Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный Исследовательский Университет
Высшая Школа Экономики»

На правах рукописи

Богачев Тихон Владимирович

ЭКСТРЕМАЛЬНЫЕ ЗАДАЧИ В НЕКОТОРЫХ ВЕРОЯТНОСТНЫХ МОДЕЛЯХ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ

РЕЗЮМЕ

диссертации на соискание ученой степени
кандидата математических наук

Научный руководитель
доктор физико-математических
наук, профессор А. В. Колесников

Москва, 2023

Введение

Область исследования

Вопросы оптимальности распределения ресурсов в системе исследуются математиками с относительно давнего времени. Множество примеров в области финансов, физики, информационных технологий и массового обслуживания показывают, что интуитивные подходы к оптимизации могут дать результаты, далекие от реального оптимума. В частности, возникают ситуации, в которых добавление дополнительного ресурса в систему приводит к ухудшению общей производительности. В качестве примера приведем знаменитый парадокс Браеса [1], описывающий теоретическую (в виде графа) конфигурацию дорог, в которой строительство новой соединительной дороги может замедлить среднее движение участника, даже если количество машин остаётся постоянным. И наоборот, закрытие одной дороги в сети Браеса позволит всем автомобилям в среднем проехать свой путь быстрее. Также выделяются и другие ситуации такого рода из области вычислительных систем [2].

В классических вопросах оптимизации рассматривают детерминированные системы, что приводит к задачам оптимального управления (см., например, [3], [4]). В системах, описывающих процессы во времени, такие, как обмен информации в коммуникационной сети, установление связей в социальной сети, движение и обмен энергии частиц, взаимодействие игроков на рынке, большое значение приобретают вероятностные модели процессов. Если берется агрегирован-

ный показатель деятельности системы за некоторый период, то мы получаем задачу, не содержащую время, как параметр. Одной из, возможно, наименее исследованных с точки зрения строгих математических результатов задач такого рода является оптимальное налогообложение. Влияние налогообложения на экономику рассматривалось на практике скорее интуитивно. Более строгие исследования в этой области были начаты специалистами по математической экономике по двум основным направлениям: влияние изменения налога на распределение различных благ в экономике (см. [5]) и вопросы оптимального подоходного налогообложения. По второй тематике упомянем весьма известную статью нобелевского лауреата Д.А. Миррлиса [6], его обзор [7], а также работы [8], [9]. Также в математической экономике выделяют товарное налогообложение (см. [10]), но конкретные математические модели не имеют строгой привязки к конкретному экономическому вопросу и фигурируют в разных смыслах в самых разных областях. Изложим постановку вопроса, исследованного автором в работах [11] (в соавторстве с С. Н. Поповой) и [12]. Нами будут рассмотрена модель подоходного налога и оптимизационная задача в этой модели, мотивированная работами [13] и [14]. Эта задача состоит в максимизации интегрального функционала на пространстве возрастающих функций при наличии нелинейного ограничения, приводящего к довольно сингулярным объектам. Поэтому, в отличие от немалого числа работ из прикладной экономики с применением эвристических методов, позволяющих дифференцировать все нужные функции и считать, что в точках экстремума они имеют нулевые производные, строгий математический анализ задачи приводит к интересным вопросам теории функций.

Мы будем считать, что экономический субъект есть абстрактный объект некого класса, полностью характеризуемого типом производительности $\theta \in \Theta \subset \mathbb{R}^n_+$. Субъекты распределены по типам согласно вероятностной мере P на Θ (вооб-

ще, мера должна быть вероятностной с точностью до нормирования, но для удобства мы будем далее считать ее просто вероятностной, это не повлияет на анализ). Обозначим через $l \in \mathbb{R}^n_+$ трудозатраты данного субъекта. В результате скалярного произведения мы имеем численный доход $y = (\theta, l)$ и полезность $U(\theta, l) = y - T(y) - f(l)$, где $T(y)$ есть налог, собираемый с дохода y, определенный регулятором, а $f(l)$ есть функция от трудозатрат, описывающая возникающие при деятельности субъекта финансовые издержки. Полезность может пониматься как чистая прибыль субъекта после понесения издержек. Также уточняется, что $f(l)$ — дважды непрерывно дифференцируемая, возрастающая, строго выпуклая функция. Исходя из соображений здравого смысла, $T(y)$ и $y - T(y)$ являются возрастающими неотрицательными непрерывными функциями. Обычно мы также предполагаем, что функция T — выпуклая, что имеет смысл, исходя из того, что налог предполагается растущим быстрее для больших доходов. Для данного типа производительности θ и фиксированной налоговой функции T мы решаем оптимизационную задачу

$$\max_{l_i>0\ \forall i} U(\theta, l) = \max_{l\in\mathbb{R}^n_+} ((\theta, l) - T(\theta, l) - f(l)). \tag{1}$$

