ПРОГРАММИРОВАНИЕ, 2023, № 6, с. 5–13

АНАЛИЗ ДАННЫХ

УДК 519.85

# АДАПТИВНЫЕ МЕТОДЫ ДЛЯ ВАРИАЦИОННЫХ НЕРАВЕНСТВ С ОТНОСИТЕЛЬНО ГЛАДКИМИ И ОТНОСИТЕЛЬНО СИЛЬНО МОНОТОННЫМИ ОПЕРАТОРАМИ


С. С. Аблаев[a,b,*] (ORCID: 0000-0002-9927-6503),
Ф. С. Стонякин[a,b,**] (ORCID: 0000-0002-9250-4438), М. С. Алкуса[a,c,***] (ORCID: 0000-0001-5470-0182),
Д. А. Пасечнюк[a,d,****] (ORCID: 0000-0002-1208-1659)

[a]Московский физико-технический институт,
141701, г. Долгопрудный, Институтский пер., 9, Россия

[b]Крымский федеральный университет им. В.И. Вернадского,
295007, г. Симферополь, проспект академика Вернадского, 4, Россия

[c]Национальный исследовательский университет “Высшая школа экономики”,
101000, г. Москва, ул. Мясницкая, д. 20, Россия

[d]Исследовательский центр доверенного искусственного интеллекта ИСП РАН,
109004, г. Москва, ул. Александра Солженицына, 25, Россия

*E-mail: seydamet.ablaev@yandex.ru

**E-mail: fedyor@mail.ru

***E-mail: mohammad.alkousa@phystech.edu

****E-mail: dmivilensky1@gmail.com

Поступила в редакцию 13.06.2023 г.
После доработки 15.07.2023 г.
Принята к публикации 20.07.2023 г.

Статья посвящена некоторым адаптивным методам для вариационных неравенств с относительно гладкими и относительно сильно монотонными операторами. Отталкиваясь от недавно предложенного проксимального варианта экстраградиентного метода для такого класса задач, мы детально исследуем метод с адаптивно подбираемыми значениями параметров. Доказана оценка скорости сходимости этого метода. Результат обобщен на класс вариационных неравенств с относительно сильно монотонными $\delta$-обобщенно гладкими операторами вариационного неравенства. Для задачи гребневой регрессии и вариационного неравенства, связанного с параллелепипедно-симплексными играми, выполнены численные эксперименты, демонстрирующие эффективность предложенной методики адаптивного подбора параметров в ходе реализации алгоритма.

**DOI:** 10.31857/S013234742306002X, **EDN:** EIFITN

## 1. ВВЕДЕНИЕ

Свойства относительной гладкости и относительной сильной монотонности для вариационных неравенств были введены довольно недавно в [1] и [2] соответственно как аналоги свойств $L$-гладкости и относительной сильной выпуклости для минимизационных задач. В данной работе, отправляясь от неадаптивного варианта проксимального зеркального метода из [2], для относительно гладких относительно сильно монотонных операторов детально исследован адаптивный алгоритм такого типа, анонсированный ранее в трудах конференции “OPTIMA-2022” [3]. Более того, по аналогии с [1] рассмотрены и аналоги такого подхода на классе вариационных неравенств с относительно сильно монотонными и $\delta$-обобщенно гладкими операторами, получены оценки качества выдаваемого решения, описывающие влияние на него параметра $\delta$.

Будем рассматривать задачу нахождения решения $x_*$ (также называемого слабым решением) вариационного неравенства

$$\max_{x\in Q}\langle g(x), x_* - x\rangle \le 0, \tag{1.1}$$

где $Q$ — выпуклое замкнутое подмножество $\mathbb{R}^n$, $g : Q \to \mathbb{R}^n$. Предположим, что удовлетворяющее (1.1) решение $x_*$ существует.

Всюду далее будем предполагать, что нам доступна некоторая выпуклая (вообще говоря, не сильно выпуклая) дифференцируемая прокс-функция $d$, порождающая расстояние, а также соответствующая ей дивергенция (расхождение) Брэгмана

$$V(y,x) := d(y) - d(x) - \langle \nabla d(x), y - x\rangle. \qquad (1.2)$$

В работе [4] выделен класс гладких задач относительно произвольной дивергенции Брэгмана (порожденной необязательно сильно выпуклыми $d$) и рассмотрен ряд примеров. Напомним следующий аналог понятия относительной сильной выпуклости функции [5] для вариационных неравенств [2].

**Определение 1.** *Оператор $g$ называется относительно $\mu$-сильно монотонным, где $\mu > 0$, если для всяких $x, y \in Q$ верно неравенство*

$$\mu V(y,x) + \mu V(x,y) \le \langle g(y) - g(x), y - x\rangle. \qquad (1.3)$$

Поясним на следующем примере, почему относительная сильная монотонность оператора вводится именно согласно (1.3).

**Пример 1.** *Если $f : Q \to \mathbb{R}$ — относительно $\mu$-сильно выпуклая дифференцируемая функция*

$$f(x) - f(y) + \mu V(x,y) \le \langle \nabla f(x), x - y\rangle \quad \forall x, y \in Q, \qquad (1.4)$$

*то*

$$f(y) - f(x) + \mu V(y,x) \le \langle \nabla f(y), y - x\rangle \quad \forall x, y \in Q. \qquad (1.5)$$

*После сложения двух последних неравенств получаем $\forall x, y \in Q$*

$$\mu V(x,y) + \mu V(y,x) \le \langle \nabla f(y) - \nabla f(x), y - x\rangle.$$

*Таким образом, неравенство* (1.3) *верно при $g(x) = \nabla f(x)$, где $\nabla f(x)$ — градиент $f$ в точке $x$.*

Относительно сильно выпуклые функционалы возникают в самых разных ситуациях [5]. Рассмотрение таких задач на ограниченных допустимых множествах естественно приводит к вариационным неравенствам с соответствующим предположением относительной сильной монотонности оператора.

