i-1}, x^{i+1}, \mathbf{K}, x^T = 0$. Подставляя эти планы в выражение для эффективности (6), получаем, что эффективность в этом случае равна $K_1 = \dfrac{a^2}{2}$.

При $a > 0,\ b > -1$ решение достигается внутри области $\Re^{T+}$ и, следовательно, оптимальные планы выражаются формулой (7), а эффективность – формулой (8).

<u>ДАС2.</u> В модели ДАС2 $1 < x_0 < T,\ L_0 = 1$, план $x^t = y^{t*}$. Таким образом, учитывая, что как только центр в момент времени $t = T - x_0 + 1$ начинает «видеть период *T*», для него согласно принципу оптимальности Беллмана оптимальными планами в последующие периоды будут являться решения задачи (1) для периода $t = T - x_0 + 1$, из (5) имеем, что оптимальные внутренние планы определяться следующей формулой:

$$(9)\quad \begin{cases} x^1 = \dfrac{a}{1-(x_0-1)b} \\ x^t = \dfrac{a}{1-(x_0-1)b}\left(1 + \dfrac{b}{1-(x_0-1)b}\right)^{t-1}, \text{ где } t = 2, \mathbf{K}, T - x_0\,. \\ x^{T-x_0+1} = \mathbf{L} = x^T = \dfrac{a}{1-(x_0-1)b}\left(1 + \dfrac{b}{1-(x_0-1)b}\right)^{T-x_0} \end{cases}$$

Подставляя планы из (9) в выражение (6), находим эффективность:

$$(10)\ K_2 = \frac{a^2\left[(1+b)\left(1+\frac{b}{1-(x_0-1)b}\right)^{2(T-x_0)+1} + 2(x_0-1)b - 1\right]}{2b(2(1-(x_0-1)b)+b)}.$$

Также как и в случае с ДАС1, определим при каких значениях параметров a и b "внутренние" планы (9) будут действительно решениями поставленной задачи.

При $a < 0$ для любого b решение достигается на границе и равно $x^t = 0,\ t = \overline{1,T}$, а эффективность K_2 в этом случае равна нулю.

При $a > 0,\ b \le -1$ решение достигается на границе области $\Re_+^T$ и оптимальными планами будут $x^i = a,\ x^1, \mathbf{K}, x^{i-1}, x^{i+1}, \mathbf{K}, x^T = 0$. Подставляя эти оптимальные планы в выражение для эффективности (6) получаем, что эффективность в этом случае равна $K_2 = \frac{a^2}{2}$.

При $a > 0,\ -1 < b < 1/(x_0 - 1)$ решение достигается внутри области $\Re^{T+}$, и, следовательно, оптимальные планы выражаются формулой (9), а эффективность – формулой (10).

При $a > 0,\ b \ge 1/(x_0 - 1)$ оптимальными будут сколь угодно большие планы (бесконечные) во всех периодах, так как, несмотря на квадратичные затраты, при данных значениях параметров a и b затраты окупаются уже во втором периоде. Эффективность в этом случае бесконечно большая. Конечно, такой результат обусловлен отсутствием ограничений, которые обычно присутствуют в реальной жизни. Например, обычно экономические агенты не готовы терпеть в первый период огромные убытки, даже если знают, что уже во втором периоде они будут компенсированы. Но более серьезное ограничение состоит в том, что множество возможных действий обычно ограничено сверху.

ДАС3. В модели ДАС3 $1 < x_0 < T,\ 1 \le m_0 \le L_0 \le x_0$, план на период обязательств определяется из (5) $x^{L_0+km_0-1}, \mathbf{K}, x^{L_0+(k+1)m_0} = y^{L_0+km_0-1*}$. Учитывая, во-первых, то, что решая задачу (1) для периода t, в котором принимается решение, мы находим план не только на этот период, но и на период обязательств, и, во вторых, принимая во внимание лемму 1, получаем оптимальные (внутренние) планы для ДАС3:

$$(11)\quad \begin{cases} x^1 = \mathbf{K} = x^{L_0} = x^0 \\ x^{L_0+1} = \mathbf{L} = x^{L_0+m_0} = x^0 b \\ \vdots \\ x^{L_0+(n-1)m_0+1} = \mathbf{L} = x^T = x^0 b^n \end{cases},$$

где

$$x^0 = \frac{a}{1-(x_0-1)b},\quad b = \left(\frac{1-b(x_0-1-L_0)}{1-(x_0-1-L_0+m_0)b}\right),$$

$$n = \text{целая часть}\left[\frac{T-x_0}{m_0}\right].$$

Подставляя планы из (11) в выражение (6), находим выражение для эффективности K_3 ДАС3:

$$(12)\quad \begin{aligned} K_3 = a\,(L_0x^0 + m_0bx^0\frac{b^n-1}{b-1} + (T-L_0-m_0n)x^0b^n) + \\ + \frac{b}{2}(L_0x^0 + m_0bx^0\frac{b^n-1}{b-1} + (T-L_0-m_0n)x^0b^n)^2 - \\ - \frac{b+1}{2}(L_0(x^0)^2 + m_0b^2(x^0)^2\frac{b^{2n}-1}{b^2-1} \\ + (T-L_0-m_0n)(x^0)^2b^{2n}) \end{aligned}$$

Анализ значений параметров a и b, при которых планы, выражаемые (11), будут действительно решениями нашей задачи, аналогичен анализу, проведенному для ДАС2 и дает такой же результат, поэтому не будем его повторять.

ДАС4. В модели ДАС4 $x_0 = T_0$, планы определяются как $x^t = y^{1*}$, то есть они одинаковы для всех периодов, и выражаются формулой:

$$(13)\quad x^t = \frac{a}{1-(T-1)b},\ t = \overline{1,T}.$$

Соответственно, эффективность модели ДАС4 будет:

(14) $$K_4 = \frac{a^2 T}{2(1-(T-1)b)}.$$

Мы не будем здесь приводить анализ допустимости решения (13), так как он полностью аналогичен анализу для ДАС2, и результаты получаются такими же, но с $x_0 = T$.

Отметим, что модели ДАС1, ДАС2, ДАС4 являются частными случаями модели ДАС3. Действительно, при $x_0 = L_0 = m_0 = 1$ ДАС3 переходит в ДАС1, при $1 < x_0 < T, L_0 = m_0 = 1$ ДАС3 переходит в ДАС2, при $x_0 = T$ ДАС3 переходит в ДАС4. Таким образом, подставляя в формулы (11) и (12) планов и эффективности для ДАС3 соответствующие значения дальновидности, горизонта обязательств и частоты принятия решения, получим формулы (7) и (8) планов и эффективности для ДАС1, формулы (9) и (10) планов и эффективности для ДАС2, формулы (13) и (14) планов и эффективности для ДАС4.

Таким образом, для того, чтобы понять какая модель лучше в смысле эффективности, достаточно исследовать поведение K_3 (далее будем пользоваться обозначением $K_3 = K(a, b, x_0, L_0, m_0, T)$) в зависимости от изменения параметров дальновидности x_0, горизонта обязательств L_0, и частоты принятия решения m_0.

На рисунке 7 изображены графики зависимости эффективностей от b для всех четырех моделей при $a > 0$. Видно, что есть две особые точки на графике, где все четыре кривые пересекаются – при $b = 0$ и $b = 1$. При этом в нуле для эффективностей K_1, K_2, K_3 происходит излом.

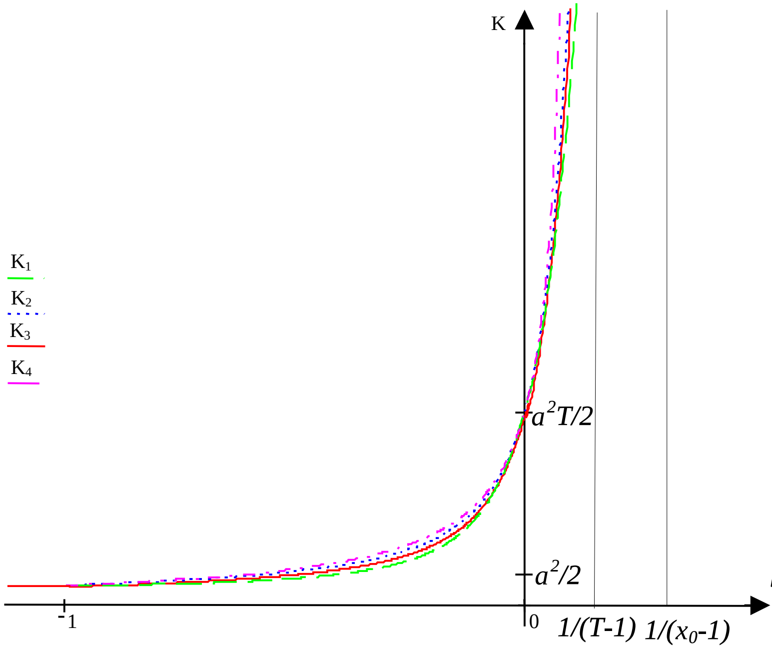

Рис. 7. Сравнение эффективности моделей ДАС1, ДАС2, ДАС3, ДАС4.

Для рассматриваемой линейной модели выполняется соотношение $K_1 \le K_3 \le K_2 \le K_4$, то есть самой эффективной является модель ДАС4 (что и должно было получиться, как было показано в Теореме 5), далее в порядке уменьшения эффективности идет модель ДАС2, потом ДАС3, и наконец наименее эффективной является модель ДАС1.

Также верно утверждение, что чем больше дальновидность, тем выше эффективность. Это продемонстрировано на рисунке 8, где для модели ДАС2 изображена зависимость эффективности K_2 от величины дальновидности x_0.

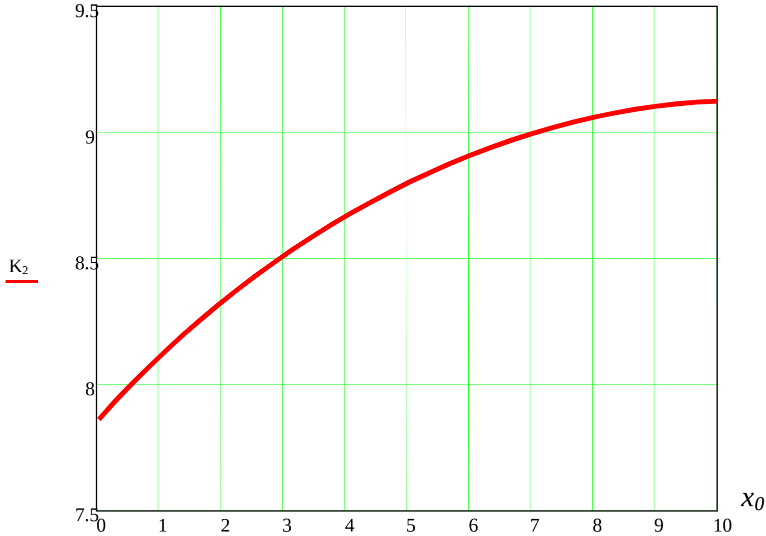

Рис. 8. Зависимость эффективности модели ДАС2 от дальновидности центра x_0 при $b = 0.05$, $T = 10$

Таким образом, можно сделать вывод о том, что в рассматриваемой модели чем больше центр информирован о будущем в каждый момент времени (если рассматривать модель ДАС3 как модель ДАС2 с переменной дальновидностью – см. лемму 1), тем выше эффективность.

Как показано на рисунке 9, при $b < 0$ оптимальные планы убывают со временем. Содержательно это означает то, что, раз сумма действий за предыдущие периоды негативно влияет на доход в этом периоде, то со временем план следует понижать, чтобы сдерживать негативное влияние на будущее. Такая ситуация может возникнуть, например, в модели загрязнения окружающей среды.

