УДК 519.8

Адаптивный аналог метода Ю.Е. Нестерова для вариационных неравенств с сильно монотонным оператором*

Ф.С. Стонякин

Крымский федеральный университет им. В.И. Вернадского, пр. Акад. Вернадского, 4, Симферополь, 295007
E-mail: fedyor@mail.ru

Стонякин Ф.С. Адаптивный аналог метода Ю.Е. Нестерова для вариационных неравенств с сильно монотонным оператором // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2019. — Т. 22, № 2. — С. 201–211.

В работе предложен адаптивный аналог метода Ю.Е. Нестерова для вариационных неравенств с сильно монотонным оператором. Основная идея предлагаемого метода — адаптивный выбор констант в максимизируемых вогнутых функционалах на каждой итерации. При этом не требуется задания точного значения такой константы, поскольку предлагаемый метод позволяет найти подходящую константу на каждой итерации. Получены оценки для параметров, определяющих качество найденного решения вариационного неравенства в зависимости от числа итераций.

DOI: 10.15372/SJNM20190206

Ключевые слова: *вариационное неравенство, сильно монотонный оператор, адаптивный метод, условие Липшица, качество решения.*

Stonyakin F.S. An adaptive analog of Nesterov's method for variational inequalities with a strongly monotone operator // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2019. — Vol. 22, № 2. — P. 201–211.

An adaptive analog of the Nesterov method for variational inequalities with a strongly monotone operator is proposed. The main idea of the method proposed is the adaptive choice of constants in maximized concave functional at each iteration. In this case there is no need in specifying an exact value of this constant, because the method proposed makes possible to find a suitable constant at each iteration. Some estimates for the parameters determining the quality of the solution of the variational inequality depending on the number of iterations have been obtained.

Keywords: *variational inequality, strongly monotone operator, adaptive method, Lipschitz condition, solution quality.*

Вариационные неравенства (ВН) нередко возникают в самых разных проблемах оптимизации и имеют многочисленные приложения в математической экономике, математическом моделировании транспортных потоков, теории игр и других разделах математики (см., например, [1, 2]).

Наиболее известным аналогом градиентного метода для ВН является экстраградиентный метод Г.М. Корпелевич [3]. Одним из современных вариантов экстраградиентного метода является проксимальный зеркальный метод А.С. Немировского [4]. Недавно Ю.Е. Нестеровым в [5] предложен новый адаптивный метод решения задач выпуклой минимизации, который в случае липшицевости градиента целевой функции не требует

*Разработка и анализ алгоритма 2 проведены при поддержке Российского научного фонда (проект № 18-71-10044), доказательство следствия 1 и разработка алгоритма 3 выполнены при поддержке Российского научного фонда (проект № 18-71-00048).

знания никакой верхней оценки $\hat{L} \geqslant L$ для этой константы Липшица L. На базе идеологии [5] в [6, замечание 5.1, п. 5] предложен похожий адаптивный аналог проксимального зеркального метода А.С. Немировского для вариационных неравенств с оператором, удовлетворяющим условию Липшица.

Хорошо известно, что в задачах оптимизации замена условия выпуклости функционала на сильную выпуклость приводит к существенно лучшей скорости сходимости методов. Аналогичный эффект имеет место и для вариационных неравенств, если оператор обладает свойством сильной монотонности. В [9, 10] был предложен метод для вариационных неравенств с сильно монотонным липшицевым оператором. Этот метод представляет собой комбинацию двойственного экстраполяционного метода [7] и методики оценочных функций (см. [8, раздел 2.2]).

В предлагаемой работе мы, в некоторой степени отталкиваясь от идеологии [6, замечание 5.1], предложим адаптивный аналог метода Ю.Е. Нестерова для вариационных неравенств с липшицевым и сильно монотонным оператором, реализация которого не требует знания никакой верхней оценки $\hat{L} \geqslant L$ константы Липшица L оператора g.

Будем рассматривать задачу нахождения решения $x^* = x^*(Q)$ вариационного неравенства

$$\langle g(x^*), x^* - y\rangle \leqslant 0 \quad \forall y \in Q, \tag{1}$$

где $g : Q \to \mathbb{R}^n$ — сильно монотонный оператор с параметром $\mu > 0$:

$$\langle g(x) - g(y), x - y\rangle \geqslant \mu\|x - y\|^2 \quad \forall x, y \in Q, \tag{2}$$

Q — выпуклое замкнутое подмножество $\mathbb{R}^n$, $\langle \cdot, \cdot \rangle$ — скалярное произведение в $\mathbb{R}^n$,

$$\|x\| = \langle Bx, x\rangle^{1/2} \tag{3}$$

есть некоторая евклидова норма в $\mathbb{R}^n$, где $B : \mathbb{R}^n \to \mathbb{R}^n$ — фиксированный оператор $B = B^\top > 0$. Будем полагать, что оператор g удовлетворяет условию Липшица:

$$\|g(x) - g(y)\|_* \leqslant L\|x - y\| \quad \forall x, y \in Q \tag{4}$$

для некоторой константы $L > 0$, $\|s\|_* = \langle s, B^{-1}s\rangle^{1/2}$.