**Пример 2.** *Параллелепипедно-симплексные игры. Рассмотрим задачу централизованного решения вариационного неравенства, связанного с параллелепипедно-симплексными играми* (*box-simplex games*) [2] *следующего вида*:

$$\max_{y\in[-1,1]^n} \max_{z\in\Delta_n} f(y,z), \quad f(y,z) := z^\top Ay - \langle b, z\rangle + \langle c, y\rangle, \qquad (1.6)$$

*где $A \in \mathbb{R}^{n\times n}$ — положительно определенная матрица, а $z \in \Delta_n$, где $\Delta_n$ — стандартный единичный симплекс в пространстве $\mathbb{R}^n$. Параллелепипедно-симплексные игры обобщают задачу $\ell_\infty$-регрессии с параллелепипедными ограничениями, имеющую следующий вид*:

$$\min_{y\in[-1,1]^n} \|Ay - b\|_\infty,$$

$\|x\|_\infty := \max_k |x_k|$, *где* $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

*Приведенная задача эквивалентна задаче* (1.1) *решения вариационного неравенства, если выбрать для $x := (y,z)^\top$*

$$\hat{g}(x) := (A^\top z + c,\ b - Ay)^\top. \qquad (1.7)$$

*Оператор $\hat{g}$ монотонен, поскольку исходная седловая задача* (1.6) *является выпуклой по $y$* (*при любом $z$*) *и вогнутой по $z$* (*при любом $y$*).

*Рассмотрим оператор $g$, его относительно сильно монотонное приближение вида*

$$g(x) := (A^T z + c + \mu_y \nabla_y d(y,z),\ b - Ay + \mu_z \nabla_z d(y,z))^\top, \qquad (1.8)$$

*где $\mu_y, \mu_z > 0$ — некоторые фиксированные параметры с достаточно малыми положительными значениями, а согласно* [2] *прокс-функция в* (1.2) *может быть выбрана следующим образом*:

$$d(x) = z^\top |A|(y)^2 + 10\|A\|_\infty \sum_{i=1}^n z_i \ln z_i,$$

*где $|\cdot|$ и $(\cdot)^2$ действуют покомпонентно, $\|A\|_\infty := \sup_{\|x\|_\infty=1} \|Ax\|_\infty$.*

*Ясно, что введенный в* (1.8) *оператор в полученной задаче решения вариационного неравенства — $\min\{\mu_y, \mu_z\}$ — сильно монотонный относительно выбранной указанным выше способом прокс-функции $d$. Поэтому к этой задаче можно непосредственно применить алгоритм* 1, *что и будет сделано далее в разделе экспериментов.*

В качестве еще одного примера рассмотрим задачу централизованной распределенной минимизации эмпирического риска в предположении схожести слагаемых [6].

**Пример 3.** *Минимизация эмпирического риска.*

*Рассмотрим задачу централизованной распределенной минимизации эмпирического риска в предположении схожести слагаемых* [6],

$$F(x) := \frac{1}{m}\sum_{j=1}^m f_j(x) = \frac{1}{nm}\sum_{j=1}^m\sum_{i=1}^n \ell(x, z_i^{(j)}) = \frac{1}{N}\sum_{i=1}^N \ell(x, z_i) \to \min_{x\in Q} \qquad (1.9)$$

*в предположении, что исходные данные представляют из себя набор из $m$ выборок размера $n$, каждая из которых хранится на одном из $m$ серверов. При этом для достаточно большого $n$ все $f_j$ есть $\mu$-сильно выпуклые и $L$-гладкие* (*удовлетворяют условию Липшица градиента с константой $L > 0$*) *функционалы, которые можно считать статистически схожими*

[6]. *Такая схожесть может быть описана в виде предположения* [6] *о том, что для всякого* $x$

$$\|\nabla^2 F(x) - \nabla^2 f_j(x)\|_2 \leq \gamma,$$

*при всяком* $j$ *для некоторого достаточно малого* $\gamma > 0$. *Здесь и далее* $\|A\|_2 := \max_{\|x\|_2 \leq 1} \|Ax\|_2$ *для евклидовой нормы* $\|\cdot\|_2$. *При этом предполагается, что существует центральный сервер* (*ему соответствует функционал* $\overline{f}$), *на который передается информация о градиентах* $f_j$ *в текущей точке, но не передается информация о значениях* $f_j$. *В* [6] *показано, что при таком допущении можно ввести прокс-функцию*

$$d(x) := \overline{f}(x) + \frac{\gamma}{2}\|x\|_2^2, \tag{1.10}$$

*и для соответствующей дивергенции Брэгмана* $V(y,x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle$ *функция* $F$ *будет относительно* 1-*гладкой и относительно* $\frac{\mu}{\mu + 2\gamma}$-*сильно выпуклой, т.е. для всяких* $x, y \in Q$ *верны неравенства*

$$F(y) \leq F(x) + \langle \nabla F(x), y - x \rangle + V(y,x),$$

$$F(y) \geq F(x) + \langle \nabla F(x), y - x \rangle + \frac{\mu}{\mu + 2\gamma} V(y,x).$$

*Это означает, что при* $\gamma \ll L$ *для задач минимизации эмпирического риска* (*целевой функционал* $L$-*гладкий и* $\mu$-*сильно выпуклый*) *можно улучшить оценку*

$$O\left(\frac{L}{\mu}\log\frac{1}{\varepsilon}\right)$$

*сложности неускоренного градиентного метода до*

$$O\left(\left(1 + \frac{\gamma}{\mu}\right)\log\frac{1}{\varepsilon}\right),$$

*используя, по-прежнему, только информацию первого порядка.*

*Отметим, что при сопоставимых значениях параметров* $\gamma$ *и* $\mu$ (*этого ввиду малости* $\gamma$ *возможно добиться, к примеру, регуляризацией задачи*) *такая оценка близка к следующей оценке сложности ускоренных методов*:

$$O\left(\sqrt{1 + \frac{\gamma}{\mu}}\log\frac{1}{\varepsilon}\right).$$

*При этом на каждой итерации метода, выполняемой центральным узлом, доступна информация о градиенте целевого функционала* $F$, *но не доступна информация о значении функционала* $F$. *Доступность информации о градиенте позволяет рассматривать поставленную задачу минимизации эмпирического риска как задачу отыскания решения вариационного неравенства с относительно гладким и относительно сильно монотонным оператором* $g = \nabla F$. *Такой подход позволяет, в частности, предложить метод с полной адаптивной настройкой на параметр гладкости задачи* (1.9), *чего не удалось добиться для метода в* [6].

