Автоматика и телемеханика, № 2, 2017

Стохастические системы, системы массового обслуживания

© 2017 г. **А.В. ГАСНИКОВ, канд. физ.-мат. наук (gasnikov@yandex.ru)**
(Московский физико-технический институт (государственный университет), Институт проблем передачи информации им. А.А. Харкевича РАН),
Е.А. КРЫМОВА, канд. физ.-мат. наук (ekkrym@gmail.com)
(Институт проблем передачи информации им. А.А. Харкевича РАН),
А.А. ЛАГУНОВСКАЯ (a.lagunovskaya@phystech.edu)
(Институт прикладной математики им. М.В. Келдыша РАН, Московский физико-технический институт (государственный университет)),
И.Н. УСМАНОВА (ilnura94@gmail.com)
(Московский физико-технический институт (государственный университет), Институт проблем передачи информации им. А.А. Харкевича РАН),
Ф.А. ФЕДОРЕНКО (f.a.fedorenko@gmail.com)
(Московский физико-технический институт (государственный университет))

СТОХАСТИЧЕСКАЯ ОНЛАЙН ОПТИМИЗАЦИЯ. ОДНОТОЧЕЧНЫЕ И ДВУХТОЧЕЧНЫЕ НЕЛИНЕЙНЫЕ МНОГОРУКИЕ БАНДИТЫ. ВЫПУКЛЫЙ И СИЛЬНО ВЫПУКЛЫЙ СЛУЧАИ[1]

Предложена безградиентная модификация метода зеркального спуска решения задач выпуклой стохастической онлайн оптимизации. Особенностью постановки является допущение, что реализации значений функции доступны с небольшими шумами. Цель данной работы – установить скорость сходимости предложенных методов и определить, при каком уровне шума факт его наличия не будет существенно сказываться на скорости сходимости.

Ключевые слова: онлайн оптимизация, безградиентные методы, неточный оракул, стохастическая оптимизация

1. Введение

Данная работа представляет собой попытку перенести результаты статьи [1] на онлайн контекст [2–10]. А именно, следуя [1] рассмотрим постановку задачи выпуклой стохастической онлайн оптимизации, в которой на каждом шаге (итерации) вместо градиента можно получать только реализацию значения, соответствующую этому шагу функции. При этом допускается, что эта

[1] Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 15-31-20571 мол_а_вед). Исследования первого и второго авторов, связанные с получением теоремы 1, выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект № 14-50-00150).

реализация доступна с шумом уровня δ, вообще говоря, не случайной природы. Рассматривается две возможности: на одном шаге (при одной реализации) получать зашумленное значение в одной точке и в двух точках. В первом случае говорят, что рассматривается задача о нелинейных многоруких бандитах (иногда добавляя "одноточечных") [7]. Во втором случае говорят о нелинейных многоруких двухточечных бандитах [7]. Принципиальная разница есть именно при таком переходе [3, 7]. Последующее увеличение числа точек не меняет принципиально картину, соответствующую двум точкам [11].

Основная идея заключается в специальном сглаживании исходной постановки задачи и использовании метода зеркального спуска [1, 7, 10, 12, 13]. Оригинальной составляющей, в частности, является предложенное в данной статье обобщение этой конструкции на случай наличия шумов. Обратим внимание на условие 1 в разделе 2 (следует сравнить, например, с [1, 14, 15]). Это условие позволило, с одной стороны, изящно распространить известные оценки на случай, когда есть шумы, см. формулы (2), (3) раздела 2, а с другой стороны, это условие хорошо подходит под специфику рассматриваемой в статье постановки (можем получать только зашумленные реализации значений функций), что демонстрируется в разделе 3. Основным результатом работы является теорема 1 раздела 3, в которой результаты статьи [1] переносятся на онлайн контекст.

Во избежание большого количества громоздких выражений опускается часть (наиболее очевидных, но громоздких) выкладок, но дается подробное описание, как они могут быть сделаны. В изложении нет стремления к общности. В частности, для большинства оценок данной работы можно не только выписать точные константы в оценках сходимости в среднем (для этого вполне достаточно написанного в данной статье), но и получить оценки вероятностей больших уклонений. Также можно накладывать более общие требования на классы изучаемых семейств функций, делая константы, характеризующие семейство, не универсальными (одинаковыми для всех шагов), а зависящими от номера шага [3, 15, 16].

Полученные оценки с учетом известных нижних оценок [2, 7, 9, 17, 18] позволяют говорить о том, что в настоящей работе предложены достаточно эффективные методы, доминирующие в ряде случаев над существующими алгоритмами.