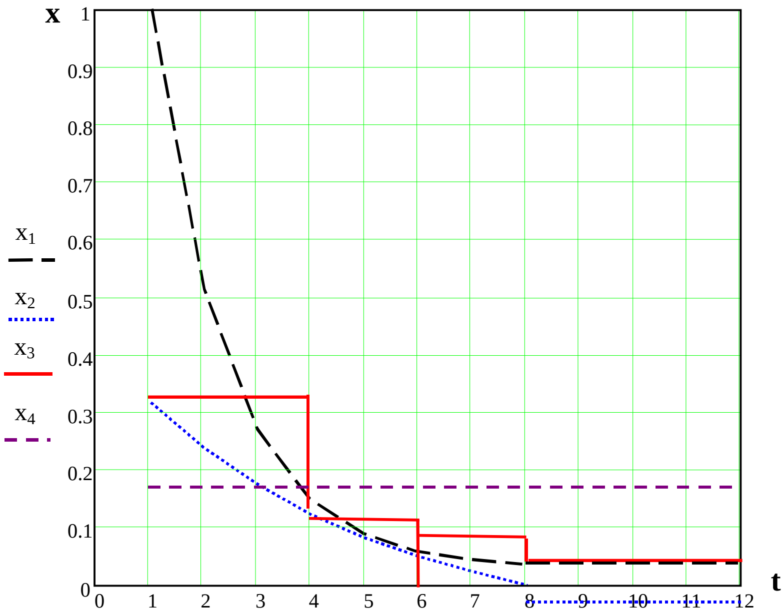

Рис. 9. Оптимальные планы для моделей ДАС1 – x_1, ДАС 2 – x_2 (x_0 = 5), ДАС3 – x_3 (x_0 = 5, L_0 = 4, m_0 = 2) и ДАС4 – x_4 в случае b < 0, T = 12

Рассмотрим несколько иллюстративных примеров.

Пример 8 (модель загрязнения окружающей среды). Рассмотрим город, построенный около крупного предприятия химической промышленности. В процессе функционирования предприятие выбрасывает вредные вещества в атмосферу, тем самым загрязняя окружающую среду. Пусть количество загрязнений линейно зависит от объема выпускаемой продукции, а степень загрязнения зависит от суммы всех выбросов, начиная с момента начала функционирования до текущего момента. Это неявно предполагает, что со временем негативный эффект от выбросов сохраняется довольно долго, не диссипируя во времени. Благополучие города зависит не только от объема выпускаемой предприятием продукции, но и от самочувствия людей, живущих в городе, а значит – от состояния окружающей среды.

Рассмотрим, как эта ситуация может быть отражена в вышеизложенной модели с накоплением. Центром является город, агентом (активным элементом) является предприятие. Производя действие – производство некоторого количества продукции в год, предприятие этим самым оказывает определенное негативное воздействие на окружающую среду, накопление которого скажется в том числе в будущих периодах. Этой модели соответствует значение $b < 0$. Если администрация города не знает к каким последствиям в будущем могут привести действия в настоящем, то есть, если центр недальновиден и соответственно действует в рамках модели ДАС1, то в первый период администрация утверждает большой план для производства. Уже во втором периоде последствия от этого действия начинают сказываться, что выражается в достаточно сильном ухудшении состояния окружающей среды. Это приводит к резкому уменьшению оптимального плана на следующий период (см. рисунок 9).

Если администрация города более информирована о вредном влиянии производства на окружающую среду, т.е. реализуется модель ДАС2 или ДАС3, то уже в первый момент времени назначается сравнительно небольшой план. Поэтому оптимальные планы для этих моделей не так резко уменьшаются в начальных периодах, как это происходит в случае недальновидного центра. Можно сказать что к реальной ситуации наиболее приближены модели ДАС2 и ДАС3, так как обычно о вредном воздействии на атмосферу руководство города и завода знает и учитывает этот фактор, вопрос в том насколько далеко вперед (в будущее) центр «заглядывает» при принятии текущих решений.

В случае полностью дальновидного центра (реализуется модель ДАС4), оптимальными являются низкие, но одинаковые планы на весь период функционирования предприятия (в качестве которого может быть выбрано характерное время диссипации загрязнений). В рассматриваемом примере такая ситуация может возникнуть только в случае, если точно известен срок жизни предприятия (или администрации) и, более того, центр обладает полной информацией о функциях затрат и дохода на весь этот период. Такие жесткие требования модели ДАС4 говорят о том, что она будет вряд ли может быть реализована на практике.

Рассмотрим другой случай, когда $b > 0$. При $b \geq 1/(x_0 - 1)$ оптимальными являются бесконечно большие планы, которые дают бесконечно большую эффективность. Для случая $\frac{1}{x_0 - 1} > b > 0$ оптимальные планы представлены на рисунке 10.

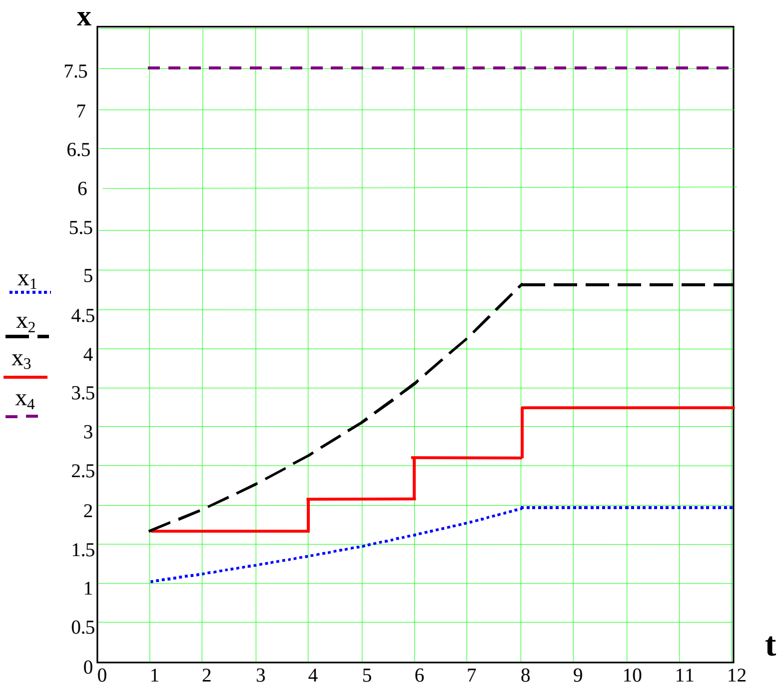

Рис. 10. Оптимальные планы для моделей ДАС1 – x_1, ДАС 2 – x_2 (x_0 = 5), ДАС3 – x_3 (x_0 = 5, L_0 = 4, m_0 = 2) и ДАС4 – x_4 в случае $1/(x_0 - 1) > b > 0$, $T = 12$

В этом случае оптимальные планы возрастают во времени. При этом план для модели с полной дальновидностью (то есть для ДАС4) изначально самый большой. •

Случай $b > 0$ может хорошо подходить для описания модели фирмы, выходящей на новый неосвоенный еще ни кем рынок сбыта.

Пример 9 (модель выхода фирмы на новый рынок). Пусть некоторая компания продвигает новый продукт на рынок, и этот продукт является уникальным, то есть у компании нет конкурентов. Проблема заключается в неосведомленности потенциальных покупателей о данном продукте.

Предположим, что спрос определяется осведомленностью покупателей о продвигаемом продукте. Чем больше продано продукта за предыдущие периоды, тем больше о нем осведомленность потребителя, и значит тем больше будет спрос на него в текущем периоде. Такое поведение укладывается в рассматриваемую модель ДАС с $b > 0$, где центром является руководство предприятия, которое назначает план, ориентируясь на потенциальный спрос.

Если руководство недальновидно, то есть не знает, что количество продукта, которое предприятие произведет сегодня и которое будет продано, положительно скажется на спросе в следующем периоде, то оно назначает небольшой план относительно плана, который был бы назначен, если руководство было бы дальновидно (см. рисунок 10). Со временем оптимальный план не уменьшается для всех четырех моделей (в случае ДАС4 он остается постоянным) так как спрос на продукцию растет и, чтобы удовлетворить этот спрос, и, соответственно, получить больший доход, надо увеличивать план.

Основной полученный выше результат о соотношении эффективностей различных моделей ДАС: $K_1 \leq K_3 \leq K_2 \leq K_4$, можно проинтерпретировать следующим образом. Чем больше компания знает о поведении потребителей в будущем (то есть обладает большей дальновидностью), тем она ведет себя более эффективным образом.

Обсудим границы применимости рассматриваемой модели в данном примере. Основное ограничение состоит в том, что функцию $g(\sum_{t=1}^{t-1} y^t)$, отражающую эффект накопления, на практике можно полагать линейной только при небольших значениях аргумента, потому что рынок сбыта не является бесконечным и как только продукт заполнит в этом рынке свою нишу, описываемый эффект перестанет действовать. Таким образом, можно говорить что предложенная модель хорошо описывает процесс выхода нового продукта на рынок в начальной стадии освоения этого рынка при условии уникальности и востребованности продукта. •

В заключение настоящего раздела отметим, что в рассматриваемой задаче считалось, что коэффициент дисконтирования равен единице, то есть будущий доход для центра для него также важен как и доход в настоящем. На практике коэффициент дисконтирования обычно меньше единицы. В этом случае граница для b (см. рисунок 7), начиная с которой оптимальными становятся бесконечно большие действия, сместится вправо.

ЗАКЛЮЧЕНИЕ

Таким образом, в настоящей работе:

- проведен обзор основных результатов теории активных систем, теории иерархических игр и теории контрактов по управлению динамическими активными системами (см. Приложение);
- дана общая постановка и введена система классификаций задач управления ДАС (раздел 1), выделены четыре базовых модели ДАС (раздел 2);
- решена задача стимулирования в многоэлементной детерминированной ДАС (теоремы 1 и 3), охарактеризовано множество планов, согласованных в ДАС (теорема 2);
- классифицированы распределения дальновидностей и горизонты принятия решений участниками ДАС, выявлен и исследован эффект обмена ролями, заключающийся в опережающем принятии решений управляемым субъектом (теорема 4);
- исследовано влияние режимов управления на эффективность управления базовыми ДАС (теоремы 5-6);
- решены задачи управления и получены оценки сравнительной эффективности различных режимов управления двух и трехпериодными ДАС (теоремы 7-8);
- изучены «эффекты накопления» в ДАС (раздел 5), что позволило описать и исследовать ряд прикладных моделей.

<u>Приложение</u>

ОБЗОР ОСНОВНЫХ РЕЗУЛЬТАТОВ ТЕОРИИ АКТИВНЫХ СИСТЕМ, ТЕОРИИ ИЕРАРХИЧЕСКИХ ИГР И ТЕОРИИ КОНТРАКТОВ ПО УПРАВЛЕНИЮ ДИНАМИЧЕСКИМИ АКТИВНЫМИ СИСТЕМАМИ

Повторяющиеся игры

Рассмотрим игру n лиц[1], стратегией каждого из которых является выбор y_i Î A_i, i Î I. Если обозначить функцию выигрыша i-го игрока $f_i(y)$, $y = (y_1, y_2, \dots, y_n)$ Î $A' = \prod_{i \in I} A_i$, то *однопериодной игрой* G называется кортеж $G = (A_1, A_2, \dots, A_n, f_1, f_2, \dots, f_n)$. *Динамической игрой* $G(T)$ (повторяющейся игрой, суперигрой и т.д.) называется игра G, повторенная T раз. Выигрыш j_i i-го игрока в суперигре есть среднее значение его выигрышей по всем периодам, то есть[2]

---

[1] *Так как ниже приводится обзор основных результатов различных научных школ и направлений по исследованию теоретико-игровых моделей управления динамическими активными системами (теория активных систем, теория игр, теория контрактов и т.д.), а в различных классах моделей для обозначения одних и тех же субъектов используются различные термины, то при описании результатов мы будем использовать терминологию, принятую в соответствующей научной школе. Во избежании путаницы следует отметить, что «равноправные» субъекты в теории игр обозначаются терминами «игрок» («второй игрок» или «производитель» в теории иерархических игр), в теории активных систем – «активный элемент» или просто «элемент», а в теории контрактов – «агент» (agent). Если рассматривается иерархическая система, то игроки, обладающие правом первого хода в рамках каждого периода функционирования, обозначаются: «центр» – в теории активных систем, «центр» или «первый игрок» – в теории иерархических игр, «начальник» (principal) – в теории контрактов. В ходе дальнейшего изложения в целях общности термины, обозначающие одно и то же понятие, будут использоваться как синонимы.*

[2] *В выражении (1) предполагается, что игроки одинаково учитывают полезности, получаемые в различных периодах. Учет будущего производится введением дисконтирующих множителей (см. ниже).*

(1) $j_i(y^{1,T}) = \frac{1}{T} \sum_{t=1}^{T} f_i(y^t)$,

где $y^t = (y_1^t, y_2^t, ..., y_n^t)$ $\hat{I}$ A' – вектор стратегий игроков в момент времени t, $y^{1,T} = (y^1, y^2, ..., y^T)$ – вектор стратегий игроков за периоды с первого по период T. «Стратегией»[1] i-го игрока в игре $G(T)$ в момент времени t является отображение $y_i^t : (A')^{t-1}$ ® A_i *истории игры* $y^{1,t-1}$, сложившейся к моменту $(t - 1)$, во множество его допустимых стратегий. Следовательно, «стратегия» i-го игрока – вектор $Y_i = (y_i^1, y_i^2, ..., y_i^T)$. Набор $Y = (Y_1, Y_2, ..., Y_n)$ индуктивно определяет в суперигре *путь* $(y^1(Y), y^2(Y), ..., y^T(Y))$, где $y^1(Y) = Y_1$, $y^t(Y) = Y^t(y^{1,t-1}(Y))$, $t > 1$.