Напомним некоторые вспомогательные оценки, понятия и результаты из [10, п. 3.2]. Отметим, что сильная монотонность g означает, что для решения x^* верны оценки при произвольном $y \in Q$:

$$\langle g(y), x^* - y\rangle + \frac{\mu}{2}\|y - x^*\|^2 \leqslant \langle g(x^*), x^* - y\rangle - \frac{\mu}{2}\|y - x^*\|^2 \leqslant 0. \tag{5}$$

Неравенства (5) приводят к идее рассматривать следующую меру близости для оценки качества найденного приближённого решения x ВН (1):

$$f(x) = \sup_{y\in Q}\left\{\langle g(y), x - y\rangle + \frac{\mu}{2}\|y - x\|^2\right\}. \tag{6}$$

Отметим основные свойства f из (6).

Теорема 1 [10]. *Функция f из (6) определена и сильно выпукла на Q с параметром μ. Более того, для всякого $x \in Q$ $f(x) \geqslant 0$ и $f(x) = 0 \Leftrightarrow x = x^*$.*

Пусть в ходе работы некоторого алгоритма образовалась последовательность $\{y_i\}_{i=0}^N \subset Q$ и $\{\lambda_i\}_{i=0}^N$ — некоторый набор положительных чисел. Тогда обозначим

$$S_N = \sum_{i=0}^{N} \lambda_i \ \text{ и } \ \widetilde{y}_N := \frac{1}{S_N} \sum_{i=0}^{N} \lambda_i y_i \text{ — усреднённый выход работы алгоритма.} \tag{7}$$

Неравенства (5) приводят к идее ввести следующую функцию зазора для оценки качества найденного решения:

$$\Delta_N := \max_{x \in Q} \left\{ \sum_{i=0}^{N} \lambda_i \left[\langle g(y_i), y_i - x \rangle - \frac{\mu}{2} \|x - y_i\|^2 \right] \right\}. \tag{8}$$

Лемма 1 [10]. *Справедливо неравенство* $f(\widetilde{y}_N) \leqslant \frac{\Delta_N}{S_N}$.

Вслед за [10] обозначим

$$\varphi_y^{\beta}(x) := \langle g(y), y - x \rangle - \frac{\beta}{2} \|x - y\|^2, \quad \Phi_k(x) := \sum_{i=0}^{k} \lambda_i \varphi_{y_i}^{\mu}(x) \tag{9}$$

для произвольного параметра $\beta > 0$, $k = 0, 1, 2, \ldots$, а также $x, y \in Q$. Ясно, что функция φ_y^{β} сильно вогнута с параметром β, а $\Phi_k(x)$ сильно вогнута с параметром μS_k. Заметим, что при этом $(k = 0, 1, 2, \ldots, N)$

$$\Delta_k = \max_{x \in Q} \Phi_k(x). \tag{10}$$

Напомним метод Ю.Е. Нестерова для ВН с липшицевым сильно монотонным оператором [9, 10]. Опишем $(k+1)$-ю итерацию этого метода $(k = 0, 1, 2, \ldots)$.

Алгоритм 1. Метод для ВН с сильно монотонным оператором:

$$x_k := \arg\max_{x \in Q} \Phi_k(x), \quad y_{k+1} := \arg\max_{x \in Q} \varphi_{x_k}^{L}(x), \quad \lambda_{k+1} := \frac{\mu}{L} S_k.$$

Выход: $\widetilde{y}_{k+1} := \frac{1}{S_{k+1}} \sum_{i=0}^{k+1} \lambda_i y_i$.

Для приведённого выше метода (алгоритм 1) согласно [10, теорема 3.2.3] в случае липшицева оператора g с константой L для числа обусловленности $\gamma = \frac{L}{\mu}$ и произвольного натурального k верны оценки:

$$\begin{aligned} \frac{\mu}{2} \|\widetilde{y}_k - x^*\|^2 \leqslant f(\widetilde{y}_k) &\leqslant \left[f(y_0) + \frac{\mu(\gamma^2 - 1)}{2} \|y_0 - x^*\|^2 \right] \exp\left(-\frac{k}{\gamma + 1} \right) \\ &\leqslant f(y_0) \gamma^2 \exp\left(-\frac{k}{\gamma + 1} \right). \end{aligned} \tag{11}$$

Замечание 1. В [10, п. 3.2.2] было проведено сравнение алгоритма 1 со стандартным проекционным методом вида

$$x_0 = \overline{x} \in Q, \tag{12}$$

$$x_{k+1} = \pi_Q(x_k - \lambda B^{-1} g(x_k)), \quad k \geq 0, \tag{13}$$

где $\pi_Q(x)$ — евклидова проекция точки x на множество Q. Как отмечено в [10, п. 3.2.2], у этого метода может быть медленная сходимость. В частности, при выборе оптимального шага $\lambda = \frac{\mu}{L^2}$ верна оценка

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 \exp\left\{-\frac{k}{\gamma^2}\right\}. \tag{14}$$

При больших значениях числа обусловленности $\gamma = \frac{L}{\mu}$ эта оценка может быть значительно хуже, чем (11). Также известно, что скорость сходимости (11) не может быть улучшена никаким чёрноящичным методом, применяемым к задаче (1), (2) (см. [10, п. 3.2.2]). В то же время, с точки зрения сложности реализации метод Ю.Е. Нестерова не будет значительно сложнее метода (12), (13): на каждой итерации требуется вычислить две проекции на множество и два значения оператора вместо одной проекции и одного значения в методе (12), (13).