2. Метод зеркального спуска для задач стохастической онлайн оптимизации с неточным оракулом

Сформулируем основную задачу стохастической онлайн оптимизации с неточным оракулом: требуется подобрать последовательность $\left\{x^k\right\} \in Q$ (Q – выпуклое множество) так, чтобы минимизировать псевдорегрет [2–10]:

$$\text{Regret}_N\left(\left\{f_k\left(\,\cdot\,\right)\right\}, \left\{x^k\right\}\right) = \frac{1}{N}\sum_{k=1}^{N} f_k\left(x^k\right) - \min_{x\in Q}\frac{1}{N}\sum_{k=1}^{N} f_k\left(x\right) \tag{1}$$

на основе доступной информации

$$\left\{\nabla_x \tilde{f}_1\left(x^1, \xi^1\right); \ldots; \nabla_x \tilde{f}_{k-1}\left(x^{k-1}, \xi^{k-1}\right)\right\}$$

при расчете x^k. Причем выполнено **условие**[2]

1) для любых $N \in \mathbb{N}$ (Ξ^{k-1} – сигма алгебра, порожденная $\xi^1, \ldots, \xi^{k-1}$)

$$\sup_{\left\{x^k = x^k\left(\xi^1, \ldots, \xi^k\right)\right\}_{k=1}^N} E\left[\frac{1}{N}\sum_{k=1}^N \left\langle E_{\xi^k}\left[\nabla_x f_k\left(x^k, \xi^k\right) - \right.\right.\right.$$
$$\left.\left.\left. - \nabla_x \tilde{f}_k\left(x^k, \xi^k\right) \middle| \Xi^{k-1}\right], x^k - x_*\right\rangle\right] \leqslant \sigma,$$

где x_* – решение задачи

$$\frac{1}{N}\sum_{k=1}^N f_k(x) \to \min_{x \in Q},$$
$$E_{\xi^k}\left[\nabla_x f_k\left(x, \xi^k\right)\right] = \nabla f_k(x).$$

Здесь случайные величины $\left\{\xi^k\right\}$ могут считаться независимыми одинаково распределенными. Онлайновость постановки задачи допускает, что на каждом шаге k функция $f_k(\,\cdot\,)$ может выбираться из рассматриваемого класса функций враждебно по отношению к используемому методу генерации последовательности $\left\{x^k\right\}$. В частности, $f_k(\,\cdot\,)$ может зависеть от

$$\left\{x^1, \xi^1, f_1(\,\cdot\,); \ldots; x^{k-1}, \xi^{k-1}, f_{k-1}(\,\cdot\,); x^k\right\}.$$

Относительно класса функций, из которого выбираются $\{f_k(\,\cdot\,)\}$, в данной работе в зависимости от контекста будут предполагаться выполненными следующие **условия**:

2) $\{f_k(\,\cdot\,)\}$ – выпуклые функции (считаем, что это условие имеет место всегда);

3) $\{f_k(\,\cdot\,)\}$ – γ_2-сильно выпуклые функции в l_2;

4) для любых $k = 1, \ldots, N$, $x \in Q$

$$E_\xi\left[\left\|\nabla_x \tilde{f}_k(x, \xi)\right\|_*^2\right] \leqslant M^2.$$

Выше (и далее в статье) используется стандартная терминология онлайн оптимизации (см., например, [2, 6, 7, 9]). Однако в отечественной литературе на данный момент имеется определенный дефицит работ по этой (достаточно популярной на западе) тематике. В связи с этим было решено “разбавить” данную статью несколькими простыми примерами, которые позволят лучше прочувствовать смысл используемых понятий.

---

[2] В частности, если

$$\left\|\nabla_x \tilde{f}_k\left(x^k, \xi^k\right) - \nabla_x f_k\left(x^k, \xi^k\right)\right\|_* \leqslant \delta, \quad \max_{x, y \in Q}\|x - y\| \leqslant R,$$

то $\sigma \leqslant \delta R$.

Пример 1 (взвешивание экспертных решений, линейные потери). Рассмотрим задачу взвешивания экспертных решений, следуя [2]. Имеется n различных Экспертов. Каждый Эксперт "играет" на рынке. Игра повторяется $N \gg 1$ раз. Пусть l_i^k – проигрыш (выигрыш со знаком минус) Эксперта i на шаге k ($\left|l_i^k\right| \leqslant M$). На каждом шаге k распределяется один доллар между Экспертами, согласно вектору

$$x^k \in Q = S_n(1) = \left\{ x \geqslant 0 : \ \sum_{i=1}^n x_i = 1 \right\}.$$

Потери, которые при этом возникают, рассчитываются по потерям Экспертов $f_k(x) = \left\langle l^k, x \right\rangle$. Целью является таким образом организовать процедуру распределения доллара на каждом шаге, чтобы суммарные потери (за N шагов) были бы минимальными. Допускается, что потери Экспертов l^k могут зависеть еще и от текущего хода x^k. Установленные далее в этом разделе результаты (формула (2) с $R^2 = \ln n$) позволяют утверждать, что если на каждом шаге можно наблюдать лишь зашумленные проигрыши Экпертов

$$\frac{\partial \tilde{f}_k\left(x^k, \xi^k\right)}{\partial x_i} = l_i^k + \xi_i^k + \delta_i^k,$$

где $\left\{\xi_i^k\right\}$ – независимые одинаково распределенные случайные величины $\xi_i^k \in N(0,1)$, $\left|\delta_i^k\right| \leqslant \delta$, то существует такой способ действий $x^k(\nabla_x \tilde{f}_1(x^1, \xi^1), \ldots$ $\ldots, \nabla_x \tilde{f}_{k-1}(x^{k-1}, \xi^{k-1}))$ (метод зеркального спуска с $\|\| = \|\|_1$, $d(x) = \ln n +$ $+ \sum_{i=1}^n x_i \ln x_i$, см. ниже), который позволяет с вероятностью не менее 0,999 после N шагов проиграть лучшему (на этом периоде $k = 1, \ldots, N$). Эксперту не более $\mathrm{O}((M+1)\sqrt{N \ln n} + \delta N)$ долларов, что означает (см. формулу (1))

$$\mathrm{Regret}_N\left(\{f_k(\cdot)\}, \{x^k\}\right) = \mathrm{O}\left((M+1)\sqrt{\frac{\ln n}{N}} + \delta\right).$$