Равновесие Нэша Y^* определяется следующим образом:

(2) $\frac{1}{T} \sum_{t=1}^{T} f_i(y^t(y^*))$ ³ $\frac{1}{T} \sum_{t=1}^{T} f_i(y^t(y^*_{-i}, y_i))$ " Y_i, i $\hat{I}$ I.

Обозначим: $N(T)$ – множество равновесных по Нэшу путей в суперигре (предположим, что $N(1)$ не пусто); $j_i^{\min}$ – гарантированный (максиминный) выигрыш i-го игрока (стратегия, обеспечивающая j_i ³ j_i^{min}, называется *индивидуально рациональной* (IR)), $\Im$ – выпуклую оболочку множества возможных значений целевых функций игроков; $\Im^*$ – подмножество множества $\Im$, состоящее из выигрышей игроков, доминирующих максиминные; $G(T-t)$ – подыгру игры $G(T)$, соответствующую последним $T - t$ периодам, где $t < T$; $Y_i(y^{1,t})$ – стратегию i-го игрока в игре $G(T - t)$ такую, что $y_i^1(y^{1,t}) = y_i^{t+1}(y^{1,t})$ и " $t < T - t$, " $a^{1,t} = (a^1, a^2, ..., a^t)$ $y_i^{t+1}(a^{1,t}) = y_i^{t+t+1}(y^{1,t}, a^{1,t})$; $Y(y^{1,t}) = (Y_1(y^{1,t}), Y_2(y^{1,t}), ..., Y_n(y^{1,t}))$.

---

[1] *Употребление кавычек обусловлено тем, что термин «стратегия» в теории игр используется в двух смыслах – как результат выбора игрока (в рассматриваемой модели – элемент множества* A_i*) и как используемое им правило принятия решений (в рассматриваемой модели – отображение имеющейся информации во множество* A_i*).*

Стратегия Y называется *согласованным с подыграми равновесием* (Subgame Perfect Equilibrium – SPE[1]) суперигры $G(T)$, если Y – равновесие Нэша в суперигре и $"\ t < T,\ "\ y^{1,t}\ \ Y(y^{1,t})$ – равновесие Нэша в подыгре $G(T\text{-}t)$. Содержательно SPE является усилением концепции равновесия Нэша для случая повторяющихся игр – требуется, чтобы для всех подыгр, заканчивающихся в момент времени T, стратегия была равновесной по Нэшу для любой истории игры, предшествующей рассматриваемой подыгре (идеология близка к принципу оптимальности Беллмана в динамическом программировании [9, 10]). SPE, в частности, обладает следующим свойством: путь, образованный последовательностью равновесных путей, является равновесным в игре, образованной последовательностью соответствующих игр.

Основная идея повторяющихся игр заключается в том, что при многократном повторении однопериодной игры удается добиться того, что выбор игроками индивидуально рациональных стратегий приводит к реализации рационального для всего коллектива исхода. В однопериодной игре это не всегда так: в общем случае, если используется некооперативная концепция равновесия (равновесие Нэша), то в однопериодной игре точка Нэша может оказаться неэффективной (по Парето) с точки зрения всех игроков. В то же время, может существовать оптимальный по Парето набор стратегий, который не является равновесным по Нэшу. Классическим примером является игра двух лиц "дилемма заключенного" (см., например, [66, 82, 107, 128]).

Многократное повторение рассматриваемой игры в некоторых случаях позволяет "оставить" игроков в Парето-оптимальной точке. Интуитивно понятно, что для этого нужно придумать механизм, который предотвращал бы отклонения, то есть наказывал бы отклонившегося игрока, причем наказывал настолько сильно, чтобы отклонение становилось невыгодным. Этой цели служит вводимая ниже стратегия наказания[2].

---

[1] *Иногда SPE переводится как «абсолютное равновесие Нэша»* [66]*, или «совершенное равновесие Нэша».*

[2] *В иерархических системах «наказание» может осуществляться центром, что иногда позволяет добиться эффективного равновесия в статике* [53].

Обозначим через $P(T)$ множество всех SPE в игре $G(T)$, обладающее следующими свойствами [87, 88, 101, 115]: это множество компактно; если некоторый путь принадлежит $P(T)$, то любой подпуть, получаемый из исходного отбрасыванием, начиная с первого момента времени, любого (меньшего T) числа стратегий, также принадлежит $P(T)$. Определим оптимальную k-периодную *стратегию наказания* i-го игрока:

(3) $$W_i^{1,k} = \min \{ \sum_{t=1}^{k} f_i(y^t) \mid y^{1,k} \hat{I} \; P(k)\}.$$

Для того чтобы $y^{1,T}$ $\hat{I}$ $P(T)$ необходимо и достаточно, чтобы $"\ i\ \hat{I}\ I,\ "\ t < T$

(4) $$W_i^{1,T-t} \leq \sum_{j=t+1}^{T} f_i(y^j),$$

то есть наказание должно быть достаточно сильным – полезность при наказании в течение всех оставшихся периодов не должна превышать то, что игрок мог бы получить не будучи наказанным [91, 92, 103, 105].

Содержательно, качественное отличие повторяющихся (многопериодных) игр от "обычных" (статических, однопериодных) заключается в том, что наличие нескольких периодов повышает ответственность игроков за свои действия – если кто-то повел себя не так как следовало, то в следующих периодах он может быть наказан остальными игроками за это отклонение. Для того, чтобы предотвращать отклонения, наказание должно быть достаточно сильным (см. (4)) и компенсировать возможный выигрыш игрока, который тот получает отклоняясь. Переключение с "нормального" режима на наказание (и быть может возвращение к исходному режиму через несколько периодов) получило название *триггерной стратегии*. Некоторые примеры того, как строить триггерные стратегии и того, как определить наилучший момент переключения (ведь не всегда можно достоверно установить факт отклонения, особенно в условиях неполной информированности), приведены в [108, 122, 131, 134, 145].

Существенной в повторяющихся играх оказывается информированность игроков. Если все игроки наблюдают все стратегии, выбранные партнерами в прошлом, то будем говорить, что имеет место *полная информированность* (perfect monitoring [133]). Если же стратегии, выбираемые в прошлом, ненаблюдаемы, а есть другая информация, например, если наблюдаемы полезности игроков[1], то имеет место *неполная информированность* (imperfect monitoring).

Основным результатом (группой результатов), полученным при исследовании повторяющихся игр является так называемая "народная теорема" (Folk Theorem (FTh)) [83, 110, 114, 139, 147 и др.]. Приведем серию теорем типа FTh [104]:

FTh1: Если игроки слабо дисконтируют будущее (коэффициенты дисконтирования близки к единице), то для любого вектора выигрышей j^* Î $\Im^*$ существует равновесие Нэша в бесконечной суперигре, в котором игроки получают выигрыши, в точности равные j^*.

Интуитивное обоснование этого результата таково. Пусть в многопериодной игре игроки выбирают стратегии y_i^* Î A_i, i Î I, обеспечивающую выигрыши j^*, до тех пор пока игрок с некоторым номером i не отклонится от соответствующей своей стратегии. В случае его отклонения в периоде k все игроки переключаются на $W_i^{k,\infty}$. Понятно, что в бесконечной игре при достаточно слабом дисконтировании моментальный выигрыш от отклонения компенсируется "вечным" наказанием.

FTh2: " j^* Î $Á^*$ в бесконечно повторяющейся игре без дисконтирования существует SPE, в котором ожидаемый выигрыш i-го игрока равен j_i^*, i Î I.

FTh3: Если некоторый вектор выплат j^* Î $Á^*$ Парето-доминирует равновесные по Нэшу выплаты в однопериодной игре, то при слабом дисконтировании в бесконечной суперигре существует SPE, в котором средний выигрыш равен j^*.

Для простоты далее будем считать, что все игроки одинаково учитывают будущее (имеют одинаковый дисконтирующий множитель) d.

---

[1] *До сих пор мы считали, что при принятии решений о выборе стратегии в каждом периоде каждый игрок одинаково учитывает будущие периоды* (*см.* (1)). *Однако, зачастую, будущие периоды учитываются с разными весами – дисконтирующими множителями.*

FTh4: Пусть $j(d)$ Í $Á^*$ множество средних выигрышей игроков в SPE бесконечно повторяемой игры, в которой игроки имеют дисконтирующий множитель d. Тогда " $d < 1$ соответствие $j(\cdot)$ полунепрерывно сверху (требование полунепрерывности нарушается при $d = 1$ (см. [104])).

В случае дисконтирования будущего справедлива

FTh5: Если $n = 2$, то " (j_1, j_2) Î $Á^*$ \$ d_0 Î $(0; 1)$: " d Î $(d_0; 1)$ существует SPE суперигры, в котором игроки получают средние выигрыши j_1 и j_2, если их дисконтирующие множители равны d.

Теорема FTh5 может быть обобщена на случай произвольного конечного числа игроков (достаточно потребовать непустоты внутренности множества $\Im^*$) [104].

На силу наказания (в сравнении выигрыша от одномоментного отклонения и дисконтированного проигрыша от наказания) существенно влияет величина дисконтирующего множителя, конечность [91] (а иногда и величина) или бесконечность T [34], а также информированность игроков. При полной информированности в суперигре может существовать равновесие Нэша, доминирующее по Парето равновесие Нэша однопериодной игры. Если игроки не дисконтируют будущие полезности, то множества равновесных векторов выплат в однопериодной и многопериодной игре совпадают. Если игроки дисконтируют будущие полезности, то все равновесия суперигры, в принципе, могут быть неэффективны (по Парето), хотя, обычно, при условии, что дисконтирующие множители не очень малы, существуют равновесия суперигры, доминирующие по Парето однопериодные.

В случае двух игроков и полной информированности равновесие в суперигре обладает следующим свойством непрерывности: любой эффективный индивидуально рациональный вектор выплат однопериодной игры может быть сколь угодно точно аппроксимирован равновесным вектором выплат суперигры. В [133] приведен пример неэффективного равновесия при наличии дисконтирования будущего, в [126], напротив, показывается, что при неполной информированности в некоторых случаях FTh оказывается верна. В условиях полной информированности при условии, что игроки не дисконтируют свои полезности (берется средняя полезность), в суперигре существует эффективное равновесие. Если же игроки дисконтируют свои полезности, то равновесие в многопериодной игре будет превосходить (по Парето) равновесие однопериодной игры [132]. В случае полной информированности факт отклонения каким-либо игроком от эффективной стратегии устанавливается тривиально, так как выбор стратегий наблюдаем. В случае неполной информированности все оказывается несколько сложнее – после каждого периода каждый игрок проверяет статистическую гипотезу, что все остальные игроки выбрали эффективные стратегии. Если один из игроков отвергает эту гипотезу, то все игроки переключаются на равновесные в однопериодной игре равновесия Нэша (эта стратегия, в общем случае, неэффективна). После заданного числа шагов (фаза наказания) все игроки возвращаются к эффективным стратегиям и опять проверяют свои гипотезы. Некоторые модели учитывают репутацию игроков если в течение длительного времени они вели себя "хорошо", то для переключения на стратегию наказания при проверке статистических гипотез требуется выполнения более жестких условий [102].

Условия и стратегии суперигры, приводящие к векторам полезностей, доминирующим однопериодное равновесие Нэша и даже более того, эффективным в однопериодной игре, для случая полной информированности приводятся в [132]. Этот же результат имеет место и для неполной информированности при некоторых дополнительных условиях (теорема 7.1 в [132]). К "недостаткам" FTh следует отнести: отсутствие предсказуемости (любой индивидуально рациональный результат может быть равновесием суперигры; FTh утверждает, что в суперигре возможно кооперативное равновесие (Парето), но непонятно каковы механизмы его достижения; наличие угрозы для того игрока, который отклоняется (или собирается отклониться), может привести к тому, что он захочет пересмотреть правила игры и т.д. [85, 92, 99, 120, 129].