Взяв точки максимума $l_{max}(\theta)$ для всех θ и, соответственно, $y_{max}(\theta) = (\theta, l_{max}(\theta))$, мы определяем правительственный доход как

$$R(T) = \int_\Theta T(y_{max}(\theta))P(d\theta). \tag{2}$$

По аналогии полная полезность в экономике может быть определена как

$$\int_\Theta U(\theta, l_{max}(\theta))P(d\theta).$$

Помимо представления системы с ограниченным ресурсом в виде вероятностного распределения элементов этой системы, выделяют такой мощный инструмент, как сетевое представление. А именно рассматривается сеть (граф) из вершин и связей (ребер) между ними, где у каждой вершины и у каждой связи есть

свои качества. В рамках сети происходят процессы возникновения и обработки некоторых условных запросов (*задач*), которые моделируются стохастическими процессами. Эти процессы требуют от сети распределения ресурса для оптимизации своей работы. В природе можно наблюдать примеры таких систем, которые распределенным образом сами оптимизируют свою деятельность, исходя из естественных законов. Поведение жидкости или газа можно приблизить до микроскопического уровня, рассматривая процесс с точки зрения механики каждой молекулы, причем ребра виртуального графа будут описывать физическое взаимодействие между частицами. На этом уровне детализации норма скорости молекулы выглядит как случайный процесс о стационарным распределением, найденным Максвеллом и позже обсуждавшимся Эрлангом ([15]). Переходя к большему масштабу, мы получим агрегированное описание, которое оперирует такими величинами, как температура и давление. Точно так же поведение электронов в энергосети можно описать в терминах случайных блужданий, и этот простой способ моделирования приводит к очень сложному поведению на макроскопическом уровне: структура потенциалов в сети резисторов такова, что она минимизирует рассеивание тепла при данном уровне протекающего тока ([16]). Локальное, случайное поведение электронов заставляет сеть в целом решать довольно сложную задачу оптимизации.

При рассмотрении коммуникационных сетей можно представить ситуацию, в которой «умная» система контроля перенаправляет соединения, устанавливаемые по заблокированным линиям. Это, в свою очередь, приводит к перенаправлению следующих соединений. В итоге запускается цепная реакция, приводящая к фатальному исходу для производительности всей сети. Таким образом, когда система слишком стремится к эффективности, она может перестараться. Эти примеры показывают, сколь важны алгоритмы маршрутизации ([17]). Также упомянем вопросы случайного роста сети и другие концепции из области

случайных графов (обзор можно найти в [18] или [19]).

Помимо упомянутых подходов к чрезвычайно широкой проблеме оптимизации устройства сетей с нагрузкой, выделяют оптимизацию процессов, происходящих в данной сети и подчиняющихся определенной вероятностной модели (систематический обзор можно найти в [20]). Мы дадим краткое введение в эту тематику для постановки вопроса, исследуемого в главе 3 данной работы. На микроуровне анализ системы базового обслуживания представляет собой теорию очередей ([21]). Для начала рассматривается единственная очередь с входящим потоком задач и процессором данной мощности, который обрабатывает эти задачи. Итоговый процесс будет иметь марковские свойства. При переходе к сети из очередей возникают концепции пуассоновских потоков в сети. Впервые эти свойства были открыты для телефонных сетей Эрлангом ([22]). Также рассматривают концепцию сети потерь (loss networks), суть которой состоит в том, что соединение между двумя вершинами требует одновременного удерживания линии на каждом промежуточном ребре в маршруте между вершинами (более подробно об этом см. [23]). Основное отличие сети потерь от сети очередей состоит в том, что в первой перегрузка сети влечет потерю соединения, в то время как во второй перегрузка ведет к увеличению задержки. И в том, и в другом типе сети можно формулировать вопросы в разных масштабах размера сети и времени. При переходе от единичного узла к сети в целом мы получаем потоки из задач, использующие ресурсы на всем пути своего следования в сети. Тогда возникают вопросы, связанные с понятиями справедливости по отношению к обслуживанию разных потоков (см. [24], [25]). Помимо прочего, в [25] исследованы жидкостные модели планирования, которые во многом стали источником постановки задач в этой диссертации. Если же мы переходим к большему временному масштабу, то уже потоки задач рассматриваются как объекты, которые появляются в системе и заканчиваются. На таком уровне

вся сеть в совокупности рассматривается как система совместного использования процессора (processor-sharing system). Между политиками планировщика на разных масштабах рассмотрения существует связь, она частично изучена, например, в [26].

В настоящей диссертации исследуется вопрос, находящийся на стыке разных масштабов. Модель, которую мы сейчас представим, исследована в работе [27] автора диссертации. В этой модели может быть рассмотрен один узел сети, либо сеть целиком, обслуживающая параллельно несколько потоков задач. Рассмотрим систему из N очередей и одного процессора мощности 1 (см. рис. 1). Есть ограничение M на максимальный размер одной очереди, другими словами, размер буфера, выделенного для одной очереди. При превышении размера буфера задача теряется. В каждый момент времени процессор определяет, каким образом распределить свою вычислительную мощность между очередями. Процессор предполагается с большим количеством логических потоков, так что может выполнять несколько задач, то есть обслуживать несколько очередей, одновременно. Но в рамках одной очереди задачи выполняются строго последовательно. Очереди формируются за счет поступающих в них извне системы задач. Поступления задач из конкретного потока j (поток задач, приходящих в очередь j) представляют собой непрерывный случайный процесс $\mathcal{A}_j(t)$. Предположим, что все задачи имеют сравнительно маленький размер, причем они прибывают в больших количествах. Тогда мы сможем рассматривать каждый поток задач как поток жидкости, льющейся со скоростью $AI_j(t)$, причем изменения скорости соответствуют конкретной траектории процесса $\mathcal{A}_j(t)$.