Статья состоит из введения, заключения и трех основных разделов. В разделе 2 рассматривается адаптивный алгоритм 1 для вариационных неравенств с относительно гладким и относительно сильно монотонным оператором. Показано, что увеличение количества отрешиваний вспомогательных подзадач, вызванное адаптивностью, некритично влияет на количество итераций. Раздел 3 содержит обобщение результатов раздела 2 на случай, когда оператор вариационного неравенства удовлетворяет условию $\delta$-обобщенной гладкости [1]. В разделе 4 приведены численные эксперименты, демонстрирующие работоспособность адаптивного подбора параметров для алгоритма 1.

## 2. АДАПТИВНЫЙ МЕТОД ДЛЯ ВАРИАЦИОННЫХ НЕРАВЕНСТВ С ОТНОСИТЕЛЬНО ГЛАДКИМ СИЛЬНО МОНОТОННЫМ ОПЕРАТОРОМ

В этом разделе, отталкиваясь от недавно предложенного в работе [2] метода, мы рассмотрим анонсированный в [3] алгоритм 1 без использования техники рестартов в случае относительно гладкого сильно монотонного оператора. Напомним, что оператор $g$ называется $L$-гладким, если при любых $x, y, z \in Q$ справедливо следующее неравенство

$$\langle g(y) - g(z), x - z \rangle \leq LV(x,z) + LV(z,y), \tag{2.1}$$

где $V(\cdot,\cdot)$ — это дивергенция Брэгмана.

**Алгоритм 1.** Адаптивный метод первого порядка для вариационных неравенств с относительно сильно монотонными операторами

**Require:** $x_0 \in Q, L_0 > 0, \mu > 0, d(\cdot), V(\cdot,\cdot)$.

1: $z_0 = \mathrm{argmin}_{u \in Q} d(u)$.

2: **for** $k \geq 0$ **do**

3: Найти наименьшее целое $i_k \geq 0$ такое, что

$$\begin{gathered}\langle g(z_k) - g(w_k), z_{k+1} - w_k \rangle \leq \\ \leq L_{k+1}\left(V(w_k, z_k) + V(z_{k+1}, w_k)\right),\end{gathered} \tag{2.2}$$

где $L_{k+1} = 2^{i_k - 1} L_k$, и

$$w_k = \arg\min_{y \in Q}\left\{\left\langle \frac{1}{L_{k+1}} g(z_k), y \right\rangle + V(y, z_k)\right\}, \tag{2.3}$$

$$z_{k+1} = \arg\min_{z\in Q}\left\{\left\langle \frac{1}{L_{k+1}} g(w_k), z\right\rangle + \right. \\ \left. + V(z, z_k) + \frac{\mu}{L_{k+1}} V(z, w_k)\right\}. \tag{2.4}$$

4: **end for**

**Ensure:** $z_k$.

**Теорема 1.** *Пусть $g$ — $\mu$-относительно сильно монотонный оператор и $z_*$ — точное решение вариационного неравенства* (1.1). *Тогда для алгоритма* 1 *верно следующее неравенство*:

$$V(z_*, z_{k+1}) \le \prod_{i=0}^{k}\left(1 + \frac{\mu}{L_{i+1}}\right)^{-1} V(z_*, z_0). \tag{2.5}$$

*Если, кроме этого, $L_0 \le 2L$, то имеет место неравенство*

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{2L}\right)^{-(k+1)} V(z_*, z_0). \tag{2.6}$$

*Доказательство.* Для $w_k$ и $z_{k+1}$, $\forall k \ge 0$, в (2.3) и (2.4) имеем

$$\frac{1}{L_{k+1}}\langle g(z_k), w_k - z_{k+1}\rangle \le V(z_{k+1}, z_k) - \\ - V(z_{k+1}, w_k) - V(w_k, z_k) \tag{2.7}$$

и

$$\frac{1}{L_{k+1}}\langle g(w_k), z_{k+1} - z_*\rangle \le V(z_*, z_k) - V(z_*, z_{k+1}) - \\ - V(z_{k+1}, z_k) + \frac{\mu}{L_{k+1}}(V(z_*, w_k) - V(z_*, z_{k+1})). \tag{2.8}$$

После сложения (2.7) и (2.8) получим

$$\frac{1}{L_{k+1}}(\langle g(z_k), w_k - z_{k+1}\rangle + \langle g(w_k), z_{k+1} - z_*\rangle) \le \\ \le -(V(w_k, z_k) + V(z_{k+1}, w_k)) + V(z_*, z_k) - \\ - V(z_*, z_{k+1}) + \frac{\mu}{L_{k+1}}(V(z_*, w_k) - V(z_*, z_{k+1})).$$

Используя неравенство (2.2), имеем

$$\langle g(z_k), z_{k+1} - z_k\rangle \le \langle g(w_k), z_{k+1} - w_k\rangle + \\ + \langle g(z_k), w_k - z_k\rangle + L_{k+1}(V(w_k, z_k) + V(z_{k+1}, w_k)),$$