Повторяющиеся иерархические игры

Особо следует отметить результаты исследования повторяющихся игр в теории иерархических игр[1]. Наибольший интерес с

---

[1] *Исторически сложилось так, что исследования по многим близким направлениям теории игр* (*в том числе и повторяющимся играм*) *в СССР и за рубежом велись параллельно, но независимо. Справедливости ради, надо отметить, что относительно многих результатов типа FTh – ср., например,* [30, 32, 41] *и* [115] (*не говоря уже о иерархических многошаго-*

точки зрения настоящей работы представляют приведенные в [41] общие результаты (см. там же соответствующий обзор), характеризующие оптимальные стратегии и выигрыш центра при повторениях игр типа Γ_1 и Γ_2 [30]. В том числе рассматривались две модели.

В первой модели[1] центр (игрок, делающий ход первым) сообщает агенту (второму игроку) свои стратегии – функции $u_i(y^{1,i})$, $i = \overline{1,T}$ (при использовании таких стратегий могут быть учтены случаи произвольного запаздывания информации, получаемой первым игроком о стратегии, выбранной вторым игроком), после чего выбор второго игрока становится "одношаговым" и заключается в определении оптимального для него при заданном управлении вектора $y^{1,T}$. Как и в статическом случае [30, 32], выделяются два режима – за выбор определенных стратегий (действий) агент поощряется, за выбор остальных действий наказывается. Таким образом, оптимальной является следующая стратегия центра – использовать «поощрения» до тех пор, пока агент в первый раз не выберет несогласованное с центром действие, после чего центр до конца игры переключается на использование стратегии наказания. Этот результат охватывает результаты, полученные для статических игр, как частные случаи, и, кроме того, позволяет получить решение задачи синтеза оптимальных управлений со стороны центра в повторяющихся иерархических играх, в которых целевой функцией агента является суммарная по периодам дисконтированная полезность (при условии, что полезность в каждом периоде зависит только от стратегий, выбранных в этом периоде) [41].

Во второй модели центр сообщает агенту свои стратегии – функции $u_i(y^{1,i})$, $i = \overline{1,T}$ – последовательно, то есть только на очередной ход, когда будет выбираться y^i. При этом решение может получено применением принципа оптимальности Беллмана – считая известными $u^{1,T-1}(\cdot)$ и $y^{1,T-1}$, центр решает статическую задачу – определения оптимального управления $u^T(\cdot, u^{1,T-1}, y^{1,T-1})$ и т.д., вплоть до первого периода.

вых играх [33, 34, 40]) *приоритет принадлежит советским или российским ученым.*

[1] *Предпочтения игроков в данной модели отражены произвольными непрерывными функциями от векторов всех стратегий за все предыдущие периоды игры.*

Многошаговые иерархические игры, описывающие управление динамической системой, состояние которой в момент времени $t+1$ зависит от ее состояния в момент времени t и управлений, выбранных центром (стратегия центра – функция от состояния системы) и агентом (агент в каждый момент времени выбирает свою стратегию при известной стратегии центра), рассматривались в [41, 66]. Данная игра сводится к антагонистической игре (определение стратегий наказания) и задаче оптимального управления.

Динамические задачи теории контрактов

В настоящем разделе рассматриваются динамические задачи теории контрактов, которые, с одной стороны, используют общие результаты анализа повторяющихся игр, а с другой – достаточно близки к динамическим моделям, исследуемым в теории активных систем (ТАС) – см. ниже.

Если предположить, что результаты деятельности АЭ в различных периодах не связаны, элементы недальновидны и отсутствуют общие ограничения на целевые функции и допустимые множества различных периодов, то получится последовательность базовых моделей теории контрактов [15, 57, 58, 125, 131], каждая из которых может исследоваться независимо.

В случае наличия общих ограничений на целевые функции, допустимые множества, параметры механизма стимулирования и т.д., при несвязанных периодах функционирования, задача стимулирования в динамической системе, по аналогии с задачей стимулирования в системе со слабо связанными элементами, может быть сведена к стандартной задаче условной оптимизации [52, 56-58].

Оба описанных выше случая представляются довольно тривиальными и редко встречаются на практике. Поэтому рассмотрим двухпериодную одноэлементную динамическую задачу теории контрактов и методы ее решения, следуя введенной в [15] терминологии.

Модель теории контрактов относится к моделям систем с внешней вероятностной неопределенностью [15, 58] и качественно заключается в следующем. Агент предпринимает некоторые действия, которые совместно с реализацией внешнего неопределенного (случайного) параметра приводят к определенным результатам деятельности. Принципиальное отличие данной модели от детерминированной заключается в том, что на момент принятия решений о выбираемых стратегиях участники системы (центр и агент) не имеют информации о будущем значении состояния природы, обладая лишь информацией о параметрическом распределении вероятностей результатов деятельности при тех или иных действиях. Кроме того, считается, что действие агента ненаблюдаемо для центра, поэтому управление может основываться лишь на наблюдаемой реализации случайной величины – результата деятельности, а участники системы предполагаются рациональными в смысле стремления максимизации своих ожидаемых полезностей.

Введем некоторые обозначения: $A = \{y_1, y_2, ..., y_n\}$, $A = \{z_1, z_2, ..., z_n\}$ (множества возможных действий и результатов деятельности конечны); $y^1 \in A$ – действие АЭ в первом периоде; $y^2 \in A$ – действие АЭ во втором периоде; $z_j \in A_0$ – результат деятельности АЭ в первом периоде; $s_j \in M$ – стимулирование АЭ в первом периоде за результат z_j, $s_{jl} \in M$ – стимулирование АЭ во втором периоде за результаты: z_j и z_l в первом и втором периодах, соответственно; $s = (s_j, s_{jl})$; $F(y^1 - s_j, y^2 - s_{jl})$ – возрастающая и вогнутая по обеим переменным целевая функция центра; $f(s_j, s_{jl}, y^1, y^2)$ – возрастающая и вогнутая по s и убывающая по y целевая функция АЭ; $p_j(y^k)$ – вероятность результата z_j при действии y^k, $k = 1, 2$, $j, l = \overline{1,n}$. Итак, ожидаемые полезности центра и агента имеют, соответственно, вид:

(1) $$EF(s, y) = \sum_{j,l=1}^{n} p_j(y^1)\, p_l(y^2)\, \Phi(y^1 - s_j, y^2 - s_{jl}),$$

(2) $$Ef(s, y) = \sum_{j,l=1}^{n} p_j(y^1)\, p_l(y^2)\, f(s_j, s_{jl}, y^1, y^2)$$

где $y = (y^1, y^2)$, E – оператор математического ожидания. По аналогии с базовой однопериодной моделью [15], задача поиска двухпериодного оптимального контракта (напомним, что контрактом называется совокупность $\{s^*; y^*\}$ оптимальной системы стимулирования и реализуемого ей действия АЭ [15, 42, 56]):

(3) $$EF(s, y^*) \to \max_{s \in M}$$

(4) $$y^* \in Arg \max_{y \in A^2} Ef(s, y)$$

может быть решена двушаговым методом[1] [15, 56, 125]. Отметим, что рассматриваемая постановка непосредственно обобщается на случай любого конечного числа периодов.

Понятно, что вычислительная сложность даже двухпериодной задачи намного выше, чем статической. Редуцировать динамическую задачу к статической удается лишь в крайне ограниченном числе случаев (см. [140] – использование условий Куна-Таккера и сведение к вариационной задаче, [119] – использование подхода первого порядка [15, 125, 132, 134]).

Так как в рассмотренной выше модели стимулирование во втором периоде зависит и от результатов первого периода, то контракт, являющийся решением задачи (3)-(4) и обладающий этим свойством, называется *контрактом с памятью* (условия независимости вознаграждений в различных периодах обычно формулируются в терминах свойств функций распределения). Если в каждом периоде АЭ стимулируется только по результатам текущего периода, то контракт называется *контрактом без памяти* [100]. Основной вопрос, возникающий при изучении динамических контрактов, заключается в выяснении преимуществ, которыми обладает динамический контракт со связанными периодами и памятью, по сравнению с последовательностью обычных однопериодных контрактов[2].

Решение однопериодной вероятностной задачи – равновесные по Нэшу платежи (значения целевых функции центра и АЭ, соответственно) – F' и f', как правило, неэффективны и доминируются по Парето другими платежами F^* и f^* [95, 132, 138] (см. описание различий FB (first-best) и SB решений (second-best) и роли неопре-

---

[1] *В двушаговом методе на первом шаге ищутся минимальные системы стимулирования, реализующие заданную пару действий (по одному для каждого периода функционирования). На втором шаге перебором по всем допустимым парам находятся оптимальная с точки зрения центра пара реализуемых действий.*

[2] *Обычно в моделях рыночной экономики предполагается, что если число АЭ "велико", то игра некооперативная, а если "мало", то – кооперативная. В динамических моделях возможность кооперации появляется именно из-за динамики – элементы имеют время "договориться" и наказать тех, кто отклоняется от соглашений (см. описание стратегий наказания выше).*

деленности в [15, 57, 118]). То есть в последовательности одноэлементных контрактов средние платежи равны F' и f', а в динамическом контракте, в соответствии с FTh, они могут достигать или приближаться[1] к F^* и f^* [140].

В то же время, если в однопериодном контракте центр может достаточно сильно наказывать АЭ (соответствующие условия на ограничения механизма стимулирования приведены в [122]), то последовательное заключение краткосрочных контрактов оказывается не менее эффективно, чем заключение долгосрочного контракта. Иными словами, если долгосрочный контракт реализует некоторую последовательность действий [15], то при "достаточно сильных" штрафах, существует оптимальная последовательность краткосрочных контрактов, реализующая ту же последовательность и дающая всем участникам те же значения ожидаемой полезности. Содержательно, возможная сила штрафов должна быть такова, чтобы за их счет достаточно сильно наказать АЭ за отклонение именно в однопериодном контракте (в динамике эту роль играют стратегии наказания, используемые в следующих периодах), то есть триггерная стратегия каждого из игроков – выбор равновесной по Парето стратегии до тех пор, пока партнер выбирает равновесную по Парето стратегию, если же партнер «переключается» на равновесие Нэша, то следует тоже переключиться на соответствующую равновесную по Нэшу стратегию. В условиях вероятностной неопределенности возникает задача идентификации – построения оптимальных для центра триггерных стратегий, то есть определения оптимальных моментов переключения на стратегию наказания по наблюдениям результатов деятельности[2] в прошлых периодах (истории игры). В [131] доказано, что при достаточно общих предположениях у центра в конечной игре существует стратегия, обес-

---

[1] *Обычно результаты об оптимальности (достижимости FB решения) требуют бесконечного повторения подыгр, а для конечного числа периодов доказывается e-оптимальность* [131]. *При отсутствии дисконтирования любое IR Парето-оптимальное распределение выигрышей в однопериодной игре (в частности – FB решение) является достижимым Парето-оптимальным распределением выигрышей в суперигре* [123, 138].

[2] *Напомним, что в задачах теории контрактов результат деятельности АЭ является случайной величиной, зависящей от ненаблюдаемого центром действия АЭ и состояния природы.*

печивающая e-Парето-оптимальные значения целевых функций. Идея доказательства этого и подобных утверждений очевидна (см. обсуждение FTh выше). При построении и проверке статистических гипотез существенным оказывается то, как АЭ дисконтирует будущее: чем меньше элемент дисконтирует будущие полезности, тем ближе можно приблизиться к эффективному равновесию в суперигре. Более того, в бесконечных играх могут существовать критические значения дисконтирующих множителей, при превышении которых равновесие в суперигре строго доминирует однопериодные равновесия Нэша [134].

Иллюстрации использования приведенных выше теоретических результатов в прикладных моделях для таких областей, как трудовые контракты, теория заключения сделок, долговые контракты, модели покупки-продажи и т.д. описаны в [89, 98, 112, 113, 116, 117, 124, 125, 136, 137, 141, 146].

Пересоглашение контрактов

Достаточно специфический класс моделей теории контрактов, обычно относимых к динамическим моделям, составляют так называемые модели *пересоглашения контрактов*, кратко рассматриваемые ниже.