Конкретный вектор распределения вычислительного ресурса $w = (w_1, \ldots, w_N)$, где $w_1 + \cdots + w_N = 1$, описывает поведение системы в конкретный момент времени, а векторное поле $w(\cdot)$ описывает всю политику обработки. Исследователи обычно рассматривают потоки поступления задач со стационарными свойства-

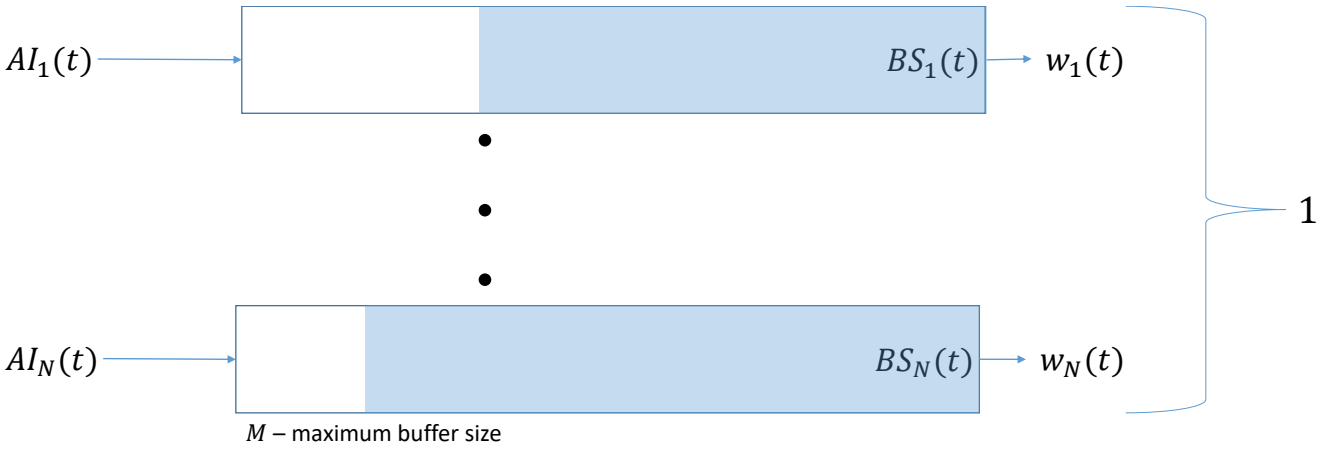

Рис. 1: Модель планировщика задач

ми, например, распределенные по закону Пуассона или более сложному закону Пуассона с марковскими интенсивностями (так называемый Markov-modulated Poisson process), и стараются обеспечить политику планировщика, которая «хороша» для таких потоков. В этом исследовании мы делаем *локальные* предположения для предстоящего периода времени длительностью T_{upd}, т.е. T_{upd} есть частота обновления политики планирования. Причина такого подхода кроется в долгосрочной устойчивости. Действительно, действуя описанным образом, мы не будем в такой степени зависимы от того, действительно ли прибытия задач хорошо оцениваются стационарными процессами в долгосрочной перспективе. Более того, в последние годы исследователи научились обнаруживать моменты марковских «скачков» с довольно неплохой точностью ([28]). Тогда мы можем считать, что в ближайшее время после момента принятия решения прибытия задач в очередь с номером i будут близки к пуассоновскому процессу с интенсивностью a_i. Более конкретно, будем считать, что каждый $\mathcal{A}_i(t)$ есть гауссовский случайный процесс, заданный следующим образом:

1. $\mathcal{A}_i(0) = 0$.
2. приращения процесса $\mathcal{A}_i(t)$ независимы.
3. при $t_1 < t_2$ верно, что $\mathcal{A}_i(t_2) - \mathcal{A}_i(t_1)$ является гауссовской случайной ве-

личиной с ожиданием $a_i(t_2 - t_1)$ и такой же дисперсией.

Это имеет смысл с точки зрения реальных приложений, где поступающие задания в течение предстоящего периода времени могут быть смоделированы пуассоновским процессом интенсивности a_i, но общая мощность и значения интенсивностей большие. Тогда дискретный процесс прибытия приближается гладким гауссовским, а с точностью до масштабирования мощность равна 1.

Обозначим через b_i начальный размер очереди $\mathcal{Q}_i(0)$. Тогда размер очереди в момент t описывается случайной величиной

$$\xi_i(t) = \mathcal{A}_i(t) + b_i - w_i t. \tag{3}$$

Если мы обрежем значения величины $\xi_i(t)$ внутрь отрезка $[0, M]$, то есть положим 0, где она меньше нуля, и M, где $\xi_i(t) > M$, то получим случайную величину $\mathcal{Q}_i(t)$, которая есть фактический размер очереди.