следовательно

$$\frac{1}{L_{k+1}}(\langle g(w_k), w_k - z_*\rangle - \mu V(z_*, w_k)) \le \\ \le V(z_*, z_k) - \left(1 + \frac{\mu}{L_{k+1}}\right) V(z_*, z_{k+1}).$$

Таким образом,

$$V(z_*, z_k) \ge \left(1 + \frac{\mu}{L_{k+1}}\right) V(z_*, z_{k+1}),$$

или

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{L_{k+1}}\right)^{-1} V(z_*, z_k).$$

Далее, по рекурсии

$$V(z_*, z_{k+1}) \le \prod_{i=0}^{k}\left(1 + \frac{\mu}{L_{i+1}}\right)^{-1} V(z_*, z_0).$$

Учитывая, что $L_i \le 2L$, $\forall i$, получаем

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{2L}\right)^{-(k+1)} V(z_*, z_0).$$

□

**Замечание 1.** *Предложенный подход позволяет настраиваться на подходящие локальные параметры $L_{k+1}$ ($k \ge 0$) на итерациях алгоритма* 1. *Однако при этом возможно увеличение числа обращений к вспомогательной подпрограмме п.* 3 *листинга. Тем не менее можно показать, что такое увеличение некритично. Действительно, ясно, что при условии $L_0 \le 2L$ имеем $L_{k+1} \le 2L$. Тогда $L_{k+1} = 2^{i_k-2}L_k$ (где $i_k$ — количество вспомогательных шагов п. 3 листинга алгоритма* 1 *на $k$-й итерации). Далее, мы получаем $\frac{L_{k+1}}{L_k} = 2^{i_k-2}$, откуда*

$$\frac{L_N}{L_0} = \frac{L_N}{L_{N-1}}\frac{L_{N-1}}{L_{N-2}}\cdots\frac{L_1}{L_0} = 2^{i_0+i_1+\ldots+i_{N-1}-2N}. \tag{2.9}$$

*С другой стороны*,

$$\frac{2L}{L_0} \ge \frac{L_N}{L_0} = 2^{i_0+i_1+\ldots+i_{N-1}-2N}, \tag{2.10}$$

*откуда получаем*

$$i_0 + i_1 + \ldots + i_{N-1} \le 2N + \log_2\frac{2L}{L_0}. \tag{2.11}$$

*Это означает, что среднее количество обращений к п.* 3 *листинга алгоритма* 1 *сопоставимо с общим количеством итераций $N$. Если вместо $L_0 \le 2L$ можно допустить лишь $L_0 \le CL$ для некоторой $C > 0$, то это приведет лишь к изменению констант в рассуждениях выше, а принципиальный вывод сохранится*.

Для примера 1 замечание 2 означает, что адаптивность не может привести к резкому росту числа коммуникаций центрального узла сети с периферийными.

## 3. ОБОБЩЕНИЕ РЕЗУЛЬТАТА НА ЗАДАЧИ С δ-ОБОБЩеННЫМ УСЛОВИЕМ ГЛАДКОСТИ ОПЕРАТОРА ВАРИАЦИОННОГО НЕРАВЕНСТВА

В этом разделе мы детально обосноуем анонсированные в [3] теоретические результаты для двух вариантов алгоритма 1 в случае, когда оператор $g$ удовлетворяет следующему условию обобщенной гладкости:

$$\langle g(y) - g(z), x - z\rangle \le LV(x,z) + LV(z,y) + \delta \quad (3.1)$$

для некоторого фиксированного $\delta > 0$ при произвольных $x, y, z \in Q$. Параметр $\delta$ указывает на возможность применения рассматриваемых подходов на классе задач с обобщенным свойством относительной гладкости для вариационного неравенства. Например, такое обобщение позволяет применять рассматриваемую в этой статье методику к вариационным неравенствам с некоторыми типами нелипшициевых операторов и седловым задачам с некоторыми типами обобщенно гладких функционалов. Более подробно мотивирующие примеры приведены в статье [1] и докладе на конференции [3].

Рассмотрим следующий алгоритм 2 и докажем для него следующий результат, анонсированный в [3].

**Алгоритм 2.** Адаптивный метод для вариационных неравенств с $\delta$-обобщенно гладкими и относительно сильно монотонными операторами

**Require:** $x_0 \in Q$, $\delta > 0$, $L_0 > 0$, $\mu > 0$, $d(\cdot)$, $V(\cdot,\cdot)$.

1: $z_0 = \arg\min_{u\in Q} d(u)$.

2: **for** $k \geqslant 0$ **do**

3: Найти наименьшее целое $i_k \ge 0$ такое, что

$$\begin{aligned}&\langle g(z_k) - g(w_k), z_{k+1} - w_k\rangle \le \\ &\le L_{k+1}\left(V(w_k, z_k) + V(z_{k+1}, w_k)\right) + \delta,\end{aligned} \quad (3.2)$$

где $L_{k+1} = 2^{i_k - 1}L_k$, и

$$w_k = \arg\min_{y\in Q}\left\{\left\langle \frac{1}{L_{k+1}} g(z_k), y\right\rangle + V(y, z_k)\right\}, \quad (3.3)$$

$$\begin{aligned}z_{k+1} = \arg\min_{z\in Q}\Bigg\{&\left\langle \frac{1}{L_{k+1}} g(w_k), z\right\rangle + V(z, z_k) + \\ &+ \frac{\mu}{L_{k+1}} V(z, w_k)\Bigg\}.\end{aligned} \quad (3.4)$$

