

УДК 004.8

## ОПТИМАЛЬНЫЙ АНАЛИЗ МЕТОДА С БАТЧИРОВАНИЕМ ДЛЯ СТОХАСТИЧЕСКИХ ВАРИАЦИОННЫХ НЕРАВЕНСТВ ВИДА КОНЕЧНОЙ СУММЫ

© 2023 г. А. Пичугин<sup>1</sup>, М. Печин<sup>1</sup>, А. Безносиков<sup>1,\*</sup>, А. Савченко<sup>2</sup>, А. Гасников<sup>1</sup>

Представлено академиком РАН А.А. Шананиным

Поступило 01.09.2023 г.

После доработки 15.09.2023 г.

Принято к публикации 18.10.2023 г.

Вариационные неравенства являются унифицированной оптимизационной постановкой, которая интересна не только сама по себе, но и потому что включает в себя задачи минимизации и поиска седловой точки. Между тем современные приложения побуждают рассматривать стохастические формулировки оптимизационных задач. В данной работе представлен анализ метода, имеющего оптимальные оценки сходимости для монотонных стохастических вариационных неравенств вида конечной суммы. В отличие от предыдущих работ, наш метод поддерживает батчирование и не теряет оптимальности оракульной сложности для любых размеров батча. Эффективность алгоритма, особенно в случае малых, но не единичных батчей, подтверждается численными экспериментами.

*Ключевые слова:* стохастическая оптимизация, вариационные неравенства, задачи вида конечной суммы, батчирование

**DOI:** 10.31857/S2686954323601598, **EDN:** BDNXBY

### 1. ВВЕДЕНИЕ

В данной работе рассматривается следующая задача вариационного неравенства:

$$\text{Найти } x^* \in \mathcal{X} \text{ такие, что } \langle F(x^*), x - x^* \rangle + g(x) - g(x^*) \geq 0, \quad \forall x \in \mathcal{X}, \quad (1)$$

где  $F$  – некоторый оператор,  $g$  – выпуклая непрерывная функция, определенная на  $\text{dom}g$ . Вариационные неравенства являются одной из популярных оптимизационных формулировок. Это связано с тем, что они включают в себя множество задач, широко встречающихся в различных областях прикладной науки. Чтобы понять важность вариационных неравенств, приведем два классических примера.

**Пример 1** (Минимизация). *Рассмотрим выпуклую задачу минимизации с регуляризатором:*

$$\min_{x \in \mathbb{R}^d} f(x) + g(x), \quad (2)$$

где  $f$  – целевая функция, а  $g$  – регуляризатор. Если взять  $F(x) := \nabla f(x)$ , то можно доказать, что

$x^* \in \text{dom}g$  является решением (1) тогда и только тогда, когда  $x^* \in \text{dom}g$  является решением (2) [1, Section 1.4.1].

**Пример 2** (Седловая задача). *Рассмотрим выпукло-вогнутую седловую задачу:*

$$\min_{x_1 \in \mathbb{R}^{d_1}} \max_{x_2 \in \mathbb{R}^{d_2}} f(x_1, x_2) + g_1(x_1) - g_2(x_2), \quad (3)$$

где  $g_1$  и  $g_2$  можно считать регуляризаторами. Определим  $F(x) := F(x_1, x_2) := [\nabla_{x_1} f(x_1, x_2), -\nabla_{x_2} f(x_1, x_2)]$  и  $g(x) = g(x_1, x_2) = g_1(x_1) + g_2(x_2)$ . Можно доказать, что  $x^* \in \text{dom}g$  является решением (1) тогда и только тогда, когда  $z^* \in \text{dom}g$  является решением (3) [1, Section 1.4.1]. Таким образом, выпукло-вогнутые седловые задачи (3) могут быть рассмотрены с помощью переформулировки в виде вариационного неравенства (1).