При $\delta = 0$ эта оценка оптимальна для данного класса задач [2]. □

Опишем метод зеркального спуска для решения задачи (1) (здесь можно следовать огромному числу литературных источников, в основном следуем [15, 19]). Введем норму $\| \ \|$ в прямом пространстве (сопряженную норму будем обозначать как $\| \ \|_*$) и прокс-функцию $d(x)$, сильно выпуклую относительно этой нормы, с константой сильной выпуклости $\geqslant 1$. Выберем точку старта

$$x^1 = \arg\min_{x \in Q} d(x),$$

считаем, что $d\left(x^1\right) = 0$, $\nabla d\left(x^1\right) = 0$.

Введем брэгмановское "расстояние"

$$V_x(y) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle.$$

Везде в дальнейшем будем считать, что

$$d(x) = V_{x^1}(x) \leqslant R^2 \quad \text{для всех} \quad x \in Q.$$

Определим оператор “проектирования” согласно этому расстоянию

$$\mathrm{Mirr}_{x^k}(g) = \arg\min_{y \in Q} \left\{ \left\langle g, y - x^k \right\rangle + V_{x^k}(y) \right\}.$$

Метод зеркального спуска (МЗС) для задачи (1) будет иметь вид, см., например, [19],

$$x^{k+1} = \mathrm{Mirr}_{x^k}\left(\alpha_k \nabla_x \tilde{f}_k\left(x^k, \xi^k\right)\right), \quad k = 1, \dots, N.$$

Тогда при выполнении условии (2) для любого $u \in Q$, $k = 1, \dots, N$ имеет место неравенство, см., например, [19],

$$\alpha_k \left\langle \nabla_x \tilde{f}_k\left(x^k, \xi^k\right), x^k - u \right\rangle \leqslant \frac{\alpha_k^2}{2} \left\| \nabla_x \tilde{f}_k\left(x^k, \xi^k\right) \right\|_*^2 + V_{x^k}(u) - V_{x^{k+1}}(u).$$

Это неравенство несложно получить в случае евклидовой прокс-структуры $d(x) = \|x\|_2^2/2$ [20] (в этом случае МЗС для задачи (1) есть просто вариант обычного метода проекции градиента). Разделим сначала выписанное неравенство на α_k и возьмем условное математическое ожидание $E_{\xi^{k+1}}\left[\,\cdot \mid \Xi^k\right]$, затем просуммируем то, что получится, по $k = 1, \dots, N$, используя условие 1. Затем возьмем от того, что получилось при суммировании, полное математическое ожидание, учитывая условие 4. В итоге, выбирая $u = x_*$, получим при условиях 1, 2, 4, $\alpha_k \equiv \alpha$ [10]

$$\begin{gathered} NE\left[\mathrm{Regret}_N\left(\{f_k(\,\cdot\,)\}, \left\{x^k\right\}\right)\right] \leqslant \\ \leqslant \frac{V_{x^1}(x_*)}{\alpha} - \frac{E\left[V_{x^{N+1}}(x_*)\right]}{\alpha} + \left(\frac{1}{2}M^2\alpha + \sigma\right)N \leqslant \\ \leqslant \frac{R^2}{\alpha} + \left(\frac{1}{2}M^2\alpha + \sigma\right)N, \end{gathered}$$

выбирая[3]

$$\alpha = \frac{R}{M}\sqrt{\frac{2}{N}},$$

получим

$$E\left[\mathrm{Regret}_N\left(\{f_k(\,\cdot\,)\}, \left\{x^k\right\}\right)\right] \leqslant MR\sqrt{\frac{2}{N}} + \sigma; \tag{2}$$

при условиях[4] 1, 3, 4, $\alpha_k \equiv (\gamma_2 k)^{-1}$, $\|\| = \|\|_2$ [9]

$$E\left[\mathrm{Regret}_N\left(\{f_k(\,\cdot\,)\}, \left\{x^k\right\}\right)\right] \leqslant \frac{M^2}{2\gamma_2 N}(1 + \ln N) + \sigma. \tag{3}$$

---

[3] Можно получить и адаптивный вариант приводимой далее оценки, для этого потребуется использовать метод двойственных усреднений [10, 19, 20].

[4] Отметим, что при условии 2 также используется неравенство

$$f_k(x^k) - f(x_*) \leqslant \langle \nabla f_k(x^k), x^k - x_* \rangle$$

при преобразовании левой части неравенства в псевдорегрет, а при условии 3 более точное неравенство

$$2(f_k(x^k) - f(x_*)) \leqslant 2\langle \nabla f_k(x^k), x^k - x_* \rangle - \gamma_2 \|x^k - x_*\|_2^2.$$

Оценки (2), (3) являются неулучшаемыми с точностью до мультипликативного числового множителя. Причем верно это и для детерминированных (нестохастических) постановок, в которых нет шумов ($\sigma = 0$), в случае оценки (2) при этом можно ограничиться классом линейных функций [2].