4: **end for**

**Ensure:** $z_k$.

**Теорема 2.** *Пусть $g$ — $\mu$-относительно сильно монотонный оператор и $z_*$ — точное решение вариационного неравенства* (1.1). *Тогда для алгоритма* 2 *справедливо следующее неравенство*:

$$\begin{aligned}V(z_*, z_{k+1}) \le &\prod_{i=1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1} V(z_*, z_0) + \\ &+ \frac{\delta}{L_{k+1} + \mu} + \sum_{j=1}^{k}\frac{\delta}{L_j + \mu}\prod_{i=j+1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1}.\end{aligned} \quad (3.5)$$

*Доказательство.* Для $w_k$ и $z_{k+1}$ при всяких $k \ge 0$ из (3.3) и (3.4) получаем, что

$$\begin{aligned}\frac{1}{L_{k+1}}\langle g(z_k), w_k - z_{k+1}\rangle &\le V(z_{k+1}, z_k) - \\ &- V(z_{k+1}, w_k) - V(w_k, z_k)\end{aligned} \quad (3.6)$$

и

$$\begin{aligned}\frac{1}{L_{k+1}}\langle g(w_k), z_{k+1} - z_*\rangle &\le V(z_*, z_k) - V(z_*, z_{k+1}) - \\ &- V(z_{k+1}, z_k) + \frac{\mu}{L_{k+1}}(V(z_*, w_k) - V(z_*, z_{k+1})).\end{aligned} \quad (3.7)$$

После сложения (3.6) и (3.7) получим

$$\begin{aligned}&\frac{1}{L_{k+1}}(\langle g(z_k), w_k - z_{k+1}\rangle + \langle g(w_k), z_{k+1} - z_*\rangle) \le \\ &\le -(V(w_k, z_k) + V(z_{k+1}, w_k)) + V(z_*, z_k) - \\ &- V(z_*, z_{k+1}) + \frac{\mu}{L_{k+1}}(V(z_*, w_k) - V(z_*, z_{k+1})).\end{aligned}$$

Используя неравенство (3.2), получим

$$\begin{aligned}&\langle g(z_k), z_{k+1} - z_k\rangle \le \langle g(w_k), z_{k+1} - w_k\rangle + \\ &+ \langle g(z_k), w_k - z_k\rangle + L_{k+1}(V(w_k, z_k) + V(z_{k+1}, w_k)) + \delta,\end{aligned}$$

следовательно

$$\begin{aligned}&\frac{1}{L_{k+1}}(\langle g(w_k), w_k - z_*\rangle - \mu V(z_*, w_k) - \delta) \le \\ &\le V(z_*, z_k) - \left(1 + \frac{\mu}{L_{k+1}}\right)V(z_*, z_{k+1}).\end{aligned}$$

Таким образом,

$$V(z_*, z_k) \ge \left(1 + \frac{\mu}{L_{k+1}}\right)V(z_*, z_{k+1}) - \frac{\delta}{L_{k+1}},$$

или

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{L_{k+1}}\right)^{-1} V(z_*, z_k) + \frac{\delta}{L_{k+1} + \mu}.$$

Далее, применяя рекурсию по $k$ в последнем неравенстве, приходим к требуемому неравенству

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{L_{k+1}}\right)^{-1} V(z_*, z_k) + \frac{\delta}{L_{k+1} + \mu} \le$$
$$\le \ldots \le \prod_{i=1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1} V(z_*, z_0) + \frac{\delta}{L_{k+1} + \mu} +$$
$$+ \sum_{j=1}^{k} \frac{\delta}{L_j + \mu} \prod_{i=j+1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1}.$$

□

Рассмотрим еще одну вариацию алгоритма 1 для вариационных неравенств с δ-обобщенно гладкими операторами. За счет модификации критерия выхода из итерации этот подход позволяет улучшить оценку качества выдаваемого методом решения в части влияния на него значения параметра $\delta > 0$.

**Алгоритм 3.** Вариант адаптивного метода для вариационных неравенств с δ-обобщенно гладкими и относительно сильно монотонными операторами

**Require:** $x_0 \in Q$, $\delta > 0$, $L_0 > 0$, $\mu > 0$, $d(\cdot)$, $V(\cdot,\cdot)$.

1: $z_0 = \arg\min_{u \in Q} d(u)$.

2: **for** $k \geqslant 0$ **do**

3: Найти наименьшее целое $i_k \ge 0$ такое, что

$$\langle g(z_k) - g(w_k), z_{k+1} - w_k \rangle \le \le L_{k+1}\left(V(w_k, z_k) + V(z_{k+1}, w_k)\right) + L_{k+1}\delta, \tag{3.8}$$

где $L_{k+1} = 2^{i_k - 1} L_k$, и

$$w_k = \arg\min_{y \in Q}\left\{\left\langle \frac{1}{L_{k+1}} g(z_k), y \right\rangle + V(y, z_k)\right\}, \tag{3.9}$$

$$z_{k+1} = \arg\min_{z \in Q}\left\{\left\langle \frac{1}{L_{k+1}} g(w_k), z \right\rangle + V(z, z_k) + \frac{\mu}{L_{k+1}} V(z, w_k)\right\}. \tag{3.10}$$

4: **end for**

**Ensure:** $z_k$.

Для алгоритма 3 докажем следующий результат, анонсированный в [3].