В то время как задачи минимизации обычно рассматриваются отдельно от вариационных неравенств, задачи поиска седловых точек очень часто изучаются именно в общности вариационных неравенств. Но даже не принимая во внимание задачи минимизации сами вариационные неравенства [2] и их частный случай – седловые задачи – имеют множество приложений в реальных приложениях. Тут, прежде всего, следует отметить классические примеры из экономики и теории игр, расцвет которых пришелся на середину

<sup>1</sup>Московский физико-технический институт, Долгопрудный, Россия

<sup>2</sup>Лаборатория искусственного интеллекта, ПАО “Сбербанк”, Москва, Россия

\*E-mail: beznosikov.an@phystech.edu

**Таблица 1.** Сравнение оракульных сложностей нахождения  $\varepsilon$ -решения монотонного стохастического вариационного неравенства вида конечной суммы с липшицевыми операторами (1) + (4). Сходимость измеряется с помощью зазора двойственности. *Обозначения:*  $\mu$  = константа сильной монотонности оператора  $F$ ,  $L$  = константы Липшица для всех  $F_i$ ,  $n$  = размер локального датасета,  $b$  = размер батча

| Источник                             | Сложность   | Дополнительные предположения  |
|--------------------------------------|---|---|
| Nemirovski et al. [7] <sup>(1)</sup> | $\mathcal{O}\left(n \frac{L}{\varepsilon}\right)$                 |   |
| Carmon et al. [31]                   | $\tilde{\mathcal{O}}\left(\sqrt{bn} \frac{L}{\varepsilon}\right)$ | $x \rightarrow \langle F(x) + \nabla g(x), x - u \rangle$ выпуклая для любой $u$ или ограниченной области |
| Alacaoglu and Malitsky [32]          | $\mathcal{O}\left(\sqrt{bn} \frac{L}{\varepsilon}\right)$         |   |
| Alacaoglu et al. [33]                | $\mathcal{O}\left(n \frac{L}{\varepsilon}\right)$                 |   |
| This paper                           | $\mathcal{O}\left(\sqrt{n} \frac{L}{\varepsilon}\right)$          |   |
| Han et al. [34]                      | $\Omega\left(\sqrt{n} \frac{L}{\varepsilon}\right)$               | нижняя оценка   |

Примечание. <sup>(1)</sup> детерминированные методы, аналогичные результаты также были получены в [27, 29].

прошлого века [1, 3, 4]. Важно отметить также классические, но более новые сюжеты из области робастной оптимизации [5–7] и обучения с учителем/без учителя [8–13]. Еще большую значимость вариационным неравенствам придают новые приложения из областей обучения с подкреплением [14, 15], состязательного обучения [16] и генеративных моделей [17–20].

Между тем с каждым годом современные прикладные задачи становятся все более сложными и имеют стохастический характер:  $F(x) = \mathbb{E}_{\xi \sim \mathcal{D}}[F_{\xi}(x)]$ . Часто такое математическое ожидание не может быть представлено в аналитической форме из-за того, что распределение  $\mathcal{D}$  является нетривиальным или даже неизвестным, поэтому прибегают к подходу Монте-Карло, который аппроксимирует интеграл с помощью выборки из распределения  $\mathcal{D}$ :

$$F(x) = \frac{1}{M} \sum_{m=1}^M F_m(x). \tag{4}$$

В алгоритмах, предназначенных для решения стохастических задач [21–24], часто избегают обращения к полным операторам или градиентам и прибегают к технике батчирования, когда выбирается одно или несколько слагаемых из (4). Поэтому очень важно, чтобы алгоритм был устойчив при использовании батчей любого размера, что подтверждается как с точки зрения теории, так и на практике. Но даже для задач минимизации не все стохастические методы обладают данным свойством [25, Раздел 5]. Это подводит нас к цели данной работы:

Разработать алгоритм решения монотонного стохастического вариационного неравенства ви-

да конченной суммы с липшицевыми слагаемыми (1) + (4). При этом алгоритм должен быть нечувствителен к размеру батчей.

**Краткий обзор существующих результатов.** Теория решения монотонных вариационных неравенств с липшицевым оператором имеет обширную историю. Естественной идеей для создания методов для решения вариационных неравенств является обобщение градиентного спуска с помощью замены градиента  $\nabla f(x)$  на оператор  $F(x)$ . Однако для монотонных операторов такой метод может расходиться. Прорывом стало создание экстраградиентного метода [26], который был широко теоретически исследован [7, 27] и впоследствии получил различные модификации и обобщения [28, 29], в том числе и для стохастического случая [30]. Что касается методов решения монотонных стохастических задач вида конечной суммы, то можно выделить работы [31–33], в которых авторы адаптируют технику редукции дисперсии для вариационных неравенств. Более того, результаты из [32] являются оптимальными [34]. Но это справедливо только для случая, когда на каждой итерации мы собираем батч размера один. Как только мы используем неединичный батч, результаты работ [31, 32] становятся хуже. Подробнее см. табл. 1.

