Задание по курсу “Стохастическая и Huge-scale оптимизация”

Гасников А.В. (ПреМоЛаб ФУПМ МФТИ, ИППИ РАН) avgasnikov@gmail.com

Программа курса доступна по ссылке

http://www.mathnet.ru/php/conference.phtml?option_lang=rus&eventID=25&confid=394

Видео-материалы

http://www.youtube.com/user/PreMoLab

Для успешной сдачи задания достаточно набрать 10 баллов

Полноценное решение задачи – это текст, с описание алгоритма, строгим его обоснованием (с числовыми константами для числа итераций и в категориях O() для общего числа арифметических операций) и результаты численных экспериментов, где это подразумевается задачей.

1. Метод зеркального спуска А.С. Немировского и различные вариации

Задача 1 (1 балл). Опишите метод зеркального спуска (МЗС) (см. главу 5 [36]) и метод двойственных усреднений (МДУ) (см. п. 2.1 диссертации [38]). В чем сходства и отличия методов?

Указание. По поводу сходства см. Приложение А3 [3]. Для обнаружения отличий – рассмотрите случай адаптивной постановки задачи (когда заранее не фиксирована точность/[число итераций], т.е. эти параметры не могут входить в размер шага исследуемого метода).

Задача 2 (2 балла). Докажите оценку (3) из п. 2 работы [6], характеризующую скорость сходимости МЗС для задач стохастической оптимизации.

Указание. См. п. 7.6.1 диссертации [14] и доказательство теоремы 4 работы [25].

Задача 3 (2 балла). Докажите аналогичную оценку для метода из п. 3 работы [6], характеризующую скорость сходимости МЗС для условных задач стохастической оптимизации.

Задача 4 (2 балла). Докажите оценку (3) из работы [20]. Получите оценки примеров 1, 2 этой же работы.

Задача 5 (2 балла). Докажите оценки (2), (3) работы [28].

Задача 6 (2 балла). Распространите результаты п. 2 работы [8] на случай не евклидовой прокс-структуры для задач оптимизации на множествах простой структуры. В частности, покажите, что МЗС и проксимальный градиентный метод (ППГМ – Algorithm 3 на стр. 36 диссертации [3]) генерируют последовательность точек, которые лежат в шаре (в смысле выбранной в прямом пространстве нормы – не обязательно евклидовой) с центром в решении задачи и радиуса, определяемого расстоянием от точки метода старта до решения (с точностью до небольшой универсальной, т.е. не зависящей от задачи, мультипликативной константы). Можно ли установить аналогичный результат для обычного прямого градиентного метода ПГМ (см. Приложение B1 [3])? Проверьте в численных экспериментах поведение числовой последовательности, генерируемой методом ПГМ, скажем с 1-нормой в прямом пространстве для плохо обусловленных задач.

Задача 7 (2 балла). Как с помощью регуляризации сделать (без дополнительной логарифмической платы в оценке скорости сходимости по функции) из МЗС для не сильно выпуклой задачи (стохастической) оптимизации сходящийся по аргументу метод?

Указание. Регуляризованную специальным образом задачу можно решать с помощью рестартов МЗС [44].

Задача 8 (2 балла). Как следует действовать, если мы хоти применить к задаче (стохастической) оптимизации МЗС, но нет априорно никакой информации о свойствах задачи?

Замечание. В отличие от п. 3 [25] в данной задаче мы не предполагаем, что есть возможность контролировать зазор двойственности.

2. Быстрый градиентный метод Ю.Е. Нестерова и различные вариации

Задача 1 (3 балла). Для задач безусловной оптимизации (выбрана евклидова прокс-структура) сравните быстрые градиентные методы и способы их получения из работ [37], п. 3.6 [9], [3], [33]. Какие варианты и при каких условиях сходятся быстрее? Постарайтесь визуализировать схему работы методов, рассматривая задачи на плоскости.

Указание. Тестировать рекомендуется на примерах наименее благоприятных функций, соответствующих нижним оценкам (см. пп. 2.1.2, 2.1.4 [37]).

Задача 2 (1 балл). По каким оценкам сходятся БГМ и МЗС, если решение задачи не единственно? В частности, множество решение образует подпространство меньшей размерности, чем исходное пространство, т.е. имеются решения сколь угодно далекие от точки страта, как ее не выбирай.