Для события $\omega \in \Omega$ и момента времени $u \in [0, T_{upd}]$ определим множество $\tau_i(\omega, u) = \{t \in [0, u] : \ \xi_i(\omega, t) \geq M\}$. Данные из очереди будут теряться в моменты времени $t \in \tau_i(\omega, u)$. Объем данных, которые не помещались в очередь за всю совокупность промежутков $\tau_i(\omega, u)$ и были удалены, оценивается штрафной функцией

$$L_i(\omega, u) = \int_{\tau_i(\omega,u)} (\xi_i(\omega, t) - M)dt. \tag{4}$$

По аналогии зададим множество $\beta_i(\omega, u) = \{t \in [0, u] : \ \xi_i(\omega, t) \leq 0\}$ и функцию бонуса

$$B_i(\omega, u) = \int_{\beta(\omega)} (-\xi_t(\omega))dt. \tag{5}$$

Полученные случайные процессы L_i и B_i с временем на $[0, T_{upd}]$ описывают объем потерь и суммарный бонус за простой до момента u.

Математическое ожидание базовой величины $\xi_i(t)$ есть

$$s_i(t) = \mathbb{E}\xi_i(t) = b_i + a_i t - w_i t. \tag{6}$$

Определение 0.0.1. *Прогнозом размера очереди* i в момент t будем называть значение $q_i(t)$, заданное ограничением значений $s_i(t)$ диапазоном $[0, M]$.

Обратим внимание, что это значение не эквивалентно расчету математического ожидания истинного размера очереди $\mathcal{Q}_i(t)$.

Ожидаемая задержка $D_i(t)$ - это время, необходимое для обработки прогнозируемой очереди $q_i(t)$ при выделенной пропускной способности. Она строится на основе базовой функции

$$d_i(t) = \frac{a_i - w_i}{w_i} t + \frac{b_i}{w_i} \tag{7}$$

после ограничения ее значений внутри границ $[0, M/w_i]$. Заметим, что если $\mathcal{Q}_i(t) = 0$, то $D_i(t) = 0$ независимо от w_i.

Замечание. Обратим внимание, что этот способ оценивания задержки не эквивалентен расчету математического ожидания процесса задержки $\mathcal{Q}_i(t)/w_i$.

Для каждого потока определяется *локальная средняя задержка*:

$$\mathcal{D}(w_i, a_i, b_i) = \frac{\int_0^{T_{upd}} D_i(t)\ dt}{T_{upd}}. \tag{8}$$

В данном подходе значения (a_i, b_i) считаются известными, вопрос лучшего определения значения a_i оставляется за скобками. Переменной управления является вектор w с ограничением $w \in \mathcal{W}$, где $\mathcal{W}$ — множество точек $w = (w_1, \ldots, w_N)$ таких, что $w_i \geq 0$ и $w_1 + \cdots + w_N = 1$, т.е. стандартный N-симплекс.

Введем различные метрики качества работы всей системы, порождающие соответствующие оптимизационные задачи.

Проблема (Минимизации суммы средних)**.**
Задача минимизации среднего арифметического средних задержек каждого потока (или, что то же самое, их суммы, поскольку суммирование идет по фик-

сированному набору) для данного состояния системы:

$$\min_{w \in \mathcal{W}} \left(\sum_{1}^{N} \mathcal{D}(w_i, a_i, b_i). \right) \tag{9}$$

Обратим внимание, что эта задача отличается от задачи минимизации общей средней задержки по всей системе.

Проблема (Минимакс средних задержек)**.**

Задача минимизации максимума из всех средних задержек:

$$\min_{w \in \mathcal{W}} \Big[\max_{i} \left(\mathcal{D}(w_i, a_i, b_i) \right) \Big]. \tag{10}$$

Далее, мы сужаем требования к вектору распределения ресурса и исходному состоянию всей системы, исходя из некоторых концептуальных соображений из области практических требований. А именно: вводится понятие *устойчивости* состояния, которое имеет тот смысл, что для этого состояния существует хотя бы одно распределение ресурса, при котором ни одна из N очередей качественно не изменит свой статус в течение ближайшего периода времени. Такие распределения мы назовем *равномерными*. Классификация статусов очередей и системы в совокупности является набором простых, но громоздких соотношений на a_i, b_i, M, T_{upd}, так что здесь мы их приводить не будем. Укажем лишь, что в результате эти концептуальные соображения в данной модели приобретают конкретный вид ограничений в виде неравенств. Так что в итоге вектор распределения ресурса, помимо нахождения на стандартном симплексе $\mathcal{W}$, должен быть внутри параллелепипеда, который обозначим через $\mathcal{R}$. Итоговая область ограничений есть $(N-1)$-мерный многогранник $\mathcal{P}$ внутри симплекса.

Цель исследования и результаты

Целью исследования является оптимизация политики работы сложной системы в различных моделях, представленных выше, либо нахождение аналитических свойств оптимальных решений. Приведем основные результаты работы.