**Теорема 3.** *Пусть $g$ μ-относительно сильно монотонный оператор и $z_*$ — точное решение вариационного неравенства* (1.1). *Тогда для алгоритма* 3 *справедливо следующее неравенство*:

$$V(z_*, z_{k+1}) \le \prod_{i=1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1} V(z_*, z_0) + \delta\left(1 + \sum_{j=1}^{k}\prod_{i=j+1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1}\right). \tag{3.11}$$

*Если, кроме этого, $L_0 \le 2L$, то имеет место неравенство*

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{2L}\right)^{-(k+1)} V(z_*, z_0) + \left(1 + \frac{2L}{\mu}\right)\delta. \tag{3.12}$$

*Доказательство.* Для $w_k$ и $z_{k+1}$ при всяких $k \ge 0$ из (3.9) и (3.10) получаем, что

$$\frac{1}{L_{k+1}}\langle g(z_k), w_k - z_{k+1} \rangle \le V(z_{k+1}, z_k) - - V(z_{k+1}, w_k) - V(w_k, z_k), \tag{3.13}$$

и

$$\frac{1}{L_{k+1}}\langle g(w_k), z_{k+1} - z_* \rangle \le V(z_*, z_k) - V(z_*, z_{k+1}) - - V(z_{k+1}, z_k) + \frac{\mu}{L_{k+1}}(V(z_*, w_k) - V(z_*, z_{k+1})). \tag{3.14}$$

После сложения (3.13) и (3.14) получим

$$\frac{1}{L_{k+1}}(\langle g(z_k), w_k - z_{k+1} \rangle + \langle g(w_k), z_{k+1} - z_* \rangle) \le$$
$$\le -(V(w_k, z_k) + V(z_{k+1}, w_k)) + V(z_*, z_k) -$$
$$-V(z_*, z_{k+1}) + \frac{\mu}{L_{k+1}}(V(z_*, w_k) - V(z_*, z_{k+1})).$$

Используя неравенство (3.8), получим

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{L_{k+1}}\right)^{-1} V(z_*, z_k) + \frac{L_{k+1}\delta}{L_{k+1} + \mu},$$

или

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{L_{k+1}}\right)^{-1} V(z_*, z_k) + \delta,$$

откуда

$$V(z_*, z_{k+1}) \le \prod_{i=1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1} V(z_*, z_0) + \delta\left(1 + \sum_{j=1}^{k}\prod_{i=j+1}^{k+1}\left(1 + \frac{\mu}{L_i}\right)^{-1}\right).$$

Далее, если $L_0 \le 2L$, то для всякого натурального $i$ верно $L_i \le 2L$, получаем

$$V(z_*, z_{k+1}) \le \left(1 + \frac{\mu}{2L}\right)^{-(k+1)} V(z_*, z_0) + \left(1 + \frac{2L}{\mu}\right)\delta,$$

что и требовалось. □

## 4. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

В данном разделе для иллюстрации работоспособности предложенной выше методики приведем результаты вычислительных экспериментов по применению алгоритма 1 к различным задачам.

Программы для проведения экспериментов были реализованы на Python 3.4, все эксперименты проводились на компьютере с процессором Intel(R) Core(TM) i7-8550U CPU (1.80GHz, 4 ядра, 8 потоков). Оперативная память компьютера составляла 8 ГБ.

1) Первой была рассмотрена задача решения вариационного неравенства, связанная с параллелепипедно-симплексными играми и описанная в примере 2. Фигурирующие в постановке задачи данные были сгенерированы случайно следующим образом: $A = BB^\top$, $B_{ij}$ есть независимые случайные величины с равномерным распределением вероятностей на [0, 0.001], $b, c \in \mathbb{R}^n$ – с равномерным распределением вероятностей на [0, 1].

К задаче размерности $n = 200$ с парами коэффициентов $\mu_z = \mu_y = 10^{-2}$ и $\mu_z = 10^{-2}$, $\mu_y = 10^{-6}$ были применены алгоритм 1, соответствующий неадаптивный вариант [2] и обычный экстраградиентный метод. Начальная точка выбиралась случайным образом, была одинаковой для всех запусков и имела компоненты с равномерным распределением вероятностей на [0, 1].

Результаты сравнения алгоритмов показаны на рис. 1. В качестве метрики качества для построения кривых сходимости используется зазор, определенный как

$$\operatorname{Gap}(y,z) := \max_{z' \in \Delta_n} f(y,z') - \min_{y' \in [-1,1]^n} f(y',z)$$

и вычисляемый приблизительно с использованием реализации солвера SLSQP из библиотеки scipy. Из рис. 1 видно, что предлагаемый адаптивный алгоритм сходится заметно быстрее как своего неадаптивного варианта алгоритма 1 с постоянным шагом, так и классического экстраградиентного метода.

2) Далее, рассмотрена распределенная централизованная задача гребневой регрессии [7], имеющая структуру (1.9) для

$$f_j(x) := \frac{1}{2s}\|A_j x - b_j\|_2^2 + \lambda\|x\|_2^2, \quad j = 1,\ldots,m, \quad (4.1)$$

для некоторого $\lambda > 0$. Таким образом, каждый агент $j$ владеет в качестве локальных данных матрицей $A_j \in \mathbb{R}^{s\times n}$ и вектором $b_j \in \mathbb{R}^s$. Данные были сгенерированы случайно в соответствии со следующими законами: $A_j \in \mathbb{R}^{s\times n}$ для каждого $j = 1,\ldots,m$ имеет компоненты с экспоненциальным распределением вероятностей с коэффициентом масштаба 1 (в данном случае $\gamma > 1$) и со стандартным распределением Коши (тогда параметр схожести слагаемых $\gamma = 10^{-2}$ в (4.1)), а элементы $b_j \in \mathbb{R}^s$ – с равномерным распределением вероятностей на [0, 1]. В качестве множества $Q$ выберем евклидов

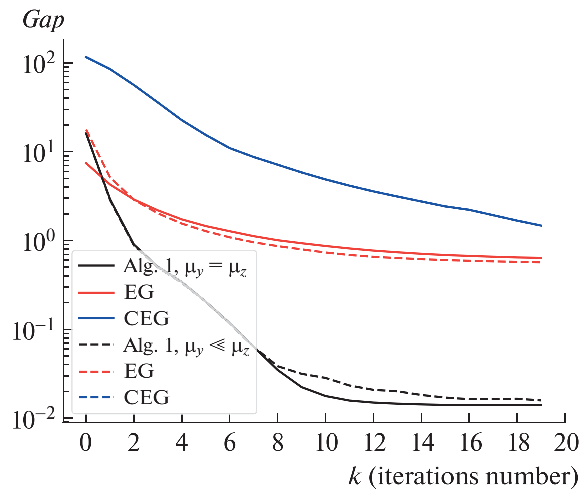


**Рис. 1.** Результаты эксперимента по сравнению с алгоритмом 1 (Alg. 1) с соответствующим неадаптивным (EG) и классическим экстраградиентным (CEG) методами для задачи (1.1), (1.8) с $\mu_y = \mu_z$ (сплошная линия) и $\mu_y \ll \mu_z$ (пунктир).