Наличие нескольких периодов функционирования, а также зависимость результата деятельности АЭ от внешнего неопределенного фактора (состояния природы) – все это обуславливает возможность пересмотра условий контракта, что должно, естественно, предусматриваться механизмом функционирования. Захотят ли стороны, подписавшие контракт, получив новую информацию, пересматривать его условия; возможно ли создать контракт, устойчивый по отношению к перезаключению (renegotiation-proof contract). Модели, в которых исследуются эти вопросы, рассматриваются в настоящем разделе.

Следует отметить, что рассмотрение контрактов с пересоглашением имеет смысл только в системах с неопределенностью, в том числе – с вероятностной неопределенностью, когда результат деятельности АЭ определяется как его действием, так и реализацией некоторой случайной величины – состояния природы. В этом случае привлекательность контрактов с пересоглашением обусловлена тем, что они позволяют реализовывать одно и то же действие

АЭ (даже в вероятностной АС) с меньшими затратами, иногда равными затратам на стимулирование в соответствующей детерминированной активной системе.

Рассмотрим одноэлементную вероятностную АС. Общепринятым в теории контрактов является следующий порядок функционирования [15]: центр выбирает функцию стимулирования и сообщает ее АЭ, элемент выбирает действие, реализуется состояние природы (a'priori, и центр, и АЭ знают лишь распределение его вероятностей), определяющее совместно с действием АЭ конкретное значение результата его деятельности; затем, в зависимости от результата деятельности, определяются значения целевых функций центра и элемента.

Пересоглашение допускается в так называемой промежуточной (interim) фазе однопериодного контракта – когда действие уже выбрано, а результат деятельности еще не наблюдается.

Фактически, центр должен предложить АЭ целое меню контрактов – каждый для определенного действия. Контракт является *защищенным от пересоглашения*, если он не перезаключается ни в одном из равновесий промежуточной стадии [94, 97]. Перезаключение контракта как бы страхует АЭ от последствий неблагоприятного для него результата деятельности, при "хорошем" действии [90, 106, 119].

Защищенным от перезаключения является контракт, принадлежащий множеству контрактов, удовлетворяющих условиям сообщения элементом в промежуточной стадии достоверной информации, условиям индивидуальной рациональности (выбираемое действие максимизирует ожидаемую полезность АЭ) и минимизирующий ожидаемые затраты центра на стимулирование [22, 58, 106].

Рассмотрим модель пересоглашения, следуя, в основном, [111], и попытаемся выяснить, какими преимуществами обладают механизмы стимулирования, предусматривающие возможность пересоглашения. Последовательность функционирования такова: центр и АЭ заключают начальный контракт; АЭ выбирает ненаблюдаемое для центра действие; центр получает от АЭ некоторую информацию о его действии; реализуется ненаблюдаемое участниками состояние природы; центр предлагает АЭ новый контракт (возможно пересоглашение); реализуется наблюдаемый центром результат деятельности АЭ, в соответствии с начальным или новым контрактом (в случае, если пересоглашение произошло) определяются полезности участников.

Возможность пересоглашения не изменяет условия реализуемости действий ни в случае, когда они наблюдаются центром (FB), ни в случае, когда они не наблюдаются (SB). То есть, достоинство контрактов с пересоглашением не в том, что они имеют более широкое множество реализуемых действий (в рамках моделей ТАС, на самом деле, при ограниченных функциях стимулирования использование пересоглашения в одноэлементной модели расширяет множество согласованных планов). Их основное преимущество – снижение затрат на стимулирование по реализации фиксированного действия (эти затраты сводятся к затратам, соответствующим детерминированному случаю).

Прокомментируем это утверждение. Пусть необходимо реализовать некоторое действие. Тогда в равновесии условие индивидуальной рациональности должно быть существенным, АЭ выберет это действие, и центр может предложить ему перезаключить исходный контракт на другой контракт, в котором АЭ выбирает то же действие и получает ту же полезность, что и в исходном контракте, а затраты на стимулирование равны затратам АЭ по выбору реализуемого действия. Таким образом, если действие элемента известно центру и он имеет возможность предложить перезаключить контракт, то множество реализуемых действий остается таким же, как и при отсутствии возможности пересоглашения, но любое действие реализуется с FB-затратами [111]. В частности, если носитель распределения результатов деятельности совпадает со всем множеством реализуемых действий, то затраты на реализацию любого действия, кроме действий с минимальными затратами, в SB-случае строго больше, чем в FB-случае [57, 125]. Значит в контрактах с пересоглашением значение целевой функции центра выше (а, следовательно, выше и эффективность механизма стимулирования), чем в контрактах без пересоглашения.

Содержательно, в контрактах с пересоглашением, в силу принципа открытого управления (в системе с одним АЭ для любого механизма существует механизм открытого управления не меньшей эффективности [17-20]), центр получает достоверную информацию о действиях, выбираемых элементом, и, следовательно, может стимулировать АЭ за действие, а не за случайный результат деятельности. Стимулирование в этом случае не менее эффективно, то есть повышение эффективности при использовании контрактов с пересоглашением происходит за счет получения центром достоверной информации о действиях элемента.

Приведенный результат позволяет сформулировать принцип защищенности от пересоглашения (renegotiation-proofness principle): в одноэлементной АС с вероятностной неопределенностью и возможностью пересоглашения без потери общности можно ограничиться рассмотрением контрактов без пересоглашения, так как все стороны могут включить результаты и последствия использования пересоглашения в первоначальный контракт [93, 121] (ср. с формулировкой и доказательством принципа выявления [16, 58, 107, 116, 117, 128]).

К сожалению, приведенный результат справедлив только в одноэлементных системах, так как в многоэлементных АС принцип выявления и утверждение о существовании для любого механизма эквивалентного механизма открытого управления не имеют места [16, 58].

В ряде случаев удается редуцировать многоэлементную или динамическую задачу к одноэлементной и статической, соответственно, и воспользоваться принципом выявления. Если, например, в многоэлементной АС неизвестные центру характеристики АЭ взаимосвязаны параметрически, то вместо решения многоэлементной задачи – сбора информации от всех АЭ, центру достаточно получить оценку параметра, то есть задача становится "одноэлементной". Аналогичный эффект агрегирования имеет место и в некоторых динамических задачах, когда, например, параметрически определяется плановая траектория [12].

Если в многоэлементной системе на промежуточной фазе центр предлагает элементам независимые контракты, то, очевидно, на этот случай результат принципа защищенности от пересоглашения обобщается непосредственно. Если же предлагаемые центром к пересоглашению контракты взаимозависимы, то неманипулируемость такого механизма требует дополнительного исследования. Поэтому вопрос о том, обладает ли пересоглашение преимуществами в многоэлементных системах, в общем случае, на сегодняшний день остается открытым (если имеются несколько АЭ и они наблюдают действия друг друга, то достаточно широкий класс механизмов (но не любой механизм!) может быть реализован (см. обзор результатов теории реализуемости и ссылки в [16, 127]) в случае, когда АЭ в промежуточной стадии посылают центру сообщения не только о себе, но и о других АЭ (всех или некоторых); при этом сообщение достоверной информации оказывается равновесием.

Отдельный класс моделей посвящен исследованию перезаключения контрактов в системах с асимметричной информированностью и сообщением информации большую роль играет информация о возможности пересоглашения [98, 113, 144].

Выше мы кратко описали пересоглашение контрактов в однопериодной модели, хотя, конечно, стадия пересоглашения может рассматриваться и как отдельный период, поэтому контракты с пересоглашением относят, как правило, к динамическим контрактам[1], хотя «полноценная» динамика (смысле, используемом в настоящей работе) в них отсутствует.

Активные системы с динамикой модели ограничений и адаптивные механизмы управления

В теории активных систем (ТАС) исследование динамики функционирования проводилось, в основном, для следующей модели.

---

[1] *В идеальной экономике все участники должны были бы заключать долговременные контракты, учитывающие все будущие возможности. Однако наличие неопределенности и недостаточная информированность на практике приводит к тому, что долгосрочные контракты встречаются достаточно редко, так как трудно учесть все возможные будущие ситуации. В описанной выше модели АЭ сообщал информацию о своем действии, не зная, какова будет реализация состояния природы, т.е. в промежуточной стадии никто из игроков не имел большей информации о неопределенных факторах, чем первоначально. Новые задачи возникают в случае, когда игроки пересматривают условия взаимоотношений в динамике, по мере поступления новой информации (см., например, использование переоценки и прогноза в модели простого АЭ* [12]). *Некоторые частные модели, учитывающие эту возможность, рассмотрены в* [92, 96, 97, 135].

В активной системе (АС), состоящей из центра и одного[1] активного элемента (АЭ), целевая функция центра в периоде t имеет вид

(1) $F^t(x^t; y^t)$,

а активного элемента

(2) $f^t(x^t; y^t)$,

где x^t – план на период t (желательное с точки зрения центра состояние АЭ), y^t – действие, выбранное АЭ в этом периоде.

Траектория $x^{1,T} = (x^1, x^2, \ldots, x^T)$ называется *плановой траекторией*, а траектория $y^{1,T} = (y^1, y^2, \ldots, y^T)$ – *траекторией реализаций*. Как и в одноэлементной статической задаче, центр выбирает систему стимулирования и устанавливает планы (на каждый период), а АЭ выбирает действие, максимизирующее его целевую функцию. Возникает вопрос – что понимать под целевой функцией АЭ в этой повторяющейся игре. Если допустимые множества не изменяются со временем и АЭ вообще не учитывает будущего (*недальновидный АЭ*), то задача сводится к набору статических задач.

Достаточно детально в ТАС были изучены так называемые *активные системы с динамикой модели ограничений*. Изменение модели ограничений (допустимых множеств) со временем учитывается зависимостью множества допустимых действий АЭ в периоде t от его действий в предыдущем периоде и от плана текущего периода, то есть $A^t = A^t(x^t, y^{t-1})$, $t \geq 2$, $A^1 = A^1(x^1)$ [2, 3, 19]. Таким образом, при известной плановой траектории недальновидный АЭ будет решать задачу поиска траектории реализаций:

(3) $f^t(x^t; y^t) \to \max\limits_{y_t \in A_t(x_t, y_{t-1})}$, $t = \overline{1, T}$.

Целевая функция дальновидного АЭ имеет вид:

(4) $\varphi_t(x^{1,T}, y^{1,T}) = f^t(x^t, y^t) + \sum\limits_{k=t+1}^{\tau} \delta^k f^k(x^k, y^k)$.

Для верхнего индекса суммирования в (4) возможны следующие варианты: $\tau = \min\{t + \tau_0, T\}$ – фиксированный горизонт τ_0 – АЭ учитывает τ_0 будущих периодов; $\tau = T$ – АЭ учитывает все будущие периоды и т.д. [18, 78-80]. То есть, дальновидный АЭ в каждом

[1] *В случае одного АЭ индекс, обозначающий его номер, будет опускаться.*

периоде t решает задачу выбора реализаций (действий – y^{t,t_0}) с целью максимизации (4).

Задача центра заключается в выборе плановой траектории, максимизирующей его целевую функцию

$$F(x^{1,T}, y^{1,T}) = \sum_{k=1}^{T} d^k f^k(x^k, y^k),$$

считая, что при выполнении условий согласования (см. ниже) реализации будут совпадать с планами [79]. Из принципа оптимальности Беллмана следует, что, если распределение дальновидности АЭ «жестко привязано» к периодам функционирования, то прогноз, сделанный в первом периоде, совпадает с реализацией в последующих периодах, а прогнозы в последующих периодах совпадают с прогнозом первого периода. Если АЭ и центр имеют различные степени дальновидности ($t_0 + 1 < T$), то АЭ не может построить прогноз на весь плановый период.

В [79] приведены условия на распределения дальновидностей, обеспечивающие совпадение реализации с планом, и показано, что в ряде случаев динамическую задачу удается свести к статической, решаемой в «расширенном» пространстве параметров.

Приведенная выше задача (3)-(4) является одним из частных случаев задачи управления динамическими активными системами. В [36, 37] в качестве одного из оснований классификации динамических задач выделялся режим управления, используемый центром. В качестве возможных режимов центр может применять *программное планирование и управление* (в рамках которого центр в начале планового периода формирует плановую траекторию[1] $x^{1,T}$ и в дальнейшем не изменяет ее), *скользящий режим* (в рамках которого центр в начале планового периода формирует плановую траекторию и в дальнейшем корректирует ее по мере поступления новой информации) и *текущий режим*, когда центр принимает в каждом периоде решения, касающиеся только этого периода (см. также настоящую работу).