1 (Налогообложение, гладкое одномерное распределение типов). В описанной модели налогообложения, связанной с вероятностным распределением экономических субъектов по типам θ, рассмотрим следующие уточнения. Во-первых, будем рассматривать θ, l в одномерном виде, т.е. в виде чисел вместо векторов. В главе 2 показано, что это не ограничивает общность в данной модели, а именно: рассмотрение векторных параметров сводится к рассмотрению норм этих векторов. Проведем замену и рассмотрим задачу в переменных y, θ, а не θ, l, что переводит $f(l)$ в $f(y, \theta)$. На перечисленные функции наложим следующие условия:

- $f(y, \theta)$ - неотрицательная непрерывная функция, где $y \geq 0$, $\theta > 0$ (или θ лежит в некотором отрезке $[\theta_{min}, \theta_{max}] \subset (0, +\infty)$),
- $f(0, \theta) = 0$ при каждом θ, функция $f(y, \theta)$ возрастает по y,
- производная $\partial_\theta f(y, \theta)$ существует и непрерывна по θ и убывает по y, существует производная $\partial_y f(y, \theta) > 0$.
- $T\colon [0, +\infty) \to [0, +\infty)$ - непрерывная функция, $T(0) = 0$ и функция $y \mapsto y - T(y)$ возрастает. Такие функции образуют класс $\mathcal{T}$.

Класс $\mathcal{T}$ можно описать как множество всех 1-липшицевых функций, которые возрастают и равны нулю в нуле. Пусть $y_T(\theta)$ — минимальная точка, в которой достигается максимум полезности (1). Такая точка существует в силу условий на f и T. Положим $y_T(0) = 0$.

Пусть борелевская вероятностная мера распределения субъектов задается плотностью: $P = p\,dx$ на $(0, +\infty)$. Будем предполагать, что либо плотность p положительна на $(0, +\infty)$, либо мера P сосредоточена на отрезке и плотность p положительна на этом отрезке. Положим

$$F(\theta) = P([\theta, +\infty)) = \int_\theta^{+\infty} p(t)\,dt.$$

Будем предполагать, что имеется P-интегрируемая локально ограниченная функция $\gamma > 0$, для которой $y - f(y, \theta) \leq 0$ при $y \geq \gamma(\theta)$.

Мы исследуем задачу нахождения величины

$$J(f, P) = \sup_{T \in \mathcal{T}} R(T), \tag{11}$$

т.е. максимизируется собранный налог. Заметим, что из сказанного выше следует, что $J(f, P) \leq \|R\|_{L^1(P)}$.

Обозначим через $\mathcal{Y}$ множество всех возрастающих непрерывных слева функций $y\colon [0, +\infty) \to [0, +\infty)$, для которых $y(0) = 0$ и при всех $\theta > 0$ выполнены неравенства

$$y(\theta) \leq \gamma(\theta), \quad \partial_y f(y(\theta), \theta) \leq 1, \quad f(y(\theta), \theta) \leq y(\theta).$$

Для всякой функции $y \in \mathcal{Y}$ введем правую обратную функцию формулой

$$v(s) = \sup\{t\colon y(t) \leq s\}.$$

Функция v также возрастает и непрерывна слева. Кроме того, $y(v(s)) = s$ в точках из множества значений y, ибо для такой точки $v(s)$ — максимальная точка в $y^{-1}(s)$.

Основной результат этой части работы состоит в следующем.

Теорема 0.0.2. *Справедливо равенство*

$$J(f, P) = \sup_{y \in \mathcal{Y}} \int \left(y(\theta) - f(y(\theta), \theta) + \int_0^\theta \partial_\theta f(y(\tau), \tau)\, d\tau \right) P(d\theta) =$$
$$= \sup_{y \in \mathcal{Y}} \int [y(\theta) - f(y(\theta), \theta)]\, P(d\theta) + \int \partial_\theta f(y(\theta), \theta) F(\theta)\, d\theta, \tag{12}$$

причем супремум можно брать по возрастающим бесконечно дифференцируемым функциям y с $y' > 0$. Приближения к супремуму по T можно получать с помощью отображений вида

$$T(s) = s - f(s, v(s)) + \int_0^{v(s)} \partial_\theta f(y(\tau), \tau)\, d\tau,$$

где $y \in \mathcal{Y}$ — бесконечно дифференцируемая функция с обратной функцией v.

С помощью теоремы 0.0.2 получено описание решения оптимизационной задачи для случая

$$f(y, \theta) = \frac{y^2}{2\theta^2}. \tag{13}$$

А именно:

$$J(f, P) = \frac{1}{2} \int \frac{p^2(\theta)\theta^3}{p(\theta)\theta + 2F(\theta)} d\theta. \tag{14}$$

2 (Кусочно-линейные налоги). Рассмотрим кусочно-линейную налоговую функцию, представленную N линейными частями. Соответствующие сегменты описываются точками разбиения $m_1 \leq \ldots \leq m_{N-1}$ и коэффициентами $k_1 \leq \ldots \leq k_N$, так что

$$T(y) = \begin{cases} k_1 y, \quad \text{если } y \leq m_1 \\ k_1 m_1 + k_2(y - m_1), \quad \text{если } m_1 \leq y \leq m_2 \\ \vdots \\ k_1 m_1 + k_2(m_2 - m_1) + \cdots + k_N(y - m_{N-1}) \text{ при } m_{N-1} \leq y. \end{cases} \tag{15}$$

Тогда верны следующие утверждения.