Пример 2. Пусть $Q = B_p^n(1)$ – единичный шар в l_p-норме. Относительно оптимального выбора нормы и прокс-структуры можно заметить следующее: если $p \geqslant 2$, то в качестве нормы $\|\ \|$ оптимально выбирать l_2-норму и евклидову прокс-структуру. Определим q из $1/p + 1/q = 1$. Пусть $1 \leqslant p \leqslant 2$, тогда $q \geqslant 2$. Если при этом $q = o(\ln n)$, то оптимально выбирать l_p-норму, а прокс-структуру задавать прокс-функцией

$$d(x) = \frac{1}{2(p-1)} \|x\|_p^2 .$$

Во всех этих случаях

$$R^2 = \max_{x \in Q} d(x) = \mathrm{O}(1) .$$

Для $q \geqslant \Omega(\ln n)$ выберем l_a-норму, где

$$a = \frac{2 \ln n}{2 \ln n - 1},$$

а прокс-структуру будем задавать прокс-функцией

$$d(x) = \frac{1}{2(a-1)} \|x\|_a^2 .$$

В этом случае $R^2 = \mathrm{O}(\ln n)$. Детали см., например, в [15, 17].

3. Одноточечные и многоточечные нелинейные многорукие бандиты

Везде в этом разделе будем считать, что все функции $f_k(x)$ и реализации $f_k(x, \xi)$ определены в Q_{μ_0} – μ_0-окрестности множества Q и удовлетворяют соответствующим условиям из раздела 2 именно в Q_{μ_0}.

Пусть требуется подобрать последовательность $\left\{x^k\right\} \in Q$ так, чтобы минимизировать псевдорегрет (1) на основе доступной информации ($m = 1, 2$)

$$\left\{\left\{\tilde{f}_1\left(x_i^1, \xi^1\right)\right\}_{i=1}^m ; \ldots ; \left\{\tilde{f}_{k-1}\left(x_i^{k-1}, \xi^{k-1}\right)\right\}_{i=1}^m\right\}$$

при расчете x^k. Будем предполагать, что имеет место следующее **условие**:

5) для любых $k \in \mathbb{N}$, $i = 1, \ldots, m$, $x_i^k \in Q_{\mu_0}$

$$\left|\tilde{f}_k\left(x_i^k, \xi^k\right) - f_k\left(x_i^k, \xi^k\right)\right| \leqslant \delta,$$

$$E_{\xi^k}\left[f_k\left(x_i^k, \xi^k\right)\right] = f_k\left(x_i^k\right),$$

$$E_{\xi^k}\left[\tilde{f}_k\left(x_i^k, \xi^k\right)^2\right] \leqslant B^2,$$

и в зависимости от контекста **условия**

6) для любых $k = 1, \dots, N$, $x, y \in Q_{\mu_0}$ (далее, как правило, это условие будет использоваться при $r = 2$, исключение сделано в табл. 2)

$$|f_k(x,\xi) - f_k(y,\xi)| \leqslant M_r(\xi)\|x-y\|_r, \quad M_r = \sqrt{E_\xi\left[M_r(\xi)^2\right]} < \infty;$$

7) для любых $k = 1, \dots, N$, $x, y \in Q_{\mu_0}$

$$\|\nabla_x f_k(x,\xi) - \nabla_x f_k(y,\xi)\|_2 \leqslant L_2(\xi)\|x-y\|_2, \quad L_2 = \sqrt{E_\xi\left[L_2(\xi)^2\right]} < \infty.$$

Введем аналоги $\nabla_x \tilde{f}_k(x,\xi)$ из раздела 2 ($\mu \leqslant \mu_0$)

$$\nabla_x \tilde{f}_k(x;e,\xi) := \frac{n}{\mu}\tilde{f}_k(x+\mu e,\xi)\,e \quad (\text{при } m=1),$$
$$\nabla_x \tilde{f}_k(x;e,\xi) := \frac{n}{\mu}\left(\tilde{f}_k(x+\mu e,\xi) - \tilde{f}_k(x,\xi)\right)e \quad (\text{при } m=2),$$

где $e \in RS_2^n(1)$, т.е. случайный вектор e равномерно распределен на сфере радиуса 1 в l_2. Считаем, что разыгрывание e происходит независимо ни от чего. Аналогично можно определить незашумленную оценку стохастического градиента $\nabla_x f_k(x;e,\xi)$, убрав в правой части тильды (волны).

Онлайновость постановки задачи допускает, что на каждом шаге k функция $f_k(\cdot)$ может выбираться из рассматриваемого класса функций враждебно по отношению к используемому методу генерации последовательности $\left\{x^k\right\}$. В частности, $f_k(\cdot)$ может зависеть от

$$\left\{x^1, \xi^1, f_1(\cdot); \dots; x^{k-1}, \xi^{k-1}, f_{k-1}(\cdot)\right\}.$$

Более того, при выборе $f_k(\cdot)$ считается полностью известной излагаемая стратегия. Подчеркнем, что поскольку стратегия рандомизированная, то речь идет об описании этой стратегии, а не о реализации. Это означает, что тому, кто подбирает $f_k(\cdot)$, известно, что $e \in RS_2^n(1)$, но неизвестно, как именно он разыгрывается. Это важная оговорка, потому что если допускать, как и в разделе 2, что на каждом шаге k реализация $e_k \in RS_2^n(1)$ становится известной тому, кто враждебно подбирает $f_k(\cdot)$, то нельзя получить оценку псевдорегрета лучше, чем $\Omega(N)$ [3]. Причины этого связаны с введением рандомизации, и на более простой задаче (линейные одноточечные многорукие бандиты) поясняются, например, в [10].