Задача 3 (4 балла). Исследуйте прямо-двойственность БГМ из предыдущей задачи в контексте возможности их использования в конструкции работы [8] вместо версии БГМ из работы [3]. Исследуйте вопрос равномерной ограниченности, генерируемых методами последовательностей (см. формулу (8) [8]).

Задача 4 (3 балла). Перенесите результаты [3] на задачи композитной (составной) оптимизации (см. п. 2.3 диссертации [38] и замечание 6).

Задача 5 (4 балла). Предложите адаптивный способ (подобный, например, способу из п. 2.3 диссертации [38]) подбора константы Липшица градиента в алгоритме AGM1 работы [3]. Предложите на основе такой модификации AGM1 универсальный быстрый градиентный метод [41].

Указание. Предварительно (для получения универсального метода) рекомендуется распространить результаты главы 4 диссертации [14] на предложенный Вами метод.

Задача 6 (10 баллов). Перенесите основные результаты предыдущих задач этого раздела на задачи стохастической оптимизации.

Указание. Идея: подставлять в методы вместо градиента – стохастический градиент и изменить политику выбора шагов. См. главу 7 диссертации [14], доказательство теоремы 4 работы [25].

Задача 7 (5 баллов). Предложите алгоритм, объединяющий в себе алгоритмы из задач 4, 5. Перенесите предложенный Вами алгоритм на сильно выпуклые задачи с помощью техники рестартов (см. п. 5 [3]). Примените последний алгоритм к задаче композитной оптимизации

$$\frac{1}{2}\|Ax-b\|_2^2+\mu\sum_{k=1}^{n}x_k\ln x_k \to \min_{\sum_{k=1}^{n}x_k=1,\, x\ge 0}.$$

Рассмотрите два случая а) $\mu\ge 0$ – мало (сильную выпуклость композита в 1-норме можно не учитывать); б) μ – достаточно большое (сильную выпуклость композита в 1-норме необходимо учитывать).

Указание. Воспользуйтесь примером 4.1.3.а диссертации [14] и тем, что на каждом шаге метода необходимо решать специальную задачу “проектирования”: считать (прокс-) градиентное отображение в композитном варианте – такая задача в случае б), к сожалению, не решается по явным формулам, поскольку для использования техники рестартов (хотим воспользоваться сильной выпуклостью) нельзя брать прокс-функцию энтропию, как это можно было делать в случае а). То есть на каждом шаге в свою очередь необходимо (приближенно) решать задачу оптимизации (правда уже более простую). Это можно делать, например, с помощью небольшой модификации подхода работы [8]. За счет сильной выпуклости функционала вспомогательной задачи тут имеет место сходимость по аргументу (в прямых переменных) в решении вспомогательной задачи. Это позволяет применять технику упомянутого примера 4.1.3.а для контроля точности решения исходной задачи исходя из точности решения вспомогательных задач.

Задача 8 (2 балла). Обобщите оценки работы [18] в сильно выпуклом случае на не евклидовы прокс-структуры.

Задача 9 (4 балла). Подобно главе 6 диссертации [14] (см. также [16]) предложите “промежуточную” модификацию AGM1 [3]. Предложите промежуточные версии производных (из AGM1) алгоритмов.

Указание. Попробуйте выбрать $\alpha_k \sim k^p/L$, $p \in [0,1]$.

3. Покомпонентные методы и безградиентные методы

Идею того, как переносить результаты предыдущего раздела с полноградиентных методов на покомпонентные методы хорошо поясняет конструкция, описанная в работе [29] (подставлять вместо градиента его покомпонентный рандомизированный вариант и изменить политику выбора шагов). Собственно, задачи тут такие: аккуратно распишите результаты, полученные 1) в п. 2 (1 балл), 2) в п. 3 (3 балла), 3) в замечании 2 (3 балла), 4) в замечании 3 (2 балла), 5) в замечании 4 (4 балла), 6) в замечании 5 (4 балла), 7) в замечании 6 (3 балла) – это наиболее полезная и при этом достаточно автономная задача из данного цикла, 8) в замечании 7 (3 балла). В примерах 2 – 4 стоит аккуратно выписать методы и оценки скорости сходимости, произвести расчеты: 9) пример 2 (2 балла), 10) пример 3 (3 балла), 11) пример 4 (3 балла)). Попробуйте с помощью отмеченной конструкции распространить основные результаты предыдущего раздела на 12) спуски по направлению (7 баллов) и 13) безградиентные методы (9 баллов). В частности, опишите методы, которые работают по оценкам из работы [18].