Теорема 0.0.3 (О виде оптимального дохода). *В случае кусочно-линейного налога оптимальный доход $y_{max}(\theta)$ является кусочно-постоянной функцией типа, т.е. полупрямая $[0, +\infty)$ делится на последовательные промежутки $I_1, \ldots, I_{2N+1}$, причем y_{max} постоянна на промежутках с четными номерами и квадратично возрастает на тех, у которых номера нечетные.*

Теорема 0.0.4 (О виде максимальной полезности). *В случае кусочно-линейного налога максимальная полезность $U_{max}(\theta)$ является непрерывной строго возрастающей функцией типа. Более того, $[0, +\infty)$ делится на последовательные промежутки $I_1, \ldots, I_{2N+1}$, причем функция U_{max} выпукла на промежутках с четными номерами и вогнута на тех, у которых номера нечетные.*

Заметим, что постановка вопроса перекликается, но не совпадает с постановками вопросов, исследованных в [29].

3. При исследовании модели распределения вычислительного ресурса в системе очередей с совместным использованием ресурса обосновано рассмотрение в качестве элемента метрики качества функции прогноза размера очереди. Конкретнее, доказана

Лемма 0.0.5. *Верно, что*

$$\mathbb{E}\xi_i(t) = \mathbb{E}\mathcal{Q}_i(t) - \Delta_{i,1}(t) + \Delta_{i,2}(t), \tag{16}$$

где $\Delta_{i,1}$ *является плотностью по времени для вычисления бонуса за простой, а* $\Delta_{i,2}$ *является плотностью потерь данных по времени. Конкретно, для каждого* $u \in [0, T_{upd}]$ *верно*

$$\begin{aligned} \mathbb{E}B_i(u) &= \int_0^u \Delta_{i,1}(t)dt \\ \mathbb{E}L_i(u) &= \int_0^u \Delta_{i,2}(t)dt. \end{aligned} \tag{17}$$

4. При исследовании прогнозов в модели системы очередей с совместным использованием ресурса получены следующие результаты.

Теорема 0.0.6 (Об оптимизации равномерной устойчивости). *Для данного устойчивого состояния* (a, b) *задача минимизации суммы средних задержек всех очередей в системе по равномерным распределениям ресурса эквивалентна задаче*

$$\min_{\substack{w_1+\cdots+w_N=1 \\ w_i^{*+} \le w_i \le w_i'}} \left(\frac{c_1}{w_1} + \cdots + \frac{c_N}{w_N} \right), \tag{18}$$

где

$$w_i^* = a_i - \frac{M - b_i}{T_{upd}}, \qquad w_i' = a_i + \frac{b_i}{T_{upd}}, \qquad c_i = \frac{a_i T_{upd}}{2} + b_i$$

и $x^+ = \max(0, x)$ *для действительных* x. *Приведенный ниже алгоритм находит точное решение данной оптимизационной задачи за конечное количество итераций.*

Алгоритм. 1. Возьмем точку v с координатами $v_i = \frac{\sqrt{c_i}}{\sum_i \sqrt{c_i}}$. Если v удовлетворяет ограничениям задачи, то v является искомым решением. В противном случае мы переходим к следующему шагу.

2. Если $N = 2$, то решение выбирается непосредственно из двух концов отрезка, соответствующего ограничениям задачи.

3. Пусть $v_{i_1}, \ldots, v_{i_l}$ — те компоненты, которые нарушают ограничения задачи, в то время как все остальные компоненты условиям задачи подчиняются. Мы будем искать решение задачи на одной из $F_1, \ldots, F_l$ – граней параллелепипеда $\mathcal{R}$, соответствующих нарушенным ограничениям. Каждой такой грани соответствует гиперплоскость, полученная при фиксации одной из переменных w_i на w_i^{*+} или w_i'. Мы выполняем следующие шаги для каждой указанной грани.

   (a) Без потери общности предположим, что мы зафиксировали компоненту w_N. Если $w_N = 0,\ 1$, то значение целевой функции на этой грани бесконечно, то есть на ней оптимум искать бессмысленно.

   (b) В противном случае мы имеем следующую задачу:

$$\min_{(w_1,\ldots,w_{N-1})\in\mathcal{P}^N_{w_N}} \left(\frac{c_1}{w_1} + \cdots + \frac{c_{N-1}}{w_{N-1}}\right), \tag{19}$$

   где $\mathcal{P}^N_{w_N} \subset \mathbb{R}^{N-1}$ - множество точек $x = (x_1, \ldots, x_{N-1})$ таких, что $(x_1, \ldots, x_{N-1}, w_N) \in \mathcal{P}$. После задания новых переменных $y_i = \frac{x_i}{1-w_N}$ имеем $y_1 + \cdots + y_{N-1} = 1$. Задача в переменных $(y_1, \ldots, y_{N-1})$ аналогична начальной. Выполним алгоритм с самого начала для новой задачи в меньшей размерности.