Сгладим исходную постановку с помощью локального усреднения по евклидову шару радиуса $\mu > 0$, который будет выбран позже,

$$f_k^\mu(x,\xi) = E_{\tilde{e}}\left[f_k(x+\mu\tilde{e},\xi)\right],$$
$$f_k^\mu(x) = E_{\tilde{e},\xi}\left[f_k(x+\mu\tilde{e},\xi)\right],$$

где $\tilde{e} \in RB_2^n(1)$, т.е. случайный вектор e равномерно распределен на шаре радиуса 1 в l_2. Заменим исходную задачу (1) задачей минимизации

$$\text{Regret}_N^{\mu}\left(\{f_k(\,\cdot\,)\},\{x^k\}\right) = \frac{1}{N}\sum_{k=1}^{N} f_k^{\mu}\left(x^k\right) - \min_{x\in Q}\frac{1}{N}\sum_{k=1}^{N} f_k^{\mu}(x). \tag{4}$$

Это делается для того, чтобы обеспечить выполнение условия 1 раздела 2, см. ниже. Будем считать, что имеют место условия 6, 7 (если условие 7 не выполнено, просто полагаем $L_2 = \infty$). Предположим также, что

$$\min\left\{M_2\mu, L_2\mu^2/2\right\} \leqslant \varepsilon/2,$$

т.е.

$$\mu \leqslant \max\left\{\frac{\varepsilon}{2M_2}, \sqrt{\frac{\varepsilon}{L_2}}\right\}, \tag{5}$$

где $\varepsilon = \varepsilon(N)$ определятся из условия (можно также сказать, что из этого условия определяется $N = N(\varepsilon)$)

$$E\left[\text{Regret}_N\left(\{f_k(\,\cdot\,)\},\{x^k\}\right)\right] \leqslant \varepsilon.$$

Из [1] следует, что при условии (5) из

$$\text{Regret}_N^{\mu}\left(\{f_k(\,\cdot\,)\},\{x^k\}\right) \leqslant \varepsilon/2$$

для тех же самых последовательностей $\{f_k(\,\cdot\,)\},\{x^k\}$ следует

$$\text{Regret}_N\left(\{f_k(\,\cdot\,)\},\{x^k\}\right) \leqslant \varepsilon.$$

Далее сконцентрируемся на минимизации сглаженной версии псевдорегрета (4), контролируя при этом выполнение условия (5).

Введенные выше $\nabla_x \tilde{f}_k(x;e,\xi)$ для задачи (4) удовлетворяют условию 1 с σ, равным соответственно

$$\sigma \leqslant E\left[\frac{\delta n}{N\mu}\sum_{k=1}^{N}|\langle e_k, r_k\rangle|\right] \leqslant \frac{2\delta R\sqrt{n}}{\mu} \quad (\text{при } m=1), \tag{6}$$

$$\sigma \leqslant E\left[\frac{2\delta n}{N\mu}\sum_{k=1}^{N}|\langle e_k, r_k\rangle|\right] \leqslant \frac{4\delta R\sqrt{n}}{\mu} \quad (\text{при } m=2), \tag{7}$$

где $E\left[r_k^2\right] \leqslant 2R^2$, $e_k \in RS_2^n(1)$ не зависит от $r_k = x^k - x_*$. Оценки (6), (7) следуют из того, что [1, 7, 13]

$$E_e\left[\nabla_x f_k(x;e,\xi)\right] = \nabla f_k^{\mu}(x,\xi),$$

и из явления концентрации равномерной меры на сфере вокруг экватора (при северном полюсе, заданном вектором r_k) [21].

Чтобы можно было воспользоваться оценками (2), (3) раздела 2, осталось в условии 4 раздела 2 оценить константу M. Выберем в прямом пространстве норму l_p, $1 \leqslant p \leqslant 2$ (см. пример 2 раздела 2). Положим $1/p + 1/q = 1$. При $m = 1$ и условии 5 имеем оценки [1]

$$M^2 \leqslant \frac{(q-1)\, n^{1+2/q} B^2}{\mu^2} \quad (\text{при } 2 \leqslant q \leqslant 2\ln n),$$
$$M^2 \leqslant \frac{4n \ln n B^2}{\mu^2} \quad (\text{при } 2\ln n < q \leqslant \infty).$$

Наиболее интересны случаи, когда $q = 2$, $q = \infty$:

(8) $$M^2 \leqslant \frac{n^2 B^2}{\mu^2} \quad (\text{при } q = 2),$$

(9) $$M^2 \leqslant \frac{4n \ln n B^2}{\mu^2} \quad (\text{при } q = \infty).$$

При $m = 2$ и выполнении условий условия 5, 6 имеем оценки [1] (случай $2 < q < \infty$ рассматривается совершенно аналогично)

$$M^2 \leqslant 3nM_2^2 + \frac{3}{4} n^2 L_2^2 \mu^2 + 12\frac{\delta^2 n^2}{\mu^2} \quad (\text{при } q = 2),$$
$$M^2 \leqslant 4\ln n M_2^2 + 3n \ln n L_2^2 \mu^2 + 48\frac{\delta^2 n \ln n}{\mu^2} \quad (\text{при } q = \infty).$$