Задача 14 (5 баллов). Перенесите результаты п. 4 работы [26], посвященного оценкам для безградиентных методов и спусков по направлению (см. также работы [18, 28, 29]) на случай общего (не “независимого”), вообще говоря, враждебного оракула, выдающего значение функции (реализацию), но, по-прежнему, с небольшим уровнем шума.

Задача 15 (5 баллов). Сопоставьте метод двойного сглаживания ((МДС), см. [42, 15, 43]) с методом из п. 3 работы [27]. Очевидно, что МДС работает быстрее. Останется ли это верным, если рассмотреть МДС в условиях шума (не случайной природы)? Определите уровень допустимого шума, при котором МДС будет быстрее метода из п. 3 работы [27].

4. Нижние оценки

Задача 1 (5 баллов). В книге [34] получены нижние оценки в категориях относительной точности, которую можно по-разному выбирать. Верно ли, что всюду в этой книге можно считать, что относительная точность считается в масштабе вариации функции (расшифруйте последнее словосочетание)? Как действовать, если множество, на котором происходит оптимизация не ограничено, но вариация функции ограничена, при этом минимум достигается (следует сравнить с упражнением 6 на стр. 157–158 [34]). Другими словами, что на самом деле входит в нижние/верхние оценки сложности: расстояние от точки старта до решения и константы гладкости или(и) вариация функции, в частности, $f\left(x^0\right) - f_*$?

Задача 2 (5 баллов). Рассматривается задача стохастической выпуклой оптимизации в $\mathbb{R}^n$ (без функциональных ограничений) с локальным оракулом (с конечной дисперсией и несмещенным стохастическим субградиентом) на единичном шаре в p-норме, $2 \le p \le \infty$ с соответствующей (этой норме) константой Липшица = 1 – тогда константа Липшица в 2-

норме не больше чем 1 (все это равномерно по шуму и по пространственной переменной). В статье [1] имеется вот такая нижняя оценка (при сформулированных условиях)

(1) $N_1(\varepsilon) \sim \frac{n^{1-2/p}}{\varepsilon^2}$, $2 \le p \le \infty$.

В книге [34] имеется следующая нижняя оценка при тех же самых условиях

(2) $N_2(\varepsilon) \sim \frac{c_p}{\varepsilon^p}$, $2 \le p < \infty$

Оценка (1) достигается на МЗС с 2-нормой и прокс-функцией $d(x) = \|x\|_2^2 / 2$ – то есть обычный евклидовой прокс-функцией. Причем тут можно еще что-то выгадать за счет того, что константа Липшица считается в 2-норме, что может быть меньше, чем константа, посчитанная в p-норме (последняя по предположению равна 1). Оценка (2) достигается на МЗС из книги [34].

Рассмотрим два диапазона:

а) $n^{-1/p} \le \varepsilon \le 1$

В этом диапазоне $N_1(\varepsilon) \succ N_2(\varepsilon)$ – то есть нижняя оценка 1 получается посильнее, но возникает нехороший момент, поскольку оценка 2 достигается (см. выше), а она лучше (меньше) чем оценка 1. При этом в работе [1] утверждается, что оценка 1 – это нижняя оценка. В этом месте возникает первое возможное противоречие.

б) $0 < \varepsilon \le n^{-1/p}$

В этом диапазоне, наоборот, $N_1(\varepsilon) \prec N_2(\varepsilon)$ – то есть нижняя оценка 2 получается посильнее, но возникает нехороший момент, поскольку оценка 1 достигается (см. выше), а она лучше (меньше) чем оценка 2. Вот этом месте возникает второе возможное противоречие.

Как бы Вы объяснили отмеченные противоречия?