4. После рекурсивного выполнения предыдущего шага на гранях $F_1, \ldots, F_l$ у нас есть минимум для каждой из них. Сравним значения оптимизируемой функции в этих точках и выберем минимум.

□

Методы исследования

В работе используются методы теории вероятностей, теории случайных процессов, действительного и функционального анализа, теории меры, выпуклой оптимизации.

Научная новизна и применимость результатов

Все перечисленные результаты исследования являются новыми. Модель, исследованная в главах 1 и 2, широко используется в математической экономике для изучения подоходного налогообложения. Полученные для гладкого варианта этой модели результаты позволяют в некоторых случаях получить явное решение задачи максимизации суммарного налога. Модель совместного использования ресурса применяется для оптимизации процессов массового обслуживания. В качестве примеров приведем обработку задач процессором, распределение и обслуживание входящих пакетов на сетевом роутере, управление соединениями в масштабах сети и т.д.

Апробация результатов

Результаты диссертации излагались на следующих конференциях и семинарах:

1. Конференция «Новые вызовы в современной теории вероятностей в пространствах высокой размерности и ее применениях в машинном обучении», 12-16 мая 2021 г., Университет Сириус, Сочи, Россия. Доклад “Mathematical Problems of Optimal Scheduling”.

2. Научно-исследовательский семинар «Стохастический анализ и его применения в экономике» под руководством профессора Колесникова А.В. и про-

фессора Конакова В.Д., Высшая Школа Экономики. Москва, 2022. Доклад «Оптимизация налогообложения и вероятностные модели».

3. Совместный семинар ИППИ и Российских исследовательских институтов компании Huawei. Серия докладов по направлению "Scheduling Optimisation" в 2020 и 2021 гг.

Публикации

Результаты диссертации опубликованы в работах [11], [12], [27]. Первые две представлены в журналах, индексируемых в системах цитирования Scopus и Web of Science.

1. Работа [11] опубликована в соавторстве с С.Н. Поповой в журнале «Математические Заметки» (Scopus Q2).

2. Работа [12] опубликована без соавторов в журнале «Известия Иркутского государственного университета. Серия Математика» (Scopus Q2).

3. Работа [27] подготовлена без соавторов и представляет собой препринт, опубликованный в системе *arXiv*.

Структура и объем работы

Диссертация изложена на 75 страницах и состоит из оглавления, введения, трех глав, излагающих результаты работы и их обоснование, заключения и списка литературы, содержащего 31 наименование.

Содержание работы

В Главе 1 исследован одномерный вариант задачи максимизации агрегированного налога с условиями гладкости, которые мы здесь дублировать не будем. В изначальной постановке задача состоит из двух ступеней оптимизации – снача-

ла каждый субъект оптимизирует свою чистую прибыль, затем максимизируется суммарный налог. В заданных же условиях задача сводится к супремуму по специально построенному классу функций. Это упрощает получение ответа, потому что такой способ не содержит промежуточного максимума. В частном случае, когда функция издержек есть

$$f(y,\theta) = \frac{y^2}{2\theta^2}, \tag{20}$$

получаем, что

$$J(f,P) = \frac{1}{2}\int \frac{p^2(\theta)\theta^3}{p(\theta)\theta + 2F(\theta)}d\theta. \tag{21}$$

В Главе 2 показано для начала, что многомерная задача элементарным образом сводится к одномерной. В основном же глава 2 посвящена поведению субъектов модели при кусочно-линейном налогообложении. Получен явный вид оптимальных трудовых затрат субъекта в зависимости от его типа производительности и параметров налоговой функции.

Глава 3 посвящена динамическому процессу в системе с ограниченным ресурсом, что сразу отличает эту постановку задачи от агрегированной статистики в первой модели. Рассматривается система из N очередей, где каждая очередь образуется в результате поступления задач в виде гауссовского процесса заданной интенсивности. Единый вычислительный центр обрабатывает все очереди, распределяя свой ресурс между очередями в каждый момент времени. То есть, способ распределения ресурса есть управление в задаче оптимизации. Обоснован выбор конкретной метрики производительности системы. Изначальные параметры системы классифицированы и для каждого из этих классов задача оптимизации работы системы представлена в виде минимизации явной функции. Для одного из классов состояния системы получен алгоритм, дающий точное решение задачи оптимизации. Корректность алгоритма доказана аналитически.

Литература

[1] Braess D. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 1968, vol. 12, pp. 258–268. English translation: On a paradox of traffic planning. *Transportation Science*, 2005, vol. 39, no. 4, pp. 446–450. https://doi.org/10.1287/trsc.1050.0127

[2] Kameda H., Altman E., Pourtallier O., Li J., Hosokawa Y. Braess-like paradoxes in distributed computer systems. *IEEE Transactions on Automatic Control*, 2000, vol. 45, no. 9, pp. 1687–1691. https://doi.org/10.1109/9.880619

[3] Алексеев В.М., Тихомиров В.М., Фомин С.В. Оптимальное управление. *Наука, М.*, 1979; 432 с.