В частности, если

(10) $$\mu \leqslant \min\left\{\max\left\{\frac{\varepsilon}{2M_2}, \sqrt{\frac{\varepsilon}{L_2}}\right\}, \frac{M_2}{L_2}\sqrt{\frac{4}{3n}}\right\}, \quad \delta \leqslant \frac{M_2\mu}{\sqrt{12n}} \quad (\text{при } q = 2),$$

(11) $$\mu \leqslant \min\left\{\max\left\{\frac{\varepsilon}{2M_2}, \sqrt{\frac{\varepsilon}{L_2}}\right\}, \frac{M_2}{L_2}\sqrt{\frac{1}{6n}}\right\}, \quad \delta \leqslant \frac{M_2\mu}{\sqrt{96n}} \quad (\text{при } q = \infty),$$

то

(12) $$M^2 \leqslant 5nM_2^2 \quad (\text{при } q = 2),$$

(13) $$M^2 \leqslant 5\ln n M_2^2 \quad (\text{при } q = \infty).$$

Полагая в (6), (7) $\sigma \leqslant \varepsilon/4$, получим дополнительно к (5) (и (10), (11) при $m = 2$) условия на δ, μ, ε

$$\frac{2\delta R\sqrt{n}}{\mu} \leqslant \frac{\varepsilon}{4} \quad (\text{при } m = 1),$$
$$\frac{4\delta R\sqrt{n}}{\mu} \leqslant \frac{\varepsilon}{4} \quad (\text{при } m = 2),$$

т.е.

$$\delta \leqslant \frac{\varepsilon\mu}{8R\sqrt{n}} \quad (\text{при } m=1), \tag{14}$$

$$\delta \leqslant \frac{\varepsilon\mu}{16R\sqrt{n}} \quad (\text{при } m=2). \tag{15}$$

Далее надо воспользоваться оценками (2), (3), добиваясь соответственно

$$MR\sqrt{\frac{2}{N}} \leqslant \frac{\varepsilon}{4}, \tag{16}$$

$$\frac{M^2}{2\gamma_2 N}\left(1+\ln N\right) \leqslant \frac{\varepsilon}{4}. \tag{17}$$

Таким образом, при $m=1$ получаем оценки на $\mu(\varepsilon)$ из (5), на $\delta(\varepsilon)$ из (14) и оценки $\mu(\varepsilon)$, на $N(\varepsilon)$ из (8), (9), (16), (17) и оценки $\mu(\varepsilon)$; при $m=2$ получаем оценки на $\mu(\varepsilon)$ из (10), (11), на $\delta(\varepsilon)$ из (10), (11), (15) и оценки $\mu(\varepsilon)$, на $N(\varepsilon)$ из (12), (13), (16), (17).

Не будем здесь выписывать то, что получается, – это довольно тривиально, но достаточно громоздко. Вместо этого резюмируем полученные в работе результаты в более наглядной форме. Для этого введем $\tilde{\mathrm{O}}(\ )$. Будем считать, что $\tilde{\mathrm{O}}(\ )$ с точностью до логарифмического множителя (от n и/или N) совпадает с $\mathrm{O}(\ )$.

Напомним (обозначения см. в разделе 2 и примере 2), что

$$x^{k+1} = \mathrm{Mirr}_{x^k}\left(\alpha_k \nabla_x \tilde{f}_k\left(x^k; e^k, \xi^k\right)\right), \quad k=1,\ldots,N,$$

где $\left\{e^k\right\}_{k=1}^N$ – независимые одинаково распределенные случайные векторы $e^k \in RS_2^n(1)$,

$$\nabla_x \tilde{f}_k\left(x^k; e^k, \xi^k\right) := \frac{n}{\mu}\tilde{f}_k\left(x^k+\mu e^k, \xi^k\right)e^k \quad (\text{при } m=1),$$

$$\nabla_x \tilde{f}_k\left(x^k; e^k, \xi^k\right) := \frac{n}{\mu}\left(\tilde{f}_k\left(x^k+\mu e^k, \xi^k\right) - \tilde{f}_k\left(x^k, \xi^k\right)\right)e^k \quad (\text{при } m=2),$$

$$\alpha_k \equiv \alpha = \frac{R}{M}\sqrt{\frac{2}{N}}$$

– в общем случае и

$$\alpha_k \equiv \left(\gamma_2 k\right)^{-1},$$

если $f_k(x)$ – γ_2-сильно выпуклые функции в l_2 (в этом случае выбирают $p = 2$).