$\ell_2$-шар в $\mathbb{R}^n$ с центром в $0 \in \mathbb{R}^n$ и радиусом 1. Вслед за [6] выберем в качестве прокс-функции

$$d(x) = f_1(x) + \frac{\gamma}{2}\|x\|_2^2,$$

где $\gamma$ оценивает сверху спектральную норму разности $\nabla^2 f_1(x) - \nabla^2 F(x)$ (ее наибольшее собственное значение) [6].

Алгоритм 1, так же как соответствующий ему неадаптивный вариант с постоянным шагом и метод зеркального спуска [6], который имеет следующий вид:

$$x_{k+1} = \arg\min_{x\in Q}\left\{\langle \nabla F(x_k), x\rangle + V(x, x_k)\right\}, \quad (4.2)$$

в качестве начальной точки имел $x_0 = (1/\sqrt{n},\ldots,1/\sqrt{n})^\top \in \mathbb{R}^n$, а коэффициенты в задаче были зафиксированы в следующих значениях: $n = 50$, $s = m = 100$.

Результаты сравнения алгоритмов представлены на рис. 2. Построенные кривые сходимости отображают изменение значения целевой функции $F(x) = \frac{1}{m}\sum_{i=1}^m f_i(x)$ с увеличением числа итераций алгоритма 1. Из рис. 2 могут быть сделаны следующие наблюдения. Во-первых, среди подходов, основанных на сведении задачи распределенной оптимизации к задаче решения вариационного неравенства, адаптивный алгоритм 1 сходится за существенно меньшее число итераций, чем его неадаптивный вариант. Во-вторых, оба

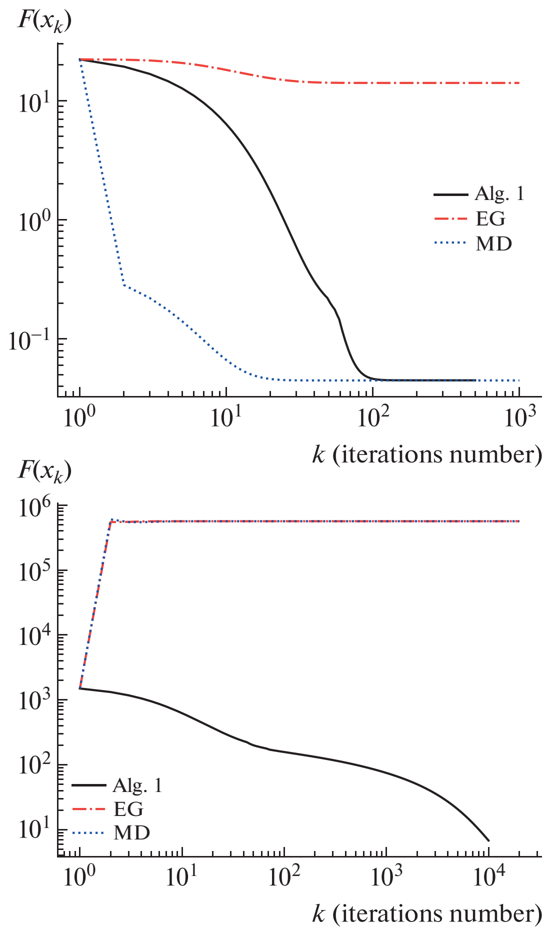


**Рис. 2.** Результаты эксперимента по сравнению алгоритма 1 (Alg. 1) с соответствующим неадаптивным методом (EG) и методом зеркального спуска (MD) для задачи (1.9), (4.1) с $\lambda = 10^{-1}$, $\gamma > 1$ (сверху) и $\lambda = 10^{-3}$, $\gamma = 10^{-2}$ (снизу).

соответствующих алгоритма уступают в эффективности подходу, решающему непосредственно исходную задачу оптимизации: можно видеть, что метод зеркального спуска наиболее эффективен среди подходов, примененных к рассмотренной задаче. Заметим также, что в наших экспериментах выбор параметра $\gamma$ меньший, чем то предписано теоретической оценкой схожести функционалов, приводил к хорошей сходимости именно исследуемого нами алгоритма 1 для вариационных неравенств. При этом оказалось, что наблюдалась расходимость соответствующего ему неадаптивного метода и метода зеркального спуска (4.2). Представляется, что отмеченные факты интересно исследовать дополнительно. На данный момент основной задачей была демонстрация преимуществ адаптивного подбора шага в алгоритме 1 по сравнению с известным ранее его неадаптивным вариантом [2].