Теперь перейдём к основным результатам работы и предложим адаптивный аналог метода Ю.Е. Нестерова для ВН (1), (2). Положим изначально $\lambda_0 := 1$, y_0 — некоторое начальное приближение искомого решения и выберем некоторое $0 < \beta_0 \leqslant 2L$, где L — константа Липшица для оператора g из (4).

Замечание 2. Ввиду сильной монотонности оператора g для произвольных различных x и y из множества Q верно $g(x) \neq g(y)$. Поэтому выполнения условия $\beta_0 \leqslant 2L$ можно добиться, выбрав

$$\beta_0 := \frac{\|g(x) - g(y)\|_*}{\|x - y\|} \tag{15}$$

для некоторых фиксированных различных x и y из Q.

Опишем $(k+1)$-ю итерацию предлагаемого метода ($k = 0, 1, 2, \ldots$).

Алгоритм 2. Адаптивный метод для ВН с сильно монотонным оператором:

1. $x_k := \arg\max_{x \in Q} \Phi_k(x)$, $\beta_{k+1} := \frac{\beta_k}{2}$.
2. $y_{k+1} := \arg\max_{x \in Q} \varphi_{x_k}^{\beta_{k+1}}(x)$.
3. Если верно
$$\|g(y_{k+1}) - g(x_k)\|_* \leqslant \sqrt{\beta_{k+1}(\beta_{k+1} + \mu)}\, \|y_{k+1} - x_k\|, \tag{16}$$
то вычисляем $\lambda_{k+1} := \frac{\mu}{\beta_{k+1}} S_k$, увеличиваем k на 1 и переходим к следующей итерации (пункт 1 алгоритма).
Иначе $\beta_{k+1} := 2\beta_{k+1}$ и переходим к пункту 2 алгоритма.
Выход: $\widetilde{y}_{k+1} := \frac{1}{S_{k+1}} \sum_{i=0}^{k+1} \lambda_i y_i$.

Замечание 3. При $\beta_{k+1} \geqslant L$ критерий выхода из итерации (16) заведомо выполнен, так как $\sqrt{L(L + \mu)} > L$ при всяком $\mu > 0$. Поэтому после завершения итерации алгоритма 2 заведомо верно неравенство

$$\beta_{k+1} < 2L. \tag{17}$$

Таким образом, константа β_{k+1} не может неограниченно увеличиться, и максимальное её значение будет сопоставимо с L.

Замечание 4. Аналогично рассуждениям из [5, с. 391] оценим количество операций в пункте 2 алгоритма 2. Пусть на $(k+1)$-й итераций их было i_{k+1}. Тогда ввиду деления на 2 в пункте 1 алгоритма 2 мы имеем

$$\beta_{k+1} = \frac{1}{2} 2^{i_{k+1}-1}\beta_k = 2^{i_{k+1}-2}\beta_k, \tag{18}$$

откуда

$$\sum_{k=0}^{N-1} i_{k+1} = 2N + \sum_{k=0}^{N-1} \log_2 \frac{\beta_{k+1}}{\beta_k} < 2N + \log_2(2L) - \log_2(\beta_0). \tag{19}$$

Таким образом, за счёт повторения вычислений в пункте 2 алгоритма 2 сложность работы предлагаемого алгоритма 2 по сравнению с алгоритмом 1 может увеличиться не более чем в 2 раза с точностью до постоянного слагаемого, зависящего от β_0 и L. Это означает, что трудоёмкость предлагаемого метода вполне сопоставима с трудоёмкостью исходного алгоритма 1. Однако при этом не требуется знания никакой константы $\hat{L} \geqslant L$. Преимуществом также является возможное существенное увеличение скорости сходимости метода в конкретных задачах (см., например, таблицу 1).