В данной статье предложено решение этой проблемы и представлен метод, оптимальный для любого размера батча. Идея основана на оптимистической схеме с моментумом из статьи [35]. Однако в указанной работе авторами были рассмотрены только сильно монотонные вариационные неравенства, тогда как в данной статье рассматриваются более общие и более сложные с точки

зрения теоретического анализа монотонные вариационные неравенства.

## 2. ПОСТАНОВКА ЗАДАЧИ И ПРЕДПОЛОЖЕНИЯ

В работе рассматривается задача (1) + (4), где  $\mathcal{X}$  – конечномерное евклидово пространство со скалярным произведением  $\langle \cdot, \cdot \rangle$  и индуцированной нормой  $\|\cdot\|$ . Для выпуклой полунепрерывной  $g$  обозначим ее область определения как  $\text{dom}g = \{x : g(x) < +\infty\}$ . Предполагаем, что функция  $g$  является такой, что оператор  $\text{prox}(x) = \arg \min_{y \in \mathcal{X}} \left\{ \alpha g(y) + \frac{1}{2} \|y - x\|^2 \right\}$  при любом  $\alpha > 0$  может быть точно вычислен “бесплатно”. Отметим важное свойство  $\text{prox}$ , а именно для  $y = \text{prox}_{\alpha g}(x)$  выполняется

$$x - y \in \partial(\alpha g)(y). \quad (5)$$

Через  $\mathbb{E}_k$  обозначим условное матожидание  $\mathbb{E}[\cdot | S^{k-1}, S^{k-2}, \dots, S^0]$ .

Каждый оператор  $F_i : \text{dom}g \rightarrow \mathbb{R}^d$  является однозначным. Основные предположения работы формулируются следующим образом.

Предположение 1.

- Решение задачи (1) + (4) существует и может быть не единственным.

- Оператор  $F$  монотонный, т.е. для любых  $u, v \in \mathbb{R}^d$  выполняется

$$\langle F(u) - F(v); u - v \rangle \geq 0.$$

- Оператор  $F$   $L$ -липшицевый, т.е. для любых  $u, v \in \mathbb{R}^d$  выполняется

$$\|F(u) - F(v)\| \leq L \|x - y\|.$$

- Оператор  $F_m$  является  $L_m$ -липшицевым.

Определим  $\bar{L}$ , как  $\bar{L}^2 = \frac{1}{M} \sum_{m=1}^M L_m^2$ .

## 3. ОСНОВНАЯ ЧАСТЬ

В основу описанного в данной статье метода положена нераспределенная версия Алгоритма 1

из статьи [35]. Данный алгоритм основан на так называемой “оптимистической” модификации [28] метода экстраградиента [26]. Суть этой модификации заключается в использовании значения оператора не только в точках текущей итерации ( $x^k$  и  $w^k$ ), но и с прошлой итерации ( $x^{k-1}$  и  $w^{k-1}$ ). Также в методе используется техника редукции дисперсии [22, 32], для чего введена дополнительную последовательность  $w^k$ . Отметим, что точка  $w^k$  (опорная точка) меняется редко (если  $p$  мало), и поэтому полный оператор  $F(w^k)$  практически не пересчитывается. Также в данной работе применен случайный отрицательный момент  $\gamma(w^k - x^k)$ , который необходим для применения техники редукции дисперсии в методах вариационных неравенств [32, 33].

В качестве критерия сходимости используем следующую функцию:

$$\text{Gap}(z) = \sup_{u \in \mathcal{C}} [\langle F(u), z - u \rangle + g(x) - g(u)]. \quad (6)$$

Здесь мы берем максимум не на всем множестве  $\mathcal{X}$ , а на  $\mathcal{C}$  – компактном подмножестве  $\mathcal{X}$ . Таким образом, мы можем рассматривать и неограниченные множества  $\mathcal{X} \subseteq \mathbb{R}^d$ . Это допустимо, поскольку рассмотренная версия критерия справедлива, если решение  $x^*$  лежит в  $\mathcal{C}$ ; подробнее см. [6]. Следующая теорема представляет сходимость предложенного метода.