[4] Асеев С.М., Вельов В.М. Другой взгляд на принцип максимума для задач оптимального управления с бесконечным горизонтом в экономике. *Успехи математических наук*, 2019. Т. 74, No. 6. С. 3–54.

[5] Harberger A. The Incidence of the Corporation Income Tax. *Journal of Political Economy*, 1962, vol. 70, pp. 215–240.

[6] Mirrlees J.A. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 1971, vol. 38, no. 2, pp. 175–208. https://doi.org/10.2307/2296779

[7] Mirrlees J.A. The theory of optimal taxation. *Handbook of Mathematical Economics*, 1986, vol. 3 (ed. by K.J. Arrow and M.D. Intriligator), Chapter 24,

pp. 1197–1249. https://doi.org/10.1016/S1573-4382(86)03006-0

[8] Hellwig M.F. Incentive problems with unidimensional hidden characteristics: a unified approach. *Econometrica*, 2010. V. 78, No. 4, 1201–1237.

[9] Saez E. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 2001. V. 68. P. 205–229.

[10] Atkinson A.B., Stiglitz J.E. The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, 1976, vol. 6, no. 1-2, pp. 55–75. https://doi.org/10.1016/0047-2727(76)90041-4

[11] Bogachev T.V., Popova S.N. On optimization of tax functions. *Mathematical Notes*, 2021, vol. 109, no. 2, pp. 170–179. https://doi.org/10.1134/S000143462101020X

[12] Bogachev T.V. Optimal Behavior of Agents in a Piecewise Linear Taxation Environment. *The Bulletin of Irkutsk State University. Series Mathematics.*, 2022, vol. 42, pp. 17–26. https://doi.org/10.26516/1997-7670.2022.42.17

[13] Steinerberger S., Tsyvinski A. Tax mechanisms and gradient flows. *arXiv*:1904.13276v1.

[14] Sachs D., Tsyvinski A., Werquin N. Nonlinear tax incidence and optimal taxation in general equilibrium. *Econometrica*, 2020, vol. 88, no. 2, pp. 469–493. https://doi.org/10.3982/ECTA14681

[15] Erlang, A. K. A proof of Maxwell's law, the principal proposition in the kinetic theory of gases, 1925. In *Brockmeyer et al, The Life and Works of A. K. Erlang*, pp. 222–226, 1948.

[16] Thomson, W., Tait, P. G. Treatise on Natural Philosophy *Cambridge*, 1879.

[17] Kelly F. P. Network routing. In *Philosophical Transactions of the Royal Society*, Vol 337, Iss. 1647, pages 343–367. 1991.

[18] Райгородский А.М. Модели случайных графов. *МЦНМО Москва*, 2016.

[19] Avrachenkov K., Dreveton M. Statistical Analysis of Networks. *Boston-Delft: now publishers*, 2022. http://dx.doi.org/10.1561/9781638280514

[20] Kelly F. P., Elena Yudovina. Stochastic networks. In *Cambridge University Press*, 2014.

[21] Asmussen S. Applied Probability and Queues, 2nd edn. Springer New York, 2003. In *Stochastic Modelling and Applied Probability*, Vol 51.

[22] Erlang, A. K. Telephone waiting times, 1920. In *Brockmeyer et al, The Life and Works of A. K. Erlang*, pp. 156–171, 1948.

[23] Kelly F. P. Loss networks. *Annals of Applied Probability*, 1991, vol. 1, no. 3, pp. 319–378.

[24] Bonald T., Massoulié L., Proutiere A. and Virtamo J. A queueing analysis of max-min fairness, proportional fairness and balanced fairness. In *Queueing Systems*, Vol 53, pages 65–84. 2006

[25] Bramson, M. Stability and Heavy Traffic Limits for Queueing Networks. *Springer*, 2006.

[26] Shah, D., Wischik, D. Switched networks with maximum weight policies: fluid approximation and multiplicative state space collapse. *Annals of Applied Probability*, 2012, vol. 22, iss. 1, pp. 70–127. http://dx.doi.org/10.1214/11-aap759

[27] Bogachev T. V. Optimization of the fluid model of scheduling: local predictions. *arXiv*:2209.04745v1, 2021.

[28] Avanzi B., Taylor G., Wong B., Xian A. Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework. *arXiv*:2003.13888v2, 2020.

[29] Apps P., Long Ngo Van, Rees R. Optimal piecewise linear income taxation. *Journal of Public Economic Theory*, 2014, vol. 16, no. 4, pp. 523–545. https://doi.org/10.1111/jpet.12070

[30] Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. 4-е изд. *Наука, М.*, 1976; 544 с.

[31] Gladkov N. A., Kolesnikov A. V. and Zimin A. P. The multistochastic Monge–Kantorovich problem. *Journal of Mathematical Analysis and Applications*, 2022; vol. 506, iss. 2, article 125666.