Справедлива следующая

Теорема 2. *При выполнении алгоритма 2 для величин Δ_k из (8) верно неравенство $\Delta_{k+1} \leqslant \Delta_k$ для всякого целого неотрицательного k.*

Доказательство. Ясно, что $\Phi_{k+1}(x) = \Phi_k(x) + \lambda_{k+1}\varphi^{\mu}_{y_{k+1}}(x)$. Тогда

$$\begin{aligned}
\Delta_{k+1} &= \max_{x\in Q}\left\{\Phi_k(x) + \lambda_{k+1}\varphi^{\mu}_{y_{k+1}}(x)\right\}\\
&\leqslant \Delta_k + \max_{x\in Q}\left\{\langle\nabla\Phi_k(x_k), x - x_k\rangle - \frac{1}{2}\mu S_k\|x-x_k\|^2 + \lambda_{k+1}\varphi^{\mu}_{y_{k+1}}(x)\right\}\\
&\leqslant \Delta_k + \max_{x\in Q}\left\{-\frac{1}{2}\mu S_k\|x-x_k\|^2 + \lambda_{k+1}\left[\langle g(y_{k+1}), y_{k+1}-x\rangle - \frac{1}{2}\mu\|x-y_{k+1}\|^2\right]\right\}.
\end{aligned}$$

В силу выбора y_{k+1} из п. 2 алгоритма 2 для всякого $x \in Q$ имеем

$$\langle -g(x_k) - \beta_{k+1}B(y_{k+1}-x_k), x - y_{k+1}\rangle \leqslant 0.$$

Далее, с учётом равенства

$$2\langle B(y_{k+1}-x_k), x-y_{k+1}\rangle = \|x-x_k\|^2 - \|y_{k+1}-x_k\|^2 - \|x-y_{k+1}\|^2 \tag{20}$$

мы получаем оценки:

$$\begin{aligned}
&\langle\, g(y_{k+1}), y_{k+1}-x\rangle - \frac{\mu}{2}\|x-y_{k+1}\|^2\\
&= \langle g(y_{k+1}) - g(x_k), y_{k+1}-x\rangle - \frac{\mu}{2}\|x-y_{k+1}\|^2 + \langle g(x_k), y_{k+1}-x\rangle\\
&\leqslant \|g(y_{k+1}) - g(x_k)\|_*\|y_{k+1}-x\| - \frac{\mu}{2}\|x-y_{k+1}\|^2 + \beta_{k+1}\langle B(y_{k+1}-x_k), x-y_{k+1}\rangle\\
&= \|g(y_{k+1}) - g(x_k)\|_*\|y_{k+1}-x\| - \frac{\mu}{2}\|x-y_{k+1}\|^2 + \frac{\beta_{k+1}}{2}\|x-x_k\|^2 - \frac{\beta_{k+1}}{2}\|y_{k+1}-x_k\|^2 -\\
&\quad \frac{\beta_{k+1}}{2}\|x-y_{k+1}\|^2 = \|g(y_{k+1}) - g(x_k)\|_*\|y_{k+1}-x\| - \frac{\beta_{k+1}+\mu}{2}\|x-y_{k+1}\|^2 +\\
&\quad \frac{\beta_{k+1}}{2}\|x-x_k\|^2 - \frac{\beta_{k+1}}{2}\|y_{k+1}-x_k\|^2
\end{aligned}$$

$$\leqslant -\frac{1}{2\mu}\left(\|g(y_{k+1})-g(x_k)\|_* - \|y_{k+1}-x)\|\right)^2 +$$
$$\frac{\|g(y_{k+1})-g(x_k)\|_*^2}{2(\beta_{k+1}+\mu)} + \frac{\beta_{k+1}}{2}\left(\|x-x_k\|^2 - \|y_{k+1}-x_k\|^2\right)$$
$$\leqslant \frac{\|g(y_{k+1})-g(x_k)\|_*^2}{2(\beta_{k+1}+\mu)} + \frac{\beta_{k+1}}{2}\left(\|x-x_k\|^2 - \|y_{k+1}-x_k\|^2\right).$$

Поэтому ввиду (16)

$$\begin{aligned}
\langle g(y_{k+1}), y_{k+1}-x\rangle - \frac{\mu}{2}\|x-y_{k+1}\|^2 \\
&\leqslant \frac{1}{2(\beta_{k+1}+\mu)}\|g(y_{k+1})-g(x_k)\|_*^2 + \frac{\beta_{k+1}}{2}\|x-x_k\|^2 - \frac{\beta_{k+1}}{2}\|y_{k+1}-x_k\|^2 \\
&\leqslant \frac{\beta_{k+1}}{2}\|x-x_k\|^2,
\end{aligned}$$

откуда с учётом $\mu S_k = \lambda_{k+1}\beta_{k+1}$ получаем требуемое. □

Следствие 1. *При выполнении алгоритма* 2 *верно неравенство* $f(\widetilde{y}_k) \leqslant \Delta_0 \exp\left(-\frac{k\mu}{\mu+\hat{\beta}}\right)$ *для всякого натурального* k, *где* $\hat{\beta}$ *определяется следующим образом*:

$$1 - \frac{\mu}{\mu+\hat{\beta}} = \sqrt[k]{\left(1-\frac{\mu}{\mu+\beta_1}\right)\left(1-\frac{\mu}{\mu+\beta_2}\right)\dots\left(1-\frac{\mu}{\mu+\beta_k}\right)}. \tag{21}$$