---

[1] *Вопросы реализуемости, согласованности и оптимальности плановых траекторий исследовались для различных режимов управления, а также для различных комбинаций распределения дальновидностей, в* [78-80].

Условия согласования.

Значительное число работ в теории активных систем посвящено исследованию задач согласованной оптимизации. Для их рассмотрения опишем кратко одноэлементную статическую задачу. Пусть система стимулирования $s(\cdot)$ зависит от параметра – плана $x\ \hat{I}\ X$ – желательного с точки зрения центра состояния АЭ – и действия АЭ $y\ \hat{I}\ A$, где X – множество допустимых планов (для простоты положим $X = A$): $s = s(x, y)$. Тогда целевая функция АЭ $f(\cdot)$, представляющая собой разность между стимулированием, получаемым от центра, и затратами агента, зависит от стимулирования, плана и действия АЭ: $f = f(s, x, y)$. Множество реализуемых действий также параметрически зависит от плана: $P(s, x) = Arg \max\limits_{y \in A} f(s, x, y)$. Изменяя планы, центр может системой стимулирования $s(\cdot, y)$ реализовать следующее множество действий: $P(s) = \bigcup\limits_{x \in X} P(s, x)$.

Обозначим $B(s) = \{x\ \hat{I}\ X \mid {}''\ y\ \hat{I}\ A\ s(x, x) - c(x)\ {}^3\ s(x, y) - c(y)\}$ *множество согласованных планов*, то есть таких планов, выполнять которые при заданной системе стимулирования для АЭ выгодно. Задавая систему стимулирования $s(x, y)$, центр имеет возможность оперативно изменять значения планов, не меняя функцию стимулирования, что достаточно привлекательно, так как особенно в динамике частые изменения целиком всего механизма управления не всегда возможны с точки зрения адаптивных свойств АЭ.

Согласованной называется система стимулирования $s\ \hat{I}\ M$, для которой выполнено $B(s) = P(s)$. Значительное внимание исследователей уделялось поиску необходимых и достаточных условий согласованности систем стимулирования, а также изучению соотношения таких свойств как согласованность и эффективность систем стимулирования – подавляющее большинство работ в ТАС на рубеже 70-80 годов было посвящено именно этой тематике. В работах по теории активных систем рассматривался целый ряд требований согласования интересов центра и АЭ, формулируемых как необходимость обеспечения требуемых соотношений между планами активных элементов и их реализациями (выбором – действиями АЭ). Среди них: механизмы, согласованные по выполнению плана (см. определение выше) в системах с полным, частичным и агрегированным планированием, x-согласованные механизмы, $D(x)$-согласованные механизмы, L-согласованные механизмы [6, 14, 18, 19] и др. – см. обзор в [20]. В упомянутых работах развиваются как методы решения задачи синтеза оптимальных механизмов функционирования, так и задачи синтеза оптимальных механизмов функционирования, согласованных по выполнению плана.

Наиболее известным и изящным достаточным условием согласованности *системы штрафов* $c(x, y)$ (для задачи стимулирования, в которой целевая функция АЭ представляет собой разность между доходом и штрафами – эта постановка является «двойственной» к описанной выше модели, в которой целевая функция АЭ определяется разностью между стимулированием и затратами [58]) является так называемое «неравенство треугольника» [18]:

$$\forall\ x, y, z \quad c(x,y) \le c(x,z) + c(z,y).$$

Описание достаточных условий согласованности можно найти в [6, 14, 19].

Вернемся к рассмотрению динамических моделей. Пусть при решении задачи планирования центр предполагает, что реализации совпадут с планами. Известно, что достаточным условием согласованности системы стимулирования в статической АС является выполнение неравенства треугольника для функций штрафов. Вопросы согласованности управления в динамических моделях типа (3)-(4) и др. исследовались в [31, 36, 78, 79]. В частности, доказано, что для согласованности в динамической модели достаточно выполнения неравенства треугольника для взвешенных сумм штрафов. Если в течение нескольких периодов штрафы не являются согласованными, то для согласования в динамике достаточно существования сильных штрафов в будущем (см. стратегии наказания выше). В упомянутых же работах исследовалась взаимосвязь между согласованностью управления в динамических моделях и распределением дальновидности участников системы при различной степени централизации.

Рассмотренная выше модель ограничений зависела от параметров, выбираемых участниками системы. Однако возможны случаи, когда допустимые множества зависят от случайных параметров (или когда, как в повторяющихся играх при неполной информированности, не все выбираемые стратегии наблюдаемы).

Следовательно возникает задача идентификации, решаемая при использовании адаптивных механизмов функционирования.

Адаптивные механизмы управления.

Основная идея адаптивных механизмов управления[1] заключается в следующем [2, 3,75].

В *механизмах с адаптивной идентификацией* проводится предварительное восстановление[2] оценочных множеств неопределенных параметров, которые затем используются при решении задачи синтеза оптимальных управлений на будущие периоды. В *адаптивных механизмах* (без идентификации) этап восстановления отсутствует, а задача синтеза решается непосредственно на основании наблюдаемых реализаций (истории игры).

Для решения задач идентификации [16] в активных системах применяются три подхода [21].

Адаптивная идентификация. Особенность задач адаптивной идентификации в активных системах состоит в том, что АЭ и центр могут иметь разные представления о том, какую модель следует получить в результате идентификации. Так как идентификация производится на основе оценки состояний активной системы, АЭ, имея определенную свободу выбора этих состояний, может влиять на результат идентификации. Для иллюстрации сказанного рассмотрим простой пример.

Пример 10. Пусть АС описывается скалярным параметром r, причем центр заинтересован в том, чтобы этот параметр был возможно больше, а АЭ заинтересован в обратном. Центр стимулирует

---

[1] *В отличие от моделей с априорной неопределенностью* [58], *в которых центр однократно (или многократно, но обладая одной и той же информацией) принимает решения в условиях неопределенности (в рамках моделей ТАС – как правило, интервальной внутренней неопределенности относительно множеств допустимых действий агентов), в динамике центр зачастую принимает решения в рамках текущей (изменяющейся при получении новой информации о существенных параметрах окружающей среды и управляемой системы в процессе ее функционирования) неопределенности.*

[2] *При построении механизмов управления с адаптивной идентификацией широко используются результаты теории адаптивного управления и теории идентификации* [11, 76, 77].

АЭ за рост наблюдаемого значения параметра. Представим целевую функцию АЭ в виде: $f(r) = A - q\, r_0 + a(r - r_0)$, где r_0 – имеющаяся у центра на начальный момент оценка параметра r, r – наблюдаемая в текущем периоде величина параметра r (остальные параметры – константы). Если центр наблюдает величину $r > r_0$, то он может уточнить оценку этого параметра в модели:

$$r_1 = r_0 + b\,(r - r_0).$$

Таким образом, при росте текущего значения параметра r АЭ выигрывает «сегодня», но поигрывает «завтра». Соизмерим сегодняшний выигрыш и завтрашний проигрыш некоторым коэффициентом γ, который характеризует степень дальновидности АЭ:

$$f_g(r) = A - q\, r_0 + a(r - r_0) + g\,(A - q\, r_1).$$

Анализ выражения для $f_g(r)$ показывает, что для того, чтобы АЭ был заинтересован в уточнении (точнее – увеличении) параметра r, должно выполняться условие $a > g\, b\, q$.

Подобные условия называются условиями прогрессивности механизма адаптивной идентификации. •

Введем ряд определений. Механизм функционирования называется *правильным*, если он обеспечивает совпадение реализации и плана (см. выше). Механизм функционирования называется *слабо прогрессивным*, если целевая функция АЭ монотонна по его действию[1]. Механизм называется *прогрессивным по плану*, если максимум целевой функции АЭ по множеству его действий, зависящему от идентифицируемого параметра, при плане, полученном в результате идентификации этого параметра, является возрастающей функцией последнего. Другими словами, правильный механизм побуждает АЭ выполнять план, слабо прогрессивный – выбирать максимальное допустимое действие на этапе реализации плана, а прогрессивный по плану – принимать и выполнять максимальные планы. *Сильно прогрессивным* называется механизм, который одновременно слабо прогрессивен и прогрессивен по плану.

В [2, 3, 72, 73, 75] рассматриваются модели АС, в которых сильной прогрессивности механизма функционирования достаточно для точной идентификации детерминированной или стохастической модели АЭ. Например, пусть множество возможных действий

---

[1] *Следует отметить, что, как правило, считается, что множество допустимых действий «монотонно» по неопределенному параметру* [75].

зависит от неизвестного центру потенциала АЭ, а потенциал, в свою очередь, зависит от управления со стороны центра и некоторой случайной величины. На основании наблюдаемой реализации центр может определить оценку потенциала с помощью той или иной рекуррентной процедуры прогнозирования [75]. Примером решения задачи адаптивного планирования может служить модель динамического простого АЭ, подробно описанная в [12]. Аналогичные процедуры используются в динамических задачах теории контрактов – оценка потенциала входит в статистическую гипотезу, проверяемую в повторяющихся играх для определения факта отклонения элемента от эффективной стратегии (см. выше).

Если предпочтения АЭ неизвестны центру или известны с некоторой погрешностью, то дополнительная информация об этих предпочтениях может быть получена следующим образом. Если выполнена гипотеза рационального поведения, то при известном управлении со стороны центра АЭ выбирает действия, максимизирующие его целевую функцию. Следовательно, если известно множество пар $W_m = \{(u_1, y_1), (u_2, y_2), \dots (u_m, y_m)\}$ управлений и действий, выбранных при этих управлениях, то на основании информации о W_m могут проверяться те или иные гипотезы о предпочтениях АЭ. Задача определения: оптимальной (по тому или иному критерию) последовательности управлений, их числа, методов «уточнения» параметров модели и т.д. называется *задачей активной адаптивной идентификации*[1] (активная адаптивная идентификация существенно использует идею дуального управления).

Пример 11. Рассмотрим следующую задачу стимулирования в одноэлементной активной системе. Целевая функция центра представляет собой разность между доходом от деятельности АЭ и затратами на стимулирование (в данном классе задач управлением со стороны центра является функция стимулирования АЭ $s(\cdot)$):

[1] *Так как рассматриваются модели социально-экономических систем, то использование активной адаптивной идентификации возможно далеко не всегда. По крайней мере, в общем случае критерий эффективности идентификации (критерий эффективности управления) должен учитывать не только эффективность управления, достигнутую в результате идентификации АС, но и потери, вызванные наличием периода активной идентификации.*

$F(s, y) = H(y) – s(y)$. Целевая функция АЭ является разностью между стимулированием и его затратами: $f(s, y) = s(y) – c(y)$.

Пусть множество допустимых действий АЭ имеет вид: $A = [0; A^+]$, где $0 < A^+ < +$¥, а затраты АЭ: $c(y) = y^2/2r$, где $r > 0$ – некоторый параметр.

Предположим, что центру известно, что множество M_1, которому заведомо принадлежит управляемая система задано в виде:

" y Î A $c^-(y)$ £ $c(y)$ £ $c^+(y)$, то есть r Î $[r^+; r^-]$.

Из условия неотрицательности целевой функции АЭ следует [58], что в условиях существующей неопределенности оптимальной (и использующей максимальный гарантированный результат) системой стимулирования будет компенсаторная система стимулирования: $s_K(y) = c^+(y)$, а оптимальным реализуемым действием действие $y^+ = Arg \max\limits_{y \in A} \{H(y) – s_K(y)\}$.

Если бы функция затрат АЭ была бы достоверно известна центру, то оптимальным было бы действие $y^* = Arg \max\limits_{y \in A} \{H(y) – c(y)\}$.

Потери эффективности, вызванные неполной информированностью, равны:

$$e = \{H(y^*) – c(y^*)\} – \{H(y^+) – c^+(y^+)\} \text{ ³ } 0.$$

Если, например, $H(y) = y$, то $e = (r – r^+) / 2$.

Итак, мы рассмотрели задачу управления в условиях интервальной неопределенности относительно параметров функции затрат АЭ. Перейдем к рассмотрению активной адаптивной идентификации.