**Теорема 1.** *Рассмотрим задачу (1) + (4) в Предположениях 1. Пусть  $\{x^k\}$  – последовательность, генерируемая Алгоритмом 3 с выбранными  $\eta, \theta, \alpha, \beta, \gamma$  следующим образом:*

$$0 < p = \gamma \leq \frac{1}{16}, \quad \eta = \min \left\{ \frac{\sqrt{\gamma b}}{8\bar{L}}, \frac{1}{8\bar{L}} \right\}.$$

*Тогда, при  $\varepsilon > 0$ , количество итераций для достижения  $\mathbb{E}[\text{Gap}(x^k)] \leq \varepsilon$  равно*

$$\mathcal{O} \left( \frac{1}{\sqrt{pb}} \frac{\bar{L}}{\varepsilon} + \frac{L}{\varepsilon} \right).$$

---

### Algorithm 1. Оптимистичный Метод с Моментом и Батчированием

---

1: **Параметры:** размер шага  $\eta > 0$ , момент  $\gamma > 0$ , вероятность  $p \in (0; 1)$ , размер батча  $b \in \{1, \dots, M\}$ , количество итераций  $K$

2: **Инициализация:** выбрать  $x^0 = w^0 = x^{-1} = w^{-1} \in \mathbb{R}^d$

3: **for**  $k = 0, 1, \dots$ , **do**

4:   Выбрать  $j_1^k, \dots, j_b^k$  случайным образом из  $\{1, \dots, M\}$  независимо и равномерно

- 5:  $S^k = \{J_1^k, \dots, J_b^k\}$
- 6:  $\Delta^k = \frac{1}{b} \sum_{j \in S^k} (F_j(x^k) - F_j(w^{k-1}) + (F_j(x^k) - F_j(x^{k-1}))) + F(w^{k-1})$
- 7:  $x^{k+1} = \text{прох}_{\eta_g}(x^k + \gamma(w^k - x^k) - \eta \Delta^k)$
- 8:  $w^{k+1} = \begin{cases} x^{k+1}, & \text{с вероятностью } p \\ w^k, & \text{с вероятностью } 1 - p \end{cases}$
- 9: **end for**

Доказательство см. в Приложении А. Отметим, что Теорема дает итерационную сложность Алгоритма 3, но интерес представляет, прежде всего, оракульная сложность (количество обращений к слагаемым  $F_i$ ). Можно заметить, что на каждой итерации в среднем вызывается  $\mathcal{O}(b + pn)$  слагаемых – каждый раз вызывается батч размера  $b$ , и с вероятностью  $p$  вычисляется полный оператор. Отсюда сразу получаем оптимальный выбор для  $p$ :

**Следствие 1.** В условиях теоремы 1 при выборе  $p = \frac{b}{n}$  получается следующая оракульная сложность для Алгоритма 1

$$\mathcal{O}\left(\sqrt{n} \frac{\bar{L}}{\varepsilon} + b \frac{L}{\varepsilon}\right).$$

Если  $b \leq \frac{\bar{L}\sqrt{n}}{L}$ , то получаем  $\mathcal{O}\left(\sqrt{n} \frac{\bar{L}}{\varepsilon}\right)$ . Этот результат является оптимальным и неулучшаемым [34].

#### 4. ЭКСПЕРИМЕНТЫ

В данном разделе проверяется эффективность работы алгоритма 1 на практике.

Рассмотрим билинейную задачу:

$$\min_{x \in \Delta^d} \max_{y \in \Delta^d} x^T A y, \tag{7}$$

где  $\Delta^d$  – единичный симплекс в  $\mathbb{R}^d$ . В данной работе используются те же экспериментальные наборы входных данных, что и в [32], в частности, рассматриваются матрицы игры “вор и полицейский” из [36] и первая тестовая матрица из [37]. Заметим, что задача (7) не имеет вид конечной суммы, как (1) + (4), но можно переписать  $A$  из (7)

следующим образом:  $A = \sum_{i=1}^d A_i$  или  $A = \sum_{i=1}^d A_i$ ,

где  $A_i$  –  $i$ -я строка  $A$ , а  $A_i$  –  $i$ -й столбец  $A$  – см. подробнее в Разделе 5.1.2 из [32].