Доказательство. Действительно, $S_0 = \lambda_0 = 1$ и для всякого $k = 0, 1, \dots$ верно

$$S_{k+1} = S_k + \lambda_{k+1} = \left(1 + \frac{\mu}{\beta_{k+1}}\right) S_k.$$

Далее, по лемме 1

$$\begin{aligned}
f(\widetilde{y}_k) &\leqslant \frac{\Delta_k}{S_k} \leqslant \Delta_0 \frac{S_0}{S_1}\frac{S_1}{S_2}\dots\frac{S_{k-1}}{S_k} \\
&= \Delta_0\left(1-\frac{\mu}{\mu+\beta_1}\right)\left(1-\frac{\mu}{\mu+\beta_2}\right)\dots\left(1-\frac{\mu}{\mu+\beta_k}\right) \\
&= \Delta_0\left(1-\frac{\mu}{\mu+\hat{\beta}}\right)^k = \Delta_0\left(1-\frac{1}{1+\frac{\hat{\beta}}{\mu}}\right)^k \leqslant \Delta_0 \exp\left(-\frac{k\mu}{\mu+\hat{\beta}}\right),
\end{aligned}$$

что и требовалось. □

Из теорем 1 и 2, а также следствия 1 вытекает следующий результат, аналогичный [10, теорема 3.2.3].

Теорема 3. *Пусть оператор* g *липшицев с константой* $L > 0$ *и сильно монотонен с параметром* $\mu > 0$. *Тогда при выполнении алгоритма* 2 *для* $\gamma = \frac{L}{\mu}$ *и всякого натурального* k *верны оценки*:

$$\begin{aligned}
\frac{\mu}{2}\|\widetilde{y}_k - x^*\|^2 \leqslant f(\widetilde{y}_k) &\leqslant \left[f(y_0) + \frac{\mu(\gamma^2-1)}{2}\|y_0 - x^*\|^2\right]\exp\left(-\frac{k}{1+\frac{\hat{\beta}}{\mu}}\right) \\
&\leqslant f(y_0)\gamma^2 \exp\left(-\frac{k}{1+\frac{\hat{\beta}}{\mu}}\right).
\end{aligned} \tag{22}$$

Отметим, что оценки (22) могут оказаться лучше (11), поскольку $\frac{\hat{\beta}}{\mu}$ может оказаться меньше γ. Далее это наглядно продемонстрировано на примере численного эксперимента для задачи (25).

Замечание 5. Рассмотрим модификацию алгоритма 2, которая исключает уменьшение константы β_{k+1} в ходе работы метода. Это даёт возможность сделать вывод о несущественном увеличении трудоёмкости по сравнению с методом Ю.Е. Нестерова (алгоритм 1). Изначально положим $\lambda_0 := 1$, y_0 — некоторое начальное приближение искомого решения и выберем некоторое $0 < \beta_0 \leqslant 2L$ (см. (15)), где L — константа Липшица для оператора g из (4). Опишем $(k+1)$-ю итерацию предлагаемой модификации алгоритма 2 $(k = 0, 1, 2, \ldots)$.

Алгоритм 3. Модификация алгоритма 2:

1. $x_k := \arg\max\limits_{x\in Q} \Phi_k(x),\ \beta_{k+1} := \beta_k$.
2. $y_{k+1} := \arg\max\limits_{x\in Q} \varphi_{x_k}^{\beta_{k+1}}(x)$.
3. Если верно
$$\|g(y_{k+1}) - g(x_k)\|_* \leqslant \sqrt{\beta_{k+1}(\beta_{k+1} + \mu)}\,\|y_{k+1} - x_k\|, \tag{23}$$
то вычисляем $\lambda_{k+1} := \frac{\mu}{\beta_{k+1}} S_k$, увеличиваем k на 1 и переходим к следующей итерации (пункт 1 алгоритма).
Иначе $\beta_{k+1} := 2\beta_{k+1}$ и переходим к пункту 2 алгоритма.
Выход: $\widetilde{y}_{k+1} := \frac{1}{S_{k+1}} \sum_{i=0}^{k+1} \lambda_i y_i$.

Поскольку β_{k+1} может лишь увеличиваться, то по сравнению с алгоритмом 1 количество вычислений, согласно пункту 2 алгоритма 3, возрастёт лишь не более чем на
$$\left\lceil \log_2 \frac{2L}{\beta_0} \right\rceil. \tag{24}$$

Для демонстрации преимуществ алгоритмов 2 и 3 по сравнению с алгоритмом 1 были проведены вычислительные эксперименты для вариационного неравенства с оператором $g : \mathbb{R}^{20} \to \mathbb{R}^{20}$ вида
$$g(x_1, x_2, \ldots, x_{20}) \\ = \left(\exp\left(x_1 + \frac{x_2}{10\exp(3)}\right), \exp\left(x_2 + \frac{x_3}{10\exp(3)}\right), \ldots, \exp\left(x_{20} + \frac{x_1}{10\exp(3)}\right)\right). \tag{25}$$