## 5. ЗАКЛЮЧЕНИЕ

В работе исследован адаптивный метод (алгоритм 1) решения вариационных неравенств с относительно гладкими и относительно сильно монотонными операторами. Получена оценка скорости сходимости рассматриваемого метода с использованием адаптивно подбираемых параметров. В частности, это дает возможность применять подход и полученные результаты к достаточно популярному классу выпукло-вогнутых седловых задач вида ($Q_x$ и $Q_y$ — выпуклые замкнутые множества)

$$\min_{x \in Q_x} \max_{y \in Q_y} f(x, y)$$

с соответствующими предположениями о гладкости функционала $f$:

$$\|\nabla_x f(x_1, y) - \nabla_x f(x_2, y)\|_2 \le L_{xx}\|x_1 - x_2\|_2,$$
$$\|\nabla_x f(x, y_1) - \nabla_x f(x, y_2)\|_2 \le L_{xy}\|y_1 - y_2\|_2,$$
$$\|\nabla_y f(x_1, y) - \nabla_y f(x_2, y)\|_2 \le L_{xy}\|x_1 - x_2\|_2,$$
$$\|\nabla_y f(x, y_1) - \nabla_y f(x, y_2)\|_2 \le L_{yy}\|y_1 - y_2\|_2,$$

$\forall x, x_1, x_2 \in Q_x,\ y, y_1, y_2 \in Q_y$ и норма $\|\cdot\|_2$ евклидова. Алгоритм 1 позволяет реализовать адаптивную настройку на все параметры гладкости $L_{xx}, L_{xy}, L_{yy} > 0$.

При этом показано (замечание 1), что дополнительные процедуры адаптивного подбора параметров в исследованных выше алгоритмах не приводят к резкому росту сложности метода. Рассмотрено приложение к адаптивному методу для задач централизованной распределенной оптимизации (задача минимизации эмпирического риска из примера 1) со схожестью слагаемых.

Также исследована более общая ситуация, когда оператор вариационного неравенства удовлетворяет условию $\delta$-обобщенной гладкости [1] (в статье [1] рассматривался случай класса только монотонных операторов). В этой связи рассмотрен алгоритм 2, получен соответствующий результат о сходимости и описано влияние параметра $\delta$ на качество выдаваемого решения. Проведены численные эксперименты по реализации алгоритма 1 для решения следующих задач: распределенная централизованная задача гребневой регрессии со схожестью функций-слагаемых и задача решения вариационного неравенства, связанного с параллелепипедно-симплексными играми. Представляется, что выполненные эксперименты продемонстрировали достаточно хорошую эффективность алгоритма 1.

## ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа над разделом 2 выполнена при частичной поддержке программы стратегического академического лидерства “Приоритет-2030” (соглашение № 075-

02-2021-1316 от 30.09.2021). Работа над разделом 4 выполнена при финансовой поддержке гранта Президента Российской Федерации для государственной поддержки ведущих научных школ НШ775.2022.1.1.

## СПИСОК ЛИТЕРАТУРЫ


1. *Stonyakin F., Tyurin A., Gasnikov A., Dvurechensky P., Agafonov A., Dvinskikh D., Alkousa M., Pasechnyuk D., Artamonov S., Piskunova V.* Inexact Relative Smoothness and Strong Convexity for Optimization and Variational Inequalities by Inexact Model // Optim. Methods and Software. 2021. V. 36. № 6. P. 1155–1201.
2. *Cohen M.B., Sidford A., Tian K.* Relative Lipschitzness in Extragradient Methods and a Direct Recipe for Acceleration. arXiv preprint https://arxiv.org/pdf/2011.06572.pdf (2020).
3. *Titov A.A., Ablaev S.S., Stonyakin F.S., Alkousa M.S., Gasnikov A.* Some Adaptive First-Order Methods for Variational Inequalities with Relatively Strongly Monotone Operators and Generalized Smoothness. In: *Olenev N., Evtushenko Y., Jaćimović M., Khachay M., Malkova V., Pospelov I.* (eds) Optimization and Applications. OPTIMA 2022. Lecture Notes in Computer Science, vol 13781. Springer, Cham, 2022.
4. *Bauschke H.H., Bolte J., Teboulle M.* A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications // Mathematics of Operations Research. 2017. V. 42 (1.2). P. 330–348.
5. *Lu H., Freund R.M., Nesterov Y.* Relatively smooth convex optimization by first-order methods, and applications // SIAM Journal on Optimization. 2018. V. 28 (1.1). P. 333–354.
6. *Hendrikx H., Xiao L., Bubeck S., Bach F., Massoulie L.* Statistically preconditioned accelerated gradient method for distributed optimization. In International conference on machine learning, 4203–4227. PMLR, 2020.
7. *Tian Y., Scutari G., Cao T., Gasnikov A.* Acceleration in Distributed Optimization under Similarity. Proceedings of The 25th International Conference on Artificial Intelligence and Statistics // PMLR. 2022. V. 151. P. 5721–5756.
8. *Jin Y., Sidford A., Tian K.* Sharper Rates for Separable Minimax and Finite Sum Optimization via Primal-Dual Extragradient Methods. In Conference on Learning Theory, 4362–4415. PMLR, 2022.


# Adaptive Methods or Variational Inequalities with Relatively Smooth and Reletively Strongly Monotone Operators

**S. S. Ablaev[a,b], F. S. Stonyakin[a,b], M. S. Alkousa[a,c], and D. A. Pasechnyk[a,d]**

[a] *Moscow Institute of Physics and Technology*
*Institutskiy per., 9, Moscow region, Dolgoprudny, 141701 Russia*
[b] *Vernadsky Crimean Federal University Academician*
*Vernadsky Avenue, 4, Simferopol, 295007 Russia*
[c] *National Research University "Higher School of Economics"*
*Myasnitskaya st., 20, Moscow, 101000 Russia*
[d] *Trusted Artificial Intelligence Research Center of ISP RAS*
*Alexander Solzhenitsyn st., 25, Moscow, 109004 Russia*

The article is devoted to some adaptive methods for variational inequalities with relatively smooth and relatively strongly monotone operators. Based on the recently proposed proximal version of the extragradient method for this class of problems, we study in detail the method with adaptively selected parameter values. An estimate for the rate of convergence of this method is proved. The result is generalized to a class of variational inequalities with relatively strongly monotone δ-generalized smooth variational inequality operators. For the problem of ridge regression and variational inequality associated with box-simplex games, numerical experiments were performed demonstrating the effectiveness of the proposed method of adaptive selection of parameters during the running of the algorithm.