Если центр устанавливает систему стимулирования $s_K(y, \tilde{r}) = y^2/2\tilde{r}$, то недальновидный АЭ выбирает действие

$$y^*(\tilde{r}) = \begin{cases} 0, \text{если } r < \tilde{r} \\ A^+, \text{если } r > \tilde{r} \end{cases}.$$

Следовательно задача свелась к определению оптимальной последовательности значений параметра $\tilde{r}$. •

Помимо задач активной адаптивной идентификации в теории активных систем рассматривались задачи встречной идентификации[1] и многоканальной идентификации[2].

При исследовании адаптивных механизмов возникают задачи выбора наилучшей процедуры прогнозирования; синтеза механизма, при котором АЭ полностью использует свой потенциал (такие механизмы получили название прогрессивных); определения реальности плановых траекторий; синтеза оптимального механизма управления и т.д. Останавливаться более подробно на описании методов решения этих задач и полученных результатов мы не будем.

<u>Последовательный синтез адаптивных механизмов функционирования.</u>

Основная идея решения задачи последовательного синтеза механизмов управления (как следует из самого названия этого класса задач управления) заключается в следующем: выделяются требования (ограничения) к механизму управления: S_1, S_2, S_3, ... ; S_i Í S,

---

[1] *Идея встречной идентификации заключается в том, что по сути идентификацией своей модели занимается сам АЭ. Результаты идентификации (значения параметров модели) он сообщает центру. Основная проблема здесь заключается в создании заинтересованности АЭ в предоставлении центру достоверной информации о параметрах модели. Методы встречной идентификации основаны, как правило, на механизмах «честной игры», которые обладают свойством неманипулируемости* [16].

[2] *Суть подхода к идентификации на основе многоканальных механизмов состоит в получении центром информации о модели АС по нескольким каналам (от нескольких источников), в числе которых могут быть управляемые активные элементы, эксперты, адаптивная компьютерная модель и др.* [14, 22]. *Центр на основе этой информации идентифицирует модель, которая и применяется для принятия решений. После того как принятое решение реализовано и оценена его эффективность, производится оценка эффективностей моделей, предложенных различными каналами (точнее – оценка эффективности решения, которое было бы принято на основе этой модели). В зависимости от сравнительной эффективности принятого решения и решений каналов производится стимулирование каналов, что побуждает последних предлагать эффективные решения, что, в свою очередь, приводит к улучшению информированности центра о параметрах управляемой системы.*

где S – множество допустимых механизмов управления. На первом шаге конструктивно определяется множество (класс) механизмов S_1 (то есть приводятся необходимые, достаточные или одновременно необходимые и достаточные условия, которым должен удовлетворять механизм, принадлежащий этому классу). Затем на втором шаге определяется множество S_2 Ç S_1, и т.д.

Возникающая при этом проблема состоит поиске таких необходимых и/или достаточных условий, описывающих тот или иной класс механизмов, которые, с одной стороны, были бы достаточно простыми и легко верифицируемыми (в первую очередь это требование относится к достаточным условиям S_i'), а с другой стороны позволяли бы получить решение задачи ($\mathbf{I}_{i} \Sigma_i'$ ¹ Æ ни на одном шаге). Подробно результаты исследования проблемы последовательного синтеза механизмов адаптивных функционирования динамических активных систем рассмотрены в [2, 3].

ЛИТЕРАТУРА

1. Айзекс Р. Дифференциальные игры. М.: Мир, 1967.
2. Андреев С.П. Синтез процедур адаптивной идентификации моделей ограничений активных элементов / Механизмы управления социально-экономическими системами. М.: Институт проблем управления, 1988. С. 32 – 36.
3. Андреев С.П. Синтез оптимальных в одном классе адаптивных механизмов функционирования активных систем // А. и Т., 1985. № 12. С. 72 – 78.
4. Ануфриев И.К., Бурков В.Н., Вилкова Н.И., Рапацкая С.Т. Модели и механизмы внутрифирменного управления. М.: ИПУ РАН, 1994. – 72 с.
5. Аткинсон Р. Человеческая память и процесс обучения. М.: Прогресс, 1980. – 528 с.
6. Ашимов А.А., Бурков В.Н., Джапаров Б.А., Кондратьев В.В. Согласованное управление активными производственными системами. М.: Наука, 1986. – 248 с.
7. Ашманов С.А. Введение в математическую экономику. М.: Наука, 1984. – 296 с.
8. Барабанов И.Н., Новиков Д.А. Механизмы управления динамическими активными системами и модели коллективного поведения / 3-я Украинская конференция по автоматическому управлению. 9-14 сент. 1996 г., Севастополь. Том 2. С. 4 – 5.
9. Беллман Р. Динамическое программирование. Москва, 1960. – 400 с.
10. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления. М.: Наука, 1969. – 120 с.
11. Болтянский В.Г. Математические методы оптимального управления. М.: Наука, 1968. – 408 с.
12. Бурков В.Н. Основы математической теории активных систем. М.: Наука, 1977. – 255 с.
13. Бурков В.Н., Горгидзе И.А., Ловецкий С.Е. Прикладные задачи теории графов. Тбилиси: Мецниереба, 1974. – 234 с.
14. Бурков В.Н., Данев Б., Еналеев А.К. и др. Большие системы: моделирование организационных механизмов. М.: Наука, 1989. – 245 с.
15. Бурков В.Н., Еналеев А.К., Новиков Д.А. Механизмы стимулирования в вероятностных моделях социально-экономических систем // Автоматика и Телемеханика. 1993. № 11. С. 3 – 30.
16. Бурков В.Н., Еналеев А.К., Новиков Д.А. Механизмы функционирования социально-экономических систем с сообщением информации // Автоматика и Телемеханика. 1996. № 3. С. 3 – 25.

17. Бурков В.Н., Ириков В.А. Модели и методы управления организационными системами. М.: Наука, 1994. – 270 с.
18. Бурков В.Н., Кондратьев В.В. Механизмы функционирования организационных систем. М.: Наука, 1981. – 384 с.
19. Бурков В.Н., Кондратьев В.В., Цыганов В.В., Черкашин А.М. Теория активных систем и совершенствование хозяйственного механизма. М.: Наука, 1984. – 272 с.
20. Бурков В.Н., Новиков Д.А. Введение в теорию активных систем. М.: ИПУ РАН, 1996. – 125 с.
21. Бурков В.Н., Новиков Д.А. Идентификация активных систем / Труды международной конференции «Идентификация систем и процессы управления». М.: ИПУ РАН, 2000. – С. 101.
22. Бурков В.Н., Новиков Д.А. Как управлять проектами. М.: Синтег, 1997. – 188 с.
23. Бурков В.Н., Новиков Д.А. Теория активных систем: состояние и перспективы. М.: Синтег, 1999 – 128 с.
24. Варшавский В.И. Коллективное поведение автоматов. М.: Наука, 1973. – 408 с.
25. Венда В.Ф. Системы гибридного интеллекта: эволюция, психология, информатика. М.: Машиностроение, 1990. – 448 с.
26. Вентцель Е.С. Элементы динамического программирования. М: Наука, 1964. – 176 с.
27. Вилкас Э.Й. Оптимальность в играх и решениях. М.: Наука. 1990. – 256 с.
28. Волкович В.Л., Михалевич В.С. Вычислительные методы исследования и проектирования сложных систем. М.: Наука, 1982. – 286 с.
29. Вудвортс Р. Экспериментальная психология. М.: Изд-во ин. лит., 1950. – 800 с.
30. Гермейер Ю.Б. Игры с непротивоположными интересами. М.: Наука, 1976. – 327 с.
31. Горгидзе И.А., Жвания В.В., Кондратьев В.В., Щепкин А.В. Правильное согласованное планирование в активных системах с динамикой модели ограничений / Планирование, оценка деятельности и стимулирование в активных системах. М.: ИПУ РАН, 1985. С. 54 – 63.
32. Горелик В.А., Кононенко А.Ф. Теоретико-игровые модели принятия решений в эколого-экономических системах. М.: Радио и связь, 1982. – 144 с.
33. Данильченко Т.Н., Мосевич К.К. Многошаговая игра двух лиц с фиксированной последовательностью ходов // ЖВМ и МФ. 1974. Т. 14. № 4. С. 1047 – 1052.

34. Данильченко Т.Н., Мосевич К.К. Многошаговая игра двух лиц при «осторожном» втором игроке и последовательной передачей информации // ЖВМ и МФ. 1974. Т. 14. № 5. С. 1323 – 1327.
35. Егоршин А.П. Управление персоналом. Н.Новгород: НИМБ, 1997. – 607 с.
36. Жвания В.В. К вопросу получения достаточных условий оптимальности правильных механизмов функционирования активных систем с динамикой модели ограничений // А. и Т. 1986. N 2. С. 160-163.
37. Жвания В.В. Оптимальный синтез систем стимулирования в активных системах с динамикой модели ограничений / Планирование, оценка деятельности и стимулирование в активных системах. М.: ИПУ РАН, 1985. С. 64 – 68.
38. Жуковский В.И., Салуквадзе М.Е. Некоторые игровые задачи управления и их приложения. Тбилиси: Мецниереба, 1998. – 462 с.
39. Колмановский В.Б. Игровые задачи управления. М.: МИЭМ, 1990. – 82 с.
40. Кононенко А.Ф. О многошаговых конфликтах с обменом информацией // ЖВМ и МФ. 1977. Т. 17. № 4. С. 922 – 931.
41. Кононенко А.Ф., Халезов А.Д., Чумаков В.В. Принятие решений в условиях неопределенности. М.: ВЦ АН СССР, 1991. – 197 с.
42. Кочиева Т.Б., Новиков Д.А. Базовые системы стимулирования. М.: Апостроф, 2000. – 108 с.
43. Кочиева Т.Б., Новиков Д.А., Титов А.С. Теоретико-игровые модели стимулирования в задачах рекрутинга / Тезисы докладов XLI научной конференции МФТИ. 27-28 ноября 1998 г. Долгопрудный, 1998. Часть II. С. 38.
44. Красс И.А. Математические модели экономической динамики. М.: Советское радио, 1976. – 280 с.
45. Крылов В.Ю., Морозов Ю.И. Кибернетические модели и психология. М.: Наука, 1984. – 174 с.
46. Лотоцкий В.А. Идентификация структур и параметров систем управления // ИКА. 1991. №3-4. С. 30 – 38.
47. Маленво Э. Лекции по микроэкономическому анализу. М.: Наука, 1985. – 392 с.
48. Месарович М., Такахара И. Общая теория систем: математические основы. М.: Мир, 1978. – 311 с.
49. Мулен Э. Кооперативное принятие решений: аксиомы и модели. М.: Мир, 1991. – 464 с.

50. Новиков Д.А. Динамика поведения систем с большим числом целенаправленных элементов // Автоматика и Телемеханика. 1996. № 4. С. 187 – 189.
51. Новиков Д.А. Закономерности итеративного научения. М.: ИПУ РАН, 1998. – 96 с.
52. Новиков Д.А. Механизмы стимулирования в динамических и многоэлементных социально-экономических системах // Автоматика и Телемеханика. 1997. № 6. С. 3 – 26.
53. Новиков Д.А. Механизмы функционирования многоуровневых организационных систем. М.: Фонд "Проблемы управления", 1999. – 150 с.
54. Новиков Д.А. Модели и механизмы управления развитием региональных образовательных систем. М.: ИПУ РАН, 2001. – 83 с.
55. Новиков Д.А. Обобщенные решения задач стимулирования в активных системах. М.: ИПУ РАН, 1998. – 68 с.
56. Новиков Д.А., Петраков С.Н. Курс теории активных систем. М.: СИНТЕГ, 1999. – 108 с.
57. Новиков Д.А. Стимулирование в вероятностных активных системах: роль неопределенности // Автоматика и Телемеханика. 1997. № 8. С. 168 – 177.
58. Новиков Д.А. Стимулирование в социально-экономических системах (базовые математические модели). М.: ИПУ РАН, 1998. – 216 с.
59. Новиков Д.А., Цветков А.В. Механизмы стимулирования в многоэлементных организационных системах. М.: ИПУ РАН, 2001. – 188 с.
60. Новиков Д.А., Цветков А.В. Механизмы функционирования организационных систем с распределенным контролем. М.: ИПУ РАН, 2001.– 118с.
61. Опойцев В.И. Равновесие и устойчивость в моделях коллективного поведения. М.: Наука, 1977. – 248 с.
62. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. М.: Наука, 1981. – 206 с.
63. Оуэн Г. Теория игр. М.: Мир, 1971. – 230 с.
64. Первозванский А.А. Курс теории автоматического управления. М.: Наука, 1986. – 616 с.
65. Перегудов Ф.И., Тарасенко Ф.П. Введение в системный анализ. М.: Высшая школа, 1989. – 367 с.
66. Петросян Л.А., Зенкевич Н.А., Семина Е.А. Теория игр. М.: Высшая школа, 1998.- 304 с.
67. Поспелов Г.С., Ириков В.А. Программно-целевое планирование и управление. М.: Советское радио, 1976. – 344 с.