В качестве множества Q выберем единичный шар с центром в нуле:
$$Q = \left\{x = (x_1, x_2, \ldots, x_{20}) \mid x_1^2 + x_2^2 + \cdots + x_{20}^2 \leqslant 1\right\}$$
для стандартной евклидовой нормы в $\mathbb{R}^{20}$ (т.е. здесь оператор B в (3) мы полагаем тождественным)
$$\|x\| := \sqrt{x_1^2 + x_2^2 + \cdots + x_{20}^2}.$$

Пусть $x = (x_1, x_2, \ldots, x_{20})$ и $y = (y_1, y_2, \ldots, y_{20})$ — два вектора из Q. Очевидно, что оператор g не является потенциальным

$$\frac{\partial}{\partial x_2}\left(\exp\left(x_1+\frac{x_2}{10\exp(3)}\right)\right)=\frac{1}{10\exp(3)}\exp\left(x_1+\frac{x_2}{10\exp(3)}\right),$$

$$\frac{\partial}{\partial x_1}\left(\exp\left(x_2+\frac{x_3}{10\exp(3)}\right)\right)=0.$$

Покажем, что оператор g удовлетворяет условию Липшица и сильно монотонен на Q. По теореме о среднем для произвольных $i,j=\overline{1,20}$ имеем

$$\begin{aligned}
&\exp\left(x_i+\frac{x_j}{10\exp(3)}\right)-\exp\left(y_i+\frac{y_j}{10\exp(3)}\right)\\
&=\exp\left(x_i+\frac{x_j}{10\exp(3)}\right)-\exp\left(y_i+\frac{x_j}{10\exp(3)}\right)+\exp\left(y_i+\frac{x_j}{10\exp(3)}\right)-\exp\left(y_i+\frac{y_j}{10\exp(3)}\right)\\
&=\exp\left(\alpha_i+\frac{x_j}{10\exp(3)}\right)(x_i-y_i)+\frac{1}{10\exp(3)}\exp\left(y_i+\frac{\gamma_j}{10\exp(3)}\right)(x_j-y_j)
\end{aligned}\tag{26}$$

для некоторых α_i и γ_j : $|\alpha_i|\leqslant 1$ и $|\gamma_j|\leqslant 1$ (α_i и γ_j лежат между x_i,y_i и x_j,y_j соответственно). Ясно, что

$$\left|\alpha_i+\frac{x_j}{10\exp(3)}\right|\leqslant\sqrt{1+\left(\frac{1}{10\exp(3)}\right)^2}\sqrt{\alpha_i^2+x_j^2}<\sqrt{2},$$

а также $\left|y_i+\frac{\gamma_j}{10\exp(3)}\right|<\sqrt{2}$. Поэтому (26) означает, что

$$\begin{aligned}
&\left|\exp\left(x_i+\frac{x_j}{10\exp(3)}\right)-\exp\left(y_i+\frac{y_j}{10\exp(3)}\right)\right|\\
&\qquad<\exp(\sqrt{2})|x_i-y_i|+\exp(\sqrt{2}-3)\frac{|x_j-y_j|}{10}<\exp(\sqrt{2})\left(|x_i-y_i|+\frac{|x_j-y_j|}{10}\right).
\end{aligned}$$

Далее, с учётом неравенства $(a+b)^2\leqslant 2(a^2+b^2)$, имеем

$$\|g(x)-g(y)\|^2<2\exp(2\sqrt{2})\left(1+\frac{1}{100}\right)\|x-y\|^2\quad\forall\, x,y\in Q,$$

$$\|g(x)-g(y)\|<\frac{\sqrt{202}\exp(\sqrt{2})}{10}\|x-y\|\quad\forall\, x,y\in Q,$$

т. е. оператор g удовлетворяет свойству Липшица с константой $L=\frac{\sqrt{202}}{10}\exp(\sqrt{2})$.

Далее, (26) означает, что для произвольных $x,y\in Q$

$$\langle g(x)-g(y),x-y\rangle=\sum_{k=1}^{20}c_k(x_k-y_k)^2+\sum_{k=1}^{19}d_k(x_k-y_k)(x_{k+1}-y_{k+1})+d_{20}(x_{20}-y_{20})(x_{1-y_1}),$$

где

$$c_k>\exp(-\sqrt{2}),\quad d_k<\frac{\exp(\sqrt{2}-3)}{10}<\frac{\exp(-\sqrt{2})}{10}$$

для всякого $k=\overline{1,20}$. Учитывая неравенство $2ab\leqslant a^2+b^2$, получаем

$$\langle g(x)-g(y),x-y\rangle > \exp(-\sqrt{2})\sum_{k=1}^{20}(x_k-y_k)^2 - \frac{\exp(-\sqrt{2})}{10}\sum_{k=1}^{20}(x_k-y_k)^2$$
$$= \frac{9}{10}\exp(-\sqrt{2})\|x-y\|^2,$$

т. е. оператор g сильно монотонен с параметром $\mu = \frac{9}{10}\exp(-\sqrt{2})$.