Задача 3 (5 баллов). Рассматривается задача гладкой выпуклой оптимизации в $\mathbb{R}^n$ (без функциональных ограничений) с локальным оракулом на единичном шаре в p-норме, $2 \le p \le \infty$ с соответствующей (этой норме) константой Липшица градиента = 1 – тогда константа Липшица градиента в 2-норме не больше чем 1. В статье [30] имеется вот такая нижняя оценка (при сформулированных условиях)

$$f(x^N) - f_* \ge \frac{C_1}{N^{1+2/p}}.$$

Причем эта оценка достигается [35]. С другой стороны, если брать обычный БГМ с евклидовой прокс-структурой для этих же задач, то верхние оценки будут иметь вид

$$f(x^N) - f_* \le \frac{C_2 n^{1-2/p}}{N^2}.$$

Причем тут можно еще что-то выгадать за счет того, что константа Липшица градиента считается в 2-норме, что может быть меньше чем в p-норме (последняя по предположению равна 1). Тем не менее, можно ли отсюда сделать вывод, что и при $N \le n$ выбранный БГМ не оптимальный метод (при $N \ge n$ БГМ доминируется методом центров тяжестей [34, 9])?

Задача 4 (А.С. Немировский, 7 баллов). Рассматривается задача гладкой выпуклой оптимизации в $\mathbb{R}^n$ (без функциональных ограничений) с локальным оракулом на единичном шаре в 1-норме, $2 \le p \le \infty$ с константой Липшица градиента в 2-норме ограниченной числом 1. Тогда нижняя оценка (при $N \le n$) будет иметь вид

$$f\left(x^N\right) - f_* \ge \frac{\tilde{C}_1}{N^3}.$$

С другой стороны, если брать обычный БГМ с KL прокс-структурой для этой же задачи, то верхняя оценка будут иметь вид

$$f\left(x^N\right) - f_* \le \frac{\tilde{C}_2 L_1}{N^2},$$

где константа Липшица градиента в 1-норме $1/n = L_2/n \le L_1 \le L_2 = 1$. Тем не менее, отсюда нельзя сделать вывод, что нижняя оценка достигается. Почему? Достигается ли эта нижняя оценка и если достигается, то на каком методе? – насколько нам известно, это пока открытая задача, поставленная А.С. Немировским. Постарайтесь продвинуться в ее решении.

Задача 5 (7 баллов). Распространите результаты предыдущих двух задач на случай, когда оракул вместо градиента может выдавать в указанной нами точке только производную по указанному нами направлению. Поставьте и решите задачу, аналогичную задаче А.С. Немировского для такого оракула.

Указание. См. таблицу 4 работы [27] и работу [28].

Задача 6 (7 баллов). Распространите технику глав 4, 5 диссертации [14] по получению нижних оценок в случае наличия шумов неслучайной природы на безградиентные методы. В частности, исследуйте вопрос о неулучшаемости оценок [18, 27, 28].

5. Игра на числе итераций и сложности итерации (развитие идеи композитной оптимизации)

Задача 1 (4 балла). Постарайтесь заменить в конструкции параграфа 2.3 книги [37] (в методе нагруженных функционалов) часть, связанную с пересчетом новой точки, оценивающей оптимальное значение функционала. А именно используйте здесь сочетание метод секущих Ньютона для внешней задачи с БГМ для внутренней задачи (заметим, что рассмотрение естественным образом должно проходить в концепции неточного оракула). Почему оценка теоремы 2.3.6 для негладкой задачи (2.47) (см. [37]) имеет вид, как если бы задача была гладкая? Не противоречит ли это нижним оценкам?

Задача 2 (3 балла). Распространите пример на случай когда вместо евклидовых шаров используются шары в других (вообще говоря, разных и не согласованных) нормах (при этом у шаров допускаются бесконечные радиусы).

Задача 3 (6 баллов). В работах [23, 24] описана конструкция суперпозиции методов с помощью концепции неточного оракула (нечто похожее нам уже встречалось в задаче 7 из раздела 2). Распространите эту конструкцию на случай, когда вместо метода Синхорна (балансировки) для двойственной внутренней задачи используется подход работы [8], т.е. двойственная внутренняя задача решается БГМ. С помощью идей двойственного сглаживания Ю.Е. Нестерова (см. главу 5 диссертации [38]) постарайтесь подобрать параметр регуляризации (коэффициент при энтропии) так, чтобы в итоговой задаче поиска барицентра Вассерштейна [24] находился с контролируемой точностью именно барицентр Вассерштейна.

Задача 4 (10 баллов). Попробуйте распространить приведенные выше результаты на покомпонентные методы, когда один или оба метода (внутренний и внешний) являются покомпонентными или даже безградиентными. Какие сочетания могут быть наиболее интересными? Приведите примеры.