Для сравнения взяты методы из табл. 1. В частности, выбраны Алгоритмы 1 и 2 из [32], Алгоритм 1 + 2 из [31]. Все алгоритмы рассматривают-

ся в евклидовой постановке с проекцией на симплекс из [38].

Параметры всех методов настраиваются, исходя из теоретических результатов, соответствующих статей – см. Раздел 6 из [32]. Все методы проверяются с различными размерами батчей. В качестве критерия сходимости используется зазор двойственности (6), который вычисляется как  $[\max_i (A^T x)_i - \min_j (A y)_j]$  ввиду решения задачи на симплексе. Критерием сравнения является количество операций (одна операция вычислительно равна вычислениям  $Ay$  и  $A^T x$ ).

Результаты отражены на рис. 1 и 2. Они показывают, что алгоритм 3 превосходит всех конкурентов.

#### 5. ВЫВОДЫ

В статье представлен алгоритм решения монотонных стохастических вариационных неравенств вида конечной суммы с липшицевыми слагаемыми. С теоретической точки зрения алгоритм является оптимальным практически для любого размера батча. В качестве направлений дальнейшей работы авторов интересует получение версии Алгоритма 3 в неевклидовом случае с произвольной дивергенцией Брэгмана.

#### Приложение

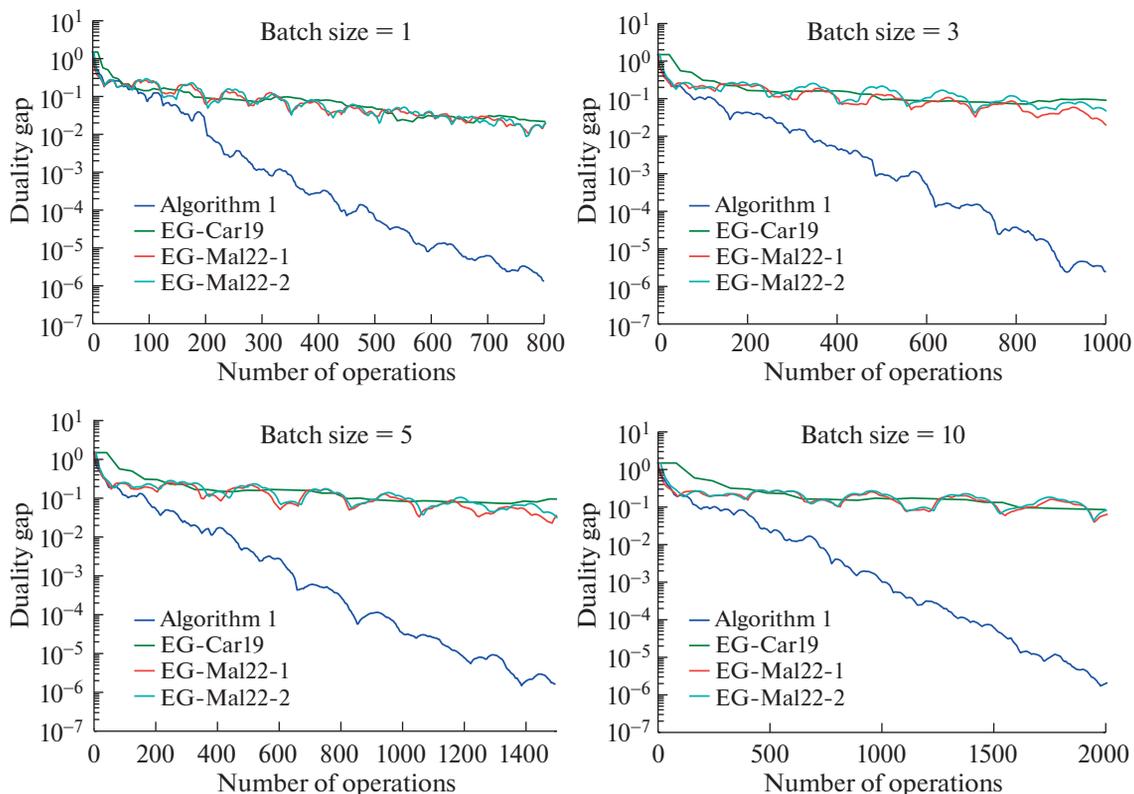
##### А. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1

Перед доказательством теоремы введем следующие леммы.