Мы применили алгоритмы 1, 2 и 3 к вариационному неравенству для оператора g из (25) с параметрами $L = \frac{\sqrt{202}}{10}\exp(\sqrt{2})$, $\mu = \frac{9}{10}\exp(-\sqrt{2})$, начального приближения $y_0 = (0.2, 0.2, \dots, 0.2) \in Q$ и в соответствии с (15)

$$\beta_0 = \frac{\|g(1,0,0,\dots,0) - g(0,1,0,\dots,0)\|}{\sqrt{2}} \tag{27}$$

для стандартной евклидовой нормы в $\mathbb{R}^{20}$.

Результаты сравнения работы алгоритмов 1 и 2 (а также алгоритмов 1 и 3) представлены в сравнительных таблицах 1 и 2, где N — количество итераций работы этих алгоритмов, время работы алгоритмов указано в миллисекундах. Все вычисления были произведены с помощью CPython 3.6.4 на компьютере с 3-ядерным процессором AMD Athlon II X3 450 с тактовой частотой 803.5 МГц на каждое ядро. ОЗУ компьютера составляла 8 Гб.

Таблица 1. Сравнение результатов работы алгоритмов 1 и 2

N	$\exp\left(\frac{-k}{1+\frac{L}{\mu}}\right)$	Время, мс	$\exp\left(\frac{-k}{1+\frac{L}{\mu}}\right)$	Время, мс	β_N	$\hat{\beta}$
3	8.9742e−01	1	3.3880e−01	2	2.1447e−01	3.8766e−01
6	8.0536e−01	1	2.0270e−02	3	2.6809e−02	2.6809e−02
9	7.2274e−01	1	4.9199e−04	4	3.3512e−03	3.9726e−02
12	6.4860e−01	2	1.2773e−05	5	4.1889e−04	1.4210e−02
15	5.8207e−01	2	4.3275e−07	5	5.2362e−05	5.1801e−03
18	5.2236e−01	2	1.7770e−08	7	6.5452e−06	1.8911e−03
21	4.6878e−01	3	8.0981e−10	8	8.1815e−07	6.8756e−04
24	4.2069e−01	3	3.8794e−11	8	1.0227e−07	2.4877e−04
27	3.7753e−01	4	1.9004e−12	9	1.2784e−08	8.9622e−05
30	3.3881e−01	4	9.3990e−14	9	1.5980e−09	3.2176e−05
33	3.0405e−01	5	4.6670e−15	10	1.9974e−10	1.1521e−05
36	2.7286e−01	5	2.3211e−16	10	2.4968e−11	4.1168e−06
39	2.4487e−01	6	1.1551e−17	12	3.1210e−12	1.4687e−06
42	2.1975e−01	7	5.7501e−19	13	3.9013e−13	5.2327e−07
45	1.9721e−01	8	2.8626e−20	14	4.8766e−14	1.8625e−07

Как видим из таблицы 1, скорость сходимости для предлагаемого нами алгоритма 2 существенно выше скорости сходимости алгоритма 1. Это получается за счёт значительного уменьшения констант β_N на итерациях в ходе работы алгоритма, а также предлагаемого нами их усреднения в (21). Из табл. 2 видим, что скорость сходимости для предлагаемого нами алгоритма 3 выше, чем для алгоритма 1, но уже не так существенно, как для алгоритма 2. При этом время работы алгоритма 3 меньше, чем время работы алгоритма 2. По сути, преимущество алгоритма 3 перед алгоритмом 1 для рассматриваемого примера определяется, прежде всего, возможностью выбора начальной константы β_0 (27) согласно предлагаемому нами способу в замечании 2 без использования какой-либо оценки $\hat{L} \geqslant L$ константы Липшица L оператора g.

Таблица 2. Сравнение результатов работы алгоритмов 1 и 3

N	$\exp\left(\frac{-k}{1+\frac{L}{\mu}}\right)$	Время, мс	$\exp\left(\frac{-k}{1+\frac{L}{\mu}}\right)$	Время, мс	β_N	$\hat{\beta}$
3	8.9742e−01	1	7.1227e−01	2	1.7158e+00	1.7158e+00
6	8.0536e−01	1	5.0732e−01	3	1.7158e+00	1.7158e+00
9	7.2274e−01	1	3.6135e−01	4	1.7158e+00	1.7158e+00
12	6.4860e−01	2	2.5738e−01	5	1.7158e+00	1.7158e+00
15	5.8207e−01	2	1.8332e−01	5	1.7158e+00	1.7158e+00
18	5.2236e−01	2	1.3057e−01	7	1.7158e+00	1.7158e+00
21	4.6878e−01	3	9.3003e−02	8	1.7158e+00	1.7158e+00
24	4.2069e−01	3	6.6243e−02	8	1.7158e+00	1.7158e+00
27	3.7753e−01	4	4.7183e−02	9	1.7158e+00	1.7158e+00
30	3.3881e−01	4	3.3607e−02	9	1.7158e+00	1.7158e+00
33	3.0405e−01	5	2.3937e−02	10	1.7158e+00	1.7158e+00
36	2.7286e−01	5	1.7049e−02	10	1.7158e+00	1.7158e+00
39	2.4487e−01	6	1.2144e−02	12	1.7158e+00	1.7158e+00
42	2.1975e−01	7	8.6496e−03	13	1.7158e+00	1.7158e+00
45	1.9721e−01	8	6.1608e−03	14	1.7158e+00	1.7158e+00

Благодарности. Автор выражает огромную благодарность Александру Владимировичу Гасникову и Юрию Евгеньевичу Нестерову, а также неизвестным рецензентам за полезные обсуждения и комментарии.