6. Невыпуклые задачи

Задача 1 (3 балла). Известно, что обычный ПГМ для безусловной задачи гладкой глобальной (не выпуклой) оптимизации сходится к одному из экстремумов (не обязательно к локальному минимуму, см. пример 1.2.2 [37]). Причем его сходимость (характеризуемая скоростью стремления квадрата евклидовой нормы градиента к нулю) не улучшаема с точностью до мультипликативной константы. Распространим ли этот результат на гладкие задачи глобальной условной оптимизации (на множествах просто структуры)? Как следует действовать, если функционал не гладкий? Можно ли так модифицировать метод (не сильно потеряв в скорости сходимости), чтобы он гарантированно сходился к локальному минимуму? Как можно увеличить скорость сходимости, если известно, что задача не только с Липшицевым градиентом, но и с Липшицевым гессианом? Насколько улучшится скорость сходимости по норме градиента при ответе на предыдущий вопрос, если дополнительно предположить, что оптимизируемая функция еще и выпуклая?

Указанием. См. [40] (логарифмические множители в оценках этой работы для выпуклой функции можно убрать).

Задача 2 (5 баллов). В приложениях (см., например, [10, 11]) возникает потребность в том, чтобы распространить результаты предыдущей задачи на случай, когда оракул нам может выдавать только зашумленный градиент (допускается шум случайной природы – несмещенный и шум не случайной природы (который дает смещение), но не большого масштаба). Постарайтесь распространить результат работы [11] по градиентному спуску для не выпуклых задач с неточным оракулом на описанный в задаче случай.

Задача 3 (9 баллов). Постарайтесь (подобно п. 7 [39]) перенести результаты предыдущих двух задач на безградиентные методы и спуски по направлению. Примените эти алгоритмы к задаче Learning Supervised PageRank [10, 11].

Задача 4 (7 баллов). Покажите, что если рассматривать класс задач оптимального управления линейных по фазовым переменным, которые сводятся к (выпуклым) Ляпуновским задачам (см. параграф 4.3 книги [2]), то можно применять изложенную выше теорию методов БГМ с неточным оракулом (для этого еще нужно выбрать в качестве гильбертова пространства $L_2(U)$ – пространство квадратично интегрируемых функций в пространстве управлений). Покажите (следуя [31]), что если ограничиваться локальной теорией, то сказанное выше можно перенести и на общие задачи оптимального управления. Концепция неточного оракула позволяет привнести сюда элемент новизны, существенно мотивированный практическими нуждами – принципиальной невозможностью (в типичных случаях нет явных формул) решать с абсолютной (очень хорошей) точностью вспомогательную задачу на каждом шаге градиентного спуска. Например, решение такой вспомогательной задачи для класса задач оптимального управления со свободным правым концом приводит к двум начальным задачам Коши для систем обыкновенных дифференциальных уравнений (важно, чтобы СОДУ для фазовых переменных и сопряженных решались, скажем, методом Эйлера, на одной и той же сетке), которые необходимо решить для вычисления градиента функционала [12]. Однако, в действительности, почти все практически интересные задачи (за редким исключением, к коим можно отнести уже упомянутый класс Ляпуновских задач [2]) в бесконечномерных пространствах не являются выпуклыми, поэтому здесь имеет смысл говорить лишь о локальной теории [32]. Если ограничиться неускоренными методами (например, ПГМ), то можно показать, что при весьма общих условиях эти методы могут быть использованы в гильбертовом пространстве в концепции неточного оракула и для невыпуклых (но гладких) функционалов, причем с аналогичными оценками скорости сходимости (отличие от выпуклого случая будет в том, что метод сходится лишь к локальному экстремуму, в бассейне притяжения которого окажется точка старта). Заметим, что задачи оптимального управления можно численно решать, построив соответствующую (аппроксимирующую) задачу оптимального управления с дискретным временем, что приводит к конечномерным задачам, для решения которых можно использовать конечномерный вариант ПГМ в невыпуклом случае (с точным оракулом). Этот путь, как правило, и предлагается в большинстве источников (см., например, [12]). Однако при таком подходе мы должны уметь (по возможности точно) решать сложную задачу оценки качества аппроксимации исходной задачи оптимального управления ее дискретным по времени вариантом. Более теоретически обоснованный способ рассуждений, по сути, приводящей к необходимости решать все те же конечномерные задачи, заключается в рассмотрении исходной задачи оптимального управления и ее решения бесконечномерным вариантом ПГМ в невыпуклом случае (с неточным оракулом). Неточность оракула существенна. Поскольку на каждой итерации этого градиентного метода необходимо решать две задачи Коши для СОДУ, что в общем случае можно сделать лишь приближенно, но с лучшим контролем точности, чем при первом подходе. Постарайтесь осуществить это план.