68. Современное состояние теории исследования операций. М.: Наука, 1979. – 464 с.
69. Столерю Л. Равновесие и экономический рост. М.: Статистика, 1974. – 302 с.
70. Трахтенгерц Э.А. Компьютерная поддержка принятия решений. М.: Синтег, 1998. – 376 с.
71. Фишер С., Дорнбуш Р., Шмалензи Р. Экономика. М.: Дело, 1993. – 864 с.
72. Цветков А.В. О выборе согласования в двухуровневой активной системе с неопределенностью / Планирование, оценка деятельности и стимулирование в активных системах. М. ИПУ РАН, 1985. С. 30 – 34.
73. Цветков А.В. Условия оптимальности согласованных механизмов функционирования при неопределенности / Неопределенность, риск, динамика в организационных системах. М.: ИПУ РАН, 1984. С. 73 – 81.
74. Цетлин М.Л. Исследования по теории автоматов и моделированию биологических систем. М.: Наука, 1969. – 316 с.
75. Цыганов В.В. Адаптивные механизмы в отраслевом управлении. М.: Наука, 1991. – 166 с.
76. Цыпкин Я.З. Адаптация и обучение в автоматических системах. М.: Наука, 1968. – 399 с.
77. Цыпкин Я.З. Основы информационной теории идентификации. М.: Наука, 1984. – 336 с.
78. Щепкин А.В. Динамические активные системы с дальновидными элементами. I. Динамическая модель активной системы // А. и Т. 1986. N 10. С. 89 – 94.
79. Щепкин А.В. Динамические активные системы с дальновидными элементами. II. Дальновидность активных элементов в динамических моделях // А. и Т. 1986. N 11. С. 82 – 94.
80. Щепкин А.В. Управление динамическими активными системами. Диссертация на соиск. уч. ст. к.т.н. М.: ИПУ РАН, 1980. – 130 с.
81. Эренберг Р.Дж., Смит Р.С. Современная экономика труда. Теория и государственная политика. М.: Изд-во МГУ, 1996. – 800 с.
82. Abreu D., Milgrom P., Pearce D. Information and timing in repeated partnership // Econometrica. 1991. Vol. 59. N 6. P. 1713 – 1733.
83. Abreu D., Dutta P., Smith L. The Folk theorem for repeated games : a NEU condition // Econometrica. 1994. Vol. 62. N 4. P. 939 – 948.
84. Abreu D. On the theory of infinitely repeated games with discounting // Econometrica. 1988. Vol. 56. N 2. P. 383 – 396.

85. Abreu D., Pearce D., Starcetti E. Toward a theory of discounted repeated games with imperfect monitoring // Econometrica. 1990. Vol. 58. N5. P. 1041 – 1063.
86. Armstrong M. Reward management. London, 2000. – 804 p.
87. Atkinson A.A., Neave E.H. An incentive scheme with desirable multiperiod properties // INFOR. 1983. V. 21. N 1. P. 76 – 83.
88. Aumann R.J., Mashler H.L. Repeated games with incomplete information. MIT Press, 1995. – 342 p.
89. Baron D., Besanko D. Commitment and fairness in a dynamic regulatory relationship // Rev. of Econ. St. 1987. V.54. N 3. P. 413 – 436.
90. Beaudry P., Poitevin M. Signaling and renegotiation in contractual relationships // Econometrica. 1993. Vol. 61. № 4. P. 745 – 781.
91. Benoit J.-P., Krishna V. Finitely repeated games // Econometrica. 1985. Vol. 53. N 4. P. 905 – 922.
92. Benoit J.-P., Krishna V. Renegotiation in finitely repeated games // Econometrica. 1993. Vol. 61. N 2. P. 303 – 323.
93. Bolton P. Renegotiation and the dynamics of contract design // European Economic Review. 1990. Vol. 34. N 2/3. P. 303 – 310.
94. Compte O. Communication in repeated games with imperfect private monitoring // Econometrica. 1998. Vol. 66. № 3. P. 597 – 626.
95. Crawford V.P. Long-term relationships governed by short-term contracts // AER. 1988. Vol. 78. N 3. P. 485 – 499.
96. Dewatripont M. Commitment through renegotiation-proof contracts with third parties // Review of economic studies. 1988. Vol. 55. N 3. P. 377 – 389.
97. Dewatripont M., Maskin E. Contract renegotiation in models of asymmetric information // European Economic Review. 1990. Vol. 34. N 2/3. P. 311 – 321.
98. Dewatripont M. Renegotiation and information revelation over time: the case of optimal labor contracts // Quarterly Journal of Economics. 1989. Vol. 104. N 3. P. 589 – 619.
99. Evans G. Sequential bargaining with correlated values // Review of economic studies. 1989. Vol. 56. N 4. P. 499 – 510.
100. Fellingham J.C., Newman D.P., Suh Y.S. Contracts without memory in multiperiod agency models // J. of Econ. Theory. 1985. V. 37. N 2. P. 340 – 355.
101. Fudenberg D., Holmstrom B., Milgrom P. Short-term contracts and long-term agency relationship // J. of Econ. Theory. 1990. V. 52. N 1. P. 194 – 206.
102. Fudenberg D., Kreps D. Reputation in the simultaneous play of multiple opponents // Review of economic studies. 1987. Vol. N 4. P. 541 – 568.

103. Fudenberg D., Levine D., Maskin E. The Folk theorem with imperfect public information // Econometrica, 1994. Vol. 62. N5. P. 997 – 1039.

104. Fudenberg D., Maskin E. The Folk theorem in repeated games with discounting or with incomplete information // Econometrica. 1986. Vol. 54. N 3. P. 533 – 554.

105. Fudenberg D., Tirole J. Sequential bargaining with incomplete information // Rev. of Econ. St. 1983. V.50. N2. P. 221 – 247.

106. Fudenberg D., Tirole J. Moral hazard and renegotiation in agency contracts // Econometrica. 1990. V.58. N 6. P. 1279 – 1319.

107. Fudenberg D., Tirole J. Game theory. Cambridge: MIT Press, 1995. – 579 p.

108. Harris M., Holmstrom B. A theory of wage dynamics // Rev. of Econ. St. 1982. V. 49. N 2. P. 315- 333.

109. Hart O.D., Moore J. Incomplete contracts and renegotiation // Econometrica. 1988. V. 56. N 4. P.755 – 785.

110. Hart O.D., Tirole J. Contract renegotiation and Coasian dynamics // Rev. of Econ. St. 1988. V.55. N4. P. 509 – 540.

111. Herman B.E., Katz M.L. Moral hazard and verifiability: the effects of renegotiation in agency // Econometrica. 1991.V. 59. N6. P. 1735 – 1753.

112. Holmstrom B. Equilibrium long-term labor contracts // Quarterly Journal of Economics. 1983. Vol. 98. N 3. Supplement. P. 23 – 54.

113. Holmstrom B., Myerson R. Efficient and durable decision rules with incomplete information // Econometrica. 1983. V.51. N6. P. 1799 – 1819.

114. Kalai E., Lahler E. Rational learning leads to Nash equilibrium / Game and Economic Theory. University of Michigan: University of Michigan Press, 1995. P. 89 – 111.

115. Kreps D., Wilson R. Sequential equilibria // Econometrica. Vol. 50. N 4. P. 863 – 894.

116. Laffont J.J. Fundamentals of public economics. Cambridge: MIT Press, 1989. – 289 p.

117. Laffont J.J. The economics of uncertainty and information. Cambridge: MIT Press, 1989. – 289 p.

118. Laffont J.-J., Tirole J. The dynamics of incentive contracts // Econometrica. 1988. V. 56. N 1. P. 7 – 29.

119. Lambert R.A. Long-term contracts and moral hazard // Bell J. of Econ. 1983. V. 14. N 3. P. 441 – 452.

120. Lehler E., Pauzner A. Repeated games differential time preferences // Econometrica. 1999. Vol. 67. № 2. P. 393 – 412.

121. Ma. C. Renegotiation and optimality in agency contracts // Review of Economic Studies. 1994. Vol. 61. N 1. P. 109 – 129.

122. Malcomson J.M., Spinnewyn F. The multiperiod principal – agent problem // Rev. of Econ. St. 1988. V. 55. N 3. P. 391 – 408.
123. Malueg D.A. Efficient outcomes in a repeated agency model with discounting // J. of Math. Econ. 1986. V.15. N 3. P. 217 – 230.
124. Marchak J., Radner R. Economic theory of teams. New Haven – London: Yale Univ. Press, 1976. – 345 p.
125. Mas-Collel A., Whinston M.D., Green J.R. Microeconomic theory. N.Y.: Oxford Univ. Press, 1995. – 981 p.
126. Matsushima H. Efficiency in repeated games with imperfect monitoring // Journal of Economic Theory. 1989. Vol. 98. N 2. P. 428-442.
127. Moore J. Implementation, contracts and renegotiation in environment with complete information / Advances in Economic Theory. Vol. 1. Cambridge: Cambridge University Press, 1992. P. 182 – 281.
128. Myerson R.B. Game theory: analysis of conflict. London: Harvard Univ. Press, 1991. – 568 p.
129. Nosal E. Contract renegotiation in a continuous state space // Economic Theory. 1997. Vol. 10. № 3. P. 413 – 436.
130. Perlman R. Labor theory. N.Y.: Wiley, 1969. – 237 p.
131. Radner R. Monitoring cooperative agreements in a repeated principal-agent relationship // Econometrica. 1981. V. 49. N 5. P. 1127 – 1148.
132. Radner R. Repeated partnership games with imperfect monitoring and no discounting // Review of economic studies. 1986. Vol. 53. N 1. P. 43 – 58.
133. Radner R., Myerson R., Maskin E. An example of a repeated partnership game with discounting and with uniformly inefficient equilibria // Rev. of Econ. St. 1986. Vol. 53. N 1. P. 59 – 69.
134. Radner R. Repeated principal-agent games with discounting // Econometrica. 1985. V. 53. N 5. P. 1173 – 1198.
135. Rey P., Salanie B. Long-term, short-term and renegotiation: on the value of commitment in contracting // Econometrica. 1990. Vol. 58. N 3. P. 597 – 619.
136. Riordan M., Sappington D. Commitment in procurement contracting // Scand. J. of Econ. 1988. V. 90. N 3. P. 357 – 372.
137. Rogerson W. Repeated moral hazard // Econometrica. 1985. Vol. 53. N 1. P. 69 – 76.
138. Rubinstein A., Yaari M.E. Repeated insurance contracts and moral hazard // J.of Econ. Theory. 1983. V. 30. N 1. P. 74 – 57.
139. Selten R. Reexamination of the perfectness concept for equilibrium points in extensive games // International Journal of Game Theory. 1975. Vol. 4. N 1. P. 22 – 55.

140. Spear S.S., Srivastava S. On repeated moral hazard with discounting // Rev. of Econ. St. 1987. V. 54. N 4. P. 599 – 617.
141. Taylor J. Aggregate dynamics and staggered contracts // Journal of Political Economy. 1980. Vol. 88. N 1. P. 1 – 23.
142. Thomas J., Worral T. Self-enforcing wage contracts // Rev. of Econ. St. 1988. V. 55. N 4. P. 541 – 554.
143. Tirole J. Incomplete contracts: where do we stand // Econometrica. 1999. Vol. 67. № 4. P. 741 – 782.
144. Tirole J. Procurement and renegotiation // Journal of Political Economy. 1986. Vol. 94. N 2. P. 235 – 259.
145. Townsend R. Optimal multiperiod contracts and the gain from enduring relationships under private information // Journal of Political Economy. 1982. Vol. 90. N 6. P. 1166 – 1186.
146. Wang G.H. Bargaining over a menu of wage contracts // Rev. of Econ. Studies. 1998. Vol. 65. № 2. P. 295 – 306.
147. Wen Q. The "Folk Theorem" for Repeated Games with Complete Information // Econometrica. 1994. Vol. 62. N 4. P. 949 – 954.