Литература

1. **Facchinei F., Pang J.S.** Finite-Dimensional Variational Inequality and Complementarity Problems. Vol. 1 and 2.—New York: Springer-Verlag, 2003.

2. **Антипин А.С., Ячимович В., Ячимович М.** Динамика и вариационные неравенства // Журн. вычисл. матем. и мат. физики.—2017.—Т. 57, № 5.—С. 783–800; Перевод: Antipin A.S., Jacimovic V., Jacimovic M. Dynamics and variational inequalities // Comput. Maths. and Math. Phys.—2017.—Vol. 57, iss. 5.—P. 784–801.

3. **Корпелевич Г.М.** Экстраградиентный метод для отыскания седловых точек и других задач // Экономика и мат. методы.—1976.—Т. 12, № 4.—С. 747–756.

4. **Nemirovski A.** Prox-method with rate of convergence O(1/T) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems // SIAM J. on Optimization.—2004.—Vol. 15.—P. 229–251.

5. **Nesterov Yu.** Universal gradient methods for convex optimization problems // Math. Program. Series A and B archive.—2015.—Vol. 152, iss. 1–2.—P. 381–404.

6. **Гасников А.В.** Современные численные методы оптимизации. Метод универсального градиентного спуска.—М.: Изд-во МФТИ, 2018.

7. **Nesterov Yu.** Dual extrapolation and its application for solving variational inequalities and related problems // Math. Program. Ser. B.—2007.—Vol. 109, iss. 2–3.—P. 319–344.

8. **Нестеров Ю.Е.** Введение в выпуклую оптимизацию.—М.: МЦНМО, 2010.

9. **Nesterov Yu., Scrimali L.** Solving strongly monotone variational and quasi-variational inequalities // Discrete and Continuous Dynamical Systems.—2007.—31.10.2139/ssrn.970903.

10. **Нестеров Ю.Е.** Алгоритмическая выпуклая оптимизация: Дис. … докт. физ.-мат. наук: 01.01.07. — Москва, 2013.

Поступила в редакцию 17 января 2018 г.
После доработки 15 ноября 2018 г.
Принята к публикации 21 января 2019 г.

Литература в транслитерации

1. **Facchinei F., Pang J.S.** Finite-Dimensional Variational Inequality and Complementarity Problems. Vol. 1 and 2. — New York: Springer-Verlag, 2003.
2. **Antipin A.S., Yachimovich V., Yachimovich M.** Dinamika i variatsionnye neravenstva // Zhurn. vychisl. matem. i mat. fiziki. — 2017. — T. 57, № 5. — S. 783–800; Perevod: Antipin A.S., Jacimovic V., Jacimovic M. Dynamics and variational inequalities // Comput. Maths. and Math. Phys. — 2017. — Vol. 57, iss. 5. — P. 784–801.
3. **Korpelevich G.M.** Ekstragradientnyy metod dlya otyskaniya sedlovyh tochek i drugih zadach // Ekonomika i mat. metody. — 1976. — T. 12, № 4. — S. 747–756.
4. **Nemirovski A.** Prox-method with rate of convergence O(1/T) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems // SIAM J. on Optimization. — 2004. — Vol. 15. — P. 229–251.
5. **Nesterov Yu.** Universal gradient methods for convex optimization problems // Math. Program. Series A and B archive. — 2015. — Vol. 152, iss. 1–2. — P. 381–404.
6. **Gasnikov A.V.** Sovremennye chislennye metody optimizatsii. Metod universal'nogo gradientnogo spuska. — M.: Izd-vo MFTI, 2018.
7. **Nesterov Yu.** Dual extrapolation and its application for solving variational inequalities and related problems // Math. Program. Ser. B. — 2007. — Vol. 109, iss. 2–3. — P. 319–344.
8. **Nesterov Yu.E.** Vvedenie v vypukluyu optimizatsiyu. — M.: MTSNMO, 2010.
9. **Nesterov Yu., Scrimali L.** Solving strongly monotone variational and quasi-variational inequalities // Discrete and Continuous Dynamical Systems. — 2007. — 31.10.2139/ssrn.970903.
10. **Nesterov Yu.E.** Algoritmicheskaya vypuklaya optimizatsiya: Dis. … dokt. fiz.-mat. nauk: 01.01.07. — Moskva, 2013.