Т е о р е м а 1. *Пусть рассматривается задача стохастической онлайн оптимизации (1) в постановке, описанной в этом разделе (в безградиентном варианте). Пусть выбрана l_p-норма, $1 \leqslant p \leqslant 2$ (см. раздел 2). Согласно этой*

Таблица 1

$m=1$	$f_k(x)$ – выпуклые функции	$f_k(x)$ – γ_2-сильно выпуклые функции в l_2 норме и $p=2$
Выполнены условия 5, 6	$\tilde{\mathrm{O}}\left(\frac{B^2M_2^2R^2n^{1+2/q}}{\varepsilon^4}\right)$	$\tilde{\mathrm{O}}\left(\frac{B^2M_2^2n^2}{\gamma_2\varepsilon^3}\right)$
Выполнены условия 5, 7	$\tilde{\mathrm{O}}\left(\frac{B^2L_2R^2n^{1+2/q}}{\varepsilon^3}\right)$	$\tilde{\mathrm{O}}\left(\frac{B^2L_2n^2}{\gamma_2\varepsilon^2}\right)$

Таблица 2

$m=2$	$f_k(x)$ – выпуклые функции	$f_k(x)$ – γ_2-сильно выпуклые функции в l_2 норме и $p=2$
Выполнены условия 5, 6	$\tilde{\mathrm{O}}\left(\frac{M_p^2R^2n^2}{\varepsilon^2}\right)$	$\tilde{\mathrm{O}}\left(\frac{M_2^2n^2}{\gamma_2\varepsilon}\right)$
Выполнены условия 5, 6, 7	$\tilde{\mathrm{O}}\left(\frac{M_2^2R^2n^{2/q}}{\varepsilon^2}\right)$	$\tilde{\mathrm{O}}\left(\frac{M_2^2n}{\gamma_2\varepsilon}\right)$

норме задана прокс-функция и расстояние Брэгмана $V_x(y)$*. Пусть* $R^2 = V_{x^1}(x_*)$*, где* x^1 *и* x_* *определены в разделе 2. Тогда*

$$E\left[\mathrm{Regret}_{N(\varepsilon)}\left(\{f_k(\cdot)\},\left\{x^k\right\}\right)\right] \leqslant \varepsilon,$$

где $N(\varepsilon)$ *определяется в табл. 1, 2.*

Обе табл. 1 и 2 заполняются исходя из описанной выше техники. Исключением является вторая строчка табл. 2, она взята из [1]. Несложно выписать точные формулы вместо $\tilde{\mathrm{O}}(\ )$ во всех полях обеих таблиц. Также несложно выписать условие на допустимый уровень шума δ, при котором мультипликативная константа в точной формуле увеличится, скажем, не более чем в 2 раза.

Оценки в третьей строчке табл. 2 неулучшаемы [11] (соответствуют нижним оценкам). Оценки во второй строчке табл. 2 неулучшаемы по ε [12, 17]. Все сказанное выше касается и стохастических, но не онлайн, постановок [12, 17].

Относительно табл. 1 имеется гипотеза, что приведенные оценки неулучшаемы по n. По ε оценки могут быть улучшены за счет ухудшения того, как входит n [18].

В заключение рассмотрим пример, демонстрирующий, что полученные в теореме 1 результаты представляются интересными не только в онлайн контексте.

Пример 3. Предположим, что “успешность” некоторого человека зависит от того, как он распоряжается своим временем. Имеется n различных родов деятельности. В k-день человек распоряжается своим временем согласно вектору $x^k \in S_n(1)$. Этот вектор отражает доли времени, уделенные соответствующим делам. В конце каждого дня человек получает “обратную связь” от “внешнего мира” вида

$$\tilde{f}\left(x^k+\mu e^k,\xi^k\right) = f\left(x^k+\mu e^k\right)\cdot\left(1+\xi^k\right),$$

где $\left\{e^k\right\}$ – независимые одинаково распределенные случайные векторы $e^k \in$ $\in RS_2^n(1)$, μ определяется согласно формуле (5), $\left\{\xi^k\right\}$ – независимые (между собой и от $\left\{e^k\right\} \in RS_2^n\left(1\right)$) одинаково распределенные случайные величины $\xi^k \in N\left(0,1\right)$, а выпуклая функция $f\left(x\right)$ со свойствами

$$|f\left(x\right)| \leqslant B, \quad |f\left(x\right) - f\left(y\right)| \leqslant M_2 \left\|x - y\right\|_2$$

правильно отражает реальное "положение дел", т.е. минимум этой функции соответствует оптимальной для данного человека конфигурации. Задача человека заключается в том, чтобы (получая каждый день описанную выше обратную связь) так организовать "процесс своего обучения" (на основе получаемой информации), чтобы как можно быстрее достичь такого состояния[5]

$$\bar{x}^N = \frac{1}{N}\sum_{k=1}^{N} x^k \in S_n\left(1\right),$$

что с вероятностью 0,999 имеет место неравенство

$$f\left(\bar{x}^N\right) - \min_{x\in S_n(1)} f\left(x\right) \leqslant \mathrm{Regret}_N\left(\left\{f\left(\,\cdot\,\right)\right\}, \left\{x^k\right\}\right) \leqslant \varepsilon.$$

Согласно теореме 1 человек может этого достичь за

$$N = \mathrm{O}\left(\frac{B^2 M_2^2 n \ln^2 n}{\varepsilon^4}\right)$$

дней. Если есть возможность получать каждый день информацию $\tilde{f}(x^k + \mu e^k, \xi^k)$ и $\tilde{f}(x^k, \xi^k)$, где μ определяется согласно формуле (11), то тогда можно улучшить оценку до

$$N = \mathrm{O}\left(\frac{M_2^2 \ln^2 n}{\varepsilon^2}\right)$$

дней – здесь предполагается также гладкость $f\left(x\right)$.