7. Прикладные задачи

Задача 1 (5 баллов). Сопоставьте различные методы решения системы линейных уравнений больших размеров. Опишите классические способы. Сравните их с методами из работы [7]. Выводы сделайте на основе теоретического анализа и численных экспериментов.

Задача 2 (6 баллов). Сопоставьте различные методы решения Google problem (поиска вектора PageRank). В частности, сопоставьте методы из работы [5, 6, 17], замечания 10, 11 [8]. Определите, какой из методов работает быстрее, и при каких условиях. Выводы сделайте на основе теоретического анализа и численных экспериментов.

Задача 3 (5 баллов). Сопоставьте различные способы поиска "проекции" точки на аффинное многообразие из работы [4]. Определите, какой из методов работает быстрее, и при каких условиях. Выводы сделайте на основе теоретического анализа и численных экспериментов.

Задача 4 (7 баллов). Сопоставьте различные способы поиска равновесий в транспортных сетях из работы [25]. Определите, какой из методов работает быстрее, и при каких условиях. Выводы сделайте на основе теоретического анализа и численных экспериментов. Предложите на основе п. 3 [25] метод поиска равновесия в модели грузоперевозок РЖД [13].

Указание. В этой и следующих двух задачах можно тестировать алгоритмы на графе [45].

Задача 5 (7 баллов). Сопоставьте различные способы поиска стохастических равновесий в транспортных сетях из работы [22]. Определите, какой из методов работает быстрее, и при каких условиях. Выводы сделайте на основе теоретического анализа и численных экспериментов.

Задача 6 (10 баллов). Реализуйте и сопоставьте два способа поиска (стохастических) равновесий в многостадийных моделях транспортных потоков [23, 24] и [21]. Какой из способов (и при каких условиях) эффективнее?

Задача 7 (7 баллов). Исходя из п. 2 работы из работы [24] (см. также задачу 3 раздела 5) и проведенного там обзора численных методов поиска энтропийно-сглаженного барицентра Вассерштейна, постарайтесь определить наилучший способ поиска "настоящего" барицентра Вассерштейна семейства вероятностных мер.

СПИСОК ЛИТЕРАТУРЫ

1. *Agarwal A., Bartlett P.L., Ravikumar P., Wainwright M.J.* Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization // IEEE Transaction of Information. 2012. V. 58. № 5. P. 3235–3249. arXiv:1009.0571
2. *Алексеев В.М., Тихомиров В.М., Фомин С.В.* Оптимальное управление. М.: Физматлит, 2005.
3. *Allen-Zhu Z., Orecchia L.* Linear coupling: An ultimate unification of gradient and mirror descent // e-print, 2014. arXiv:1407.1537
4. *Anikin A., Dvurechensky P., Gasnikov A., Golov A., Gornov A., Maximov Yu., Mendel M., Spokoiny V.* Modern efficient numerical approaches to regularized regression problems in application to traffic demands matrix calculation from link loads // Proceedings of International conference ITAS-2015. Russia, Sochi, September, 2015. arXiv:1508.00858
5. *Аникин А.С., Гасников А.В., Горнов А.Ю., Камзолов Д.И., Максимов Ю.В., Нестеров Ю.Е.* Эффективные численные методы решения задачи PageRank для дважды разреженных матриц // Труды МФТИ. 2015. Т. 7. № 4. С. 74–94. arXiv:1508.07607
6. *Аникин А.С., Гасников А.В., Горнов А.Ю.* Рандомизация и разреженность в задачах huge-scale оптимизации на примере работы метода зеркального спуска // Труды МФТИ. 2016. Т. 8. arXiv:1602.00594 (подана)
7. *Аникин А.С., Гасников А.В., Горнов А.Ю.* О неускоренных эффективных методах решения разреженных задач квадратичной оптимизации // Труды МФТИ. 2016. Т. 8. arXiv:1602.01124 (подана)
8. *Аникин А.С., Гасников А.В., Двуреченский П.Е., Тюрин А.И., Чернов А.В.* Двойственные подходы к задачам минимизации сильно выпуклых функционалов простой структуры при аффинных ограничениях // ЖВМ и МФ. 2016. Т. 56. arXiv:1602.01686 (подана)
9. *Bubeck S.* Convex optimization: algorithms and complexity // In Foundations and Trends in Machine Learning. 2015. V. 8. no. 3-4. P. 231–357. arXiv:1405.4980
10. *Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., Zhukovskii M.* Learning supervised PageRank with gradient-free optimization methods // e-print, 2014. arXiv:1411.4282
11. *Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., Zhukovskii M.* Learning Supervised PageRank with Gradient-Based and Gradient-Free Optimization Methods // e-print, 2016. arXiv:1603.00717
12. *Васильев Ф.П.* Методы оптимизации. М.: МЦНМО, 2011.
13. *Ващенко М.П., Гасников А.В., Молчанов Е.Г., Поспелова Л.Я., Шананин А.А.* Вычислимые модели и численные методы для анализа тарифной политики железнодорожных грузоперевозок. М.: ВЦ РАН, 2014. arXiv:1501.02205
14. *Devolder O.* Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization. CORE UCL, PhD thesis, March 2013. http://www.ucllouvain.be/cps/ucl/doc/core/documents/coredp2011_70web.pdf
15. *Duchi J.C., Jordan M.I., Wainwright M.J., Wibisono A.* Optimal rates for zero-order convex optimization: the power of two function evaluations // IEEE Transaction of Information. 2015. V. 61. № 5. P. 2788–2806.
16. *Dvurechensky P., Gasnikov A.* Stochastic Intermediate Gradient Method for Convex Problems with Inexact Stochastic Oracle // Journal Optimization Theory and Applications. 2016. arXiv:1411.2876 (submitted)