4. Заключение

В работе предложены эффективные методы нулевого порядка (также говорят прямые методы или безградиентные методы) для задач выпуклой стохастической онлайн оптимизации. Методы строились на базе обычного метода зеркального спуска для задач стохастической оптимизации. Вместо стохастического градиента в метод зеркального спуска подставлялись специальные дискретные аналоги, аппроксимирующие стохастический градиент. При правильном пересчете размера шага получаются эффективные методы.

[5] Заметим, что с достаточной точностью и доверительным уровнем (при $N \gg 1$) можно считать, что

$$\bar{x}^N \simeq \frac{1}{N}\sum_{k=1}^{N}\left(x^k + \mu e^k\right).$$

СПИСОК ЛИТЕРАТУРЫ

1. *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 // АиТ. 2016. № 10. P. 57–78.
2. *Lugosi G., Cesa-Bianchi N.* Prediction, Learning and Games. N.Y.: Cambridge Univer. Press, 2006.
3. *Agarwal A., Dekel O., Xiao L.* Optimal Algorithm for Online Convex Optimization with Multi-Point Bandit Feedback // COLT. 2010. P. 28–40.
4. *Sridharan K.* Learning from an Optimization Viewpoint // PhD Thesis, Toyota Technol. Institut. Chicago, 2011. arXiv:1204.4145
5. *Bubeck S.* Introduction to Online Optimization. Princeton Univer.: Lecture Notes, 2011. http://www.princeton.edu/~sbubeck/BubeckLectureNotes.pdf
6. *Shalev-Shwartz S.* Online Learning and Online Convex Optimization // Foundat. Trends Machin. Learning. 2011. V. 4. No. 2. P. 107–194. http://www.cs.huji.ac.il/~shais/papers/OLsurvey.pdf
7. *Bubeck S., Cesa-Bianchi N.* Regret Analysis of Stochastic and Nonstochastic Multi-Armed Bandit Problems // Foundat. Trends Machin. Learning. 2012. V. 5. No. 1. P. 1–122. http://www.princeton.edu/~sbubeck/SurveyBCB12.pdf
8. *Rakhlin A., Sridharan K.* Statistical Learning Theory and Sequential Prediction // e-print, 2014. http://stat.wharton.upenn.edu/~rakhlin/book_draft.pdf
9. *Hazan E.* Introduction to online convex optimization // e-print, 2015. http://ocobook.cs.princeton.edu/OCObook.pdf
10. *Гасников А.В., Нестеров Ю.Е., Спокойный В.Г.* Об эффективности одного метода рандомизации зеркального спуска в задачах онлайн оптимизации // ЖВМ и МФ. 2015. Т. 55. № 4. С. 55–71. arXiv:1410.3118
11. *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 Transact. Inform. 2015. V. 61. No. 5. P. 2788–2806. http://www.eecs.berkeley.edu/~wainwrig/Papers/DucZero15.pdf
12. *Немировский А.С., Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. М.: Наука, 1979.
13. *Flaxman A.D., Kalai A.T., McCahan H.B.* Online Convex Optimization in the Bandit Setting: Gradient Descent without a Gradient // Proc. 16 Annual ACM-SIAM sympos Discret. Algorithm. 2005. P. 385–394. http://research.microsoft.com/en-us/um/people/adum/publications/2005-Online_Convex_Optimization_in_the_Bandit_Setting.pdf
14. *Juditsky A., Nemirovski A.* First Order Methods for Nonsmooth Convex Large-Scale Optimization, I, II. Optim. Machine Learning. Ed. S. Sra, S. Nowozin, S. Wright. MIT Press, 2012.
15. *Гасников А.В., Двуреченский П.Е., Нестеров Ю.Е.* Стохастические градиентные методы с неточным оракулом // Тр. МФТИ. 2016. Т. 8. № 1. С. 41–91. arxiv:1411.4218
16. *Nemirovski A.* Lectures on Modern Convex Optimization Analysis, Algorithms, and Engineering Applications. Philadelphia: SIAM, 2013. http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf
17. *Agarwal A., Bartlett P.L., Ravikumar P., Wainwright M.J.* Information-Theoretic Lower Bounds on the Oracle Complexity of Stochastic Convex Optimization // IEEE Transact. Inform. 2012. V. 58. P. 3235–3249. arXiv:1009.0571

18. *Bubeck S., Eldan R.* Multi-Scale Exploration of Convex Functions and Bandit Convex Optimization // e-print, 2015. http://research.microsoft.com/en-us/um/people/sebubeck/ConvexBandits.pdf
19. *Allen-Zhu Z., Orecchia L.* Linear Coupling: An Ultimate Unification of Gradient and Mirror Descent // e-print, 2014. arXiv:1407.1537
20. *Nesterov Y.* Primal-Dual Subgradient Methods for Convex Problems // Math. Program. Ser. B. 2009. V. 120(1). P. 261–283.
21. *Ledoux M.* Concentration of Measure Phenomenon. Providence, RI, Amer. Math. Soc., 2001 (Math. Surveys Monogr. V. 89).

Статья представлена к публикации членом редколлегии П.С. Щербаковым.

Поступила в редакцию 16.10.2014