17. *Гасников А.В., Дмитриев Д.Ю.* Об эффективных рандомизированных алгоритмах поиска вектора PageRank // ЖВМ и МФ. Т. 55. № 3. 2015. С. 355–371. arXiv:1410.3120
18. *Гасников А.В., Двуреченский П.Е., Камзолов Д.И.* Градиентные и прямые методы с неточным оракулом для задач стохастической оптимизации // Динамика систем и процессы управления. Труды Международной конференции, посвящено 90-летию со дня рождения академика Н.Н. Красовского. Екатеринбург, 15 – 20 сентября 2014. Издательство: Институт математики и механики УрО РАН им. Н.Н. Красовского (Екатеринбург), 2015. С. 111–117. arXiv:1502.06259
19. *Гасников А.В.* Об эффективной вычислимости конкурентных равновесий в транспортно-экономических моделях // Математическое моделирование. 2015. Т. 27. № 12. С. 121–136. arXiv:1410.3123
20. *Гасников А.В., Лагуновская А.А., Морозова Л.Э.* О связи имитационной логит динамики в популяционной теории игр и метода зеркального спуска в онлайн оптимизации на примере задачи выбора кратчайшего маршрута // Труды МФТИ. 2015. Т. 7. № 4. С. 104–113. arXiv:1511.02398
21. *Гасников А.В., Гасникова Е.В., Мациевский С.В., Усик И.В.* О связи моделей дискретного выбора с разномасштабными по времени популяционными играми загрузок // Труды МФТИ. 2015. Т. 7. № 4. С. 129–142. arXiv:1511.02390
22. *Гасников А.В., Гасникова Е.В., Двуреченский П.Е., Ершов Е.И., Лагуновская А.А.* Поиск стохастических равновесий в транспортных моделях равновесного распределения потоков // Труды МФТИ. 2015. Т. 7. № 4. С. 114–128. arXiv:1505.07492
23. *Гасников А.В., Двуреченский П.Е., Камзолов Д.И., Нестеров Ю.Е., Спокойный В.Г., Стецюк П.И., Суворикова А.Л., Чернов А.В.* Поиск равновесий в многостадийных транспортных моделях // Труды МФТИ. 2015. Т. 7. № 4. С. 143–155. arXiv:1506.00292; https://mipt.ru/upload/medialibrary/ffe/143-155.pdf
24. *Гасников А.В., Двуреченский П.Е., Спокойный В.Г., Стецюк П.И., Суворикова А.Л.* Суперпозиция метода балансировки и универсального градиентного метода для поиска энтропийно-сглаженного барицентра Вассерштейна и равновесий в многостадийных моделях транспортных потоков // Труды МФТИ. 2016. Т. 8. arXiv:1506.00292 (принята к печати)
25. *Гасников А.В., Двуреченский П.Е., Дорн Ю.В., Максимов Ю.В*. Численные методы поиска равновесного распределения потоков в модели Бэкмана и модели стабильной динамики // Математическое моделирование. 2016. Т. 28. (принята к печати) arXiv:1506.00293
26. *Гасников А.В., Двуреченский П.Е., Нестеров Ю.Е.* Стохастические градиентные методы с неточным оракулом // Труды МФТИ. 2016. Т. 8. arxiv:1411.4218
27. *Гасников А.В., Лагуновская А.А., Усманова И.Н., Федоренко Ф.А.* Безградиентные прокс-методы с неточным оракулом для негладких задач выпуклой стохастической оптимизации на симплексе // Автоматика и телемеханика. 2016. arXiv:1412.3890 (принята к печати)
28. *Гасников А.В., Крымова Е.А., Лагуновская А.А., Усманова И.Н., Федоренко Ф.А.* Стохастическая онлайн оптимизация. Одноточечные и двухточечные нелинейные многорукие бандиты. Выпуклый и сильно выпуклый случаи // Автоматика и телемеханика. 2016. arXiv:1509.01679 (подана)
29. *Гасников А.В., Двуреченский П.Е., Усманова И.Н.* О нетривиальности быстрых (ускоренных) рандомизированных методов // Труды МФТИ. 2016. Т. 8. arXiv:1508.02182 (принята к печати)
30. *Guzman C., Nemirovski A.* On lower complexity bounds for large-scale smooth convex optimization // Journal of Complexity. 2015. arXiv:1307.5001

31. *Halkin H.* Liapounov's theorem of the range of a vector measure and Pontryagin's maximum principle // Arch. Rat. Mech. Anal. 1962. V. 10. P. 296–304.
32. *Левитин Е.С., Поляк Б.Т.* Методы минимизации при наличии ограничений // ЖВМ и МФ. 1966. Т. 6. № 5. С. 787–823.
33. *Lin H., Mairal J., Harchaoui Z.* A universal catalyst for first-order optimization // Advances in Neural Information Processing Systems (NIPS), 2015. https://hal.inria.fr/hal-01160728
34. *Немировский А.С., Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. М.: Наука, 1979. http://www2.isye.gatech.edu/~nemirovs/Lect_EMCO.pdf
35. *Немировский А.С., Нестеров Ю.Е.* Оптимальные методы гладкой выпуклой оптимизации // ЖВМ и МФ. 1985. Т. 25. № 3. С. 356–369.
36. *Nemirovski A.* Lectures on modern convex optimization analysis, algorithms, and engineering applications. Philadelphia: SIAM, 2013. http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf
37. *Нестеров Ю.Е.* Введение в выпуклую оптимизацию. М.: МЦНМО, 2010.
38. *Нестеров Ю.Е.* Алгоритмическая выпуклая оптимизация. Диссертация на соискание степени д.ф.-м.н. по специальности 01.01.07 – вычислительная математика. Долгопрудный, МФТИ 26 декабря 2013 г. 367 стр. http://www.mathnet.ru/php/seminars.phtml?option_lang=rus&presentid=8313
39. *Nesterov Yu.* Random gradient-free minimization of convex functions // CORE Discussion Paper 2011/1. 2011.
40. *Nesterov Yu.* How to make the gradients small // OPTIMA 88. 2012. P. 10–11. http://www.mathopt.org/Optima-Issues/optima88.pdf
41. *Nesterov Yu.* Universal gradient methods for convex optimization problems // CORE Discussion Paper 2013/63. 2013. Math. Program. Ser. A. 2015. V. 152 P. 381–404. https://www.uclouvain.be/cps/ucl/doc/core/documents/coredp2013_26web.pdf
42. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983; М.: УРСС, 2014.
43. *Shamir O.* An Optimal Algorithm for Bandit and Zero-Order Convex Optimization with Two-Point Feedback // e-print, 2015. arXiv:1507.08752
44. *Juditsky A., Nesterov Yu.* Deterministic and stochastic primal-dual subgradient algorithms for uniformly convex minimization // Stoch. System. 2014. V. 4. no. 1. P. 44–80. arXiv:1401.1792
45. http://www.bgu.ac.il/~bargera/tntp/

21 марта 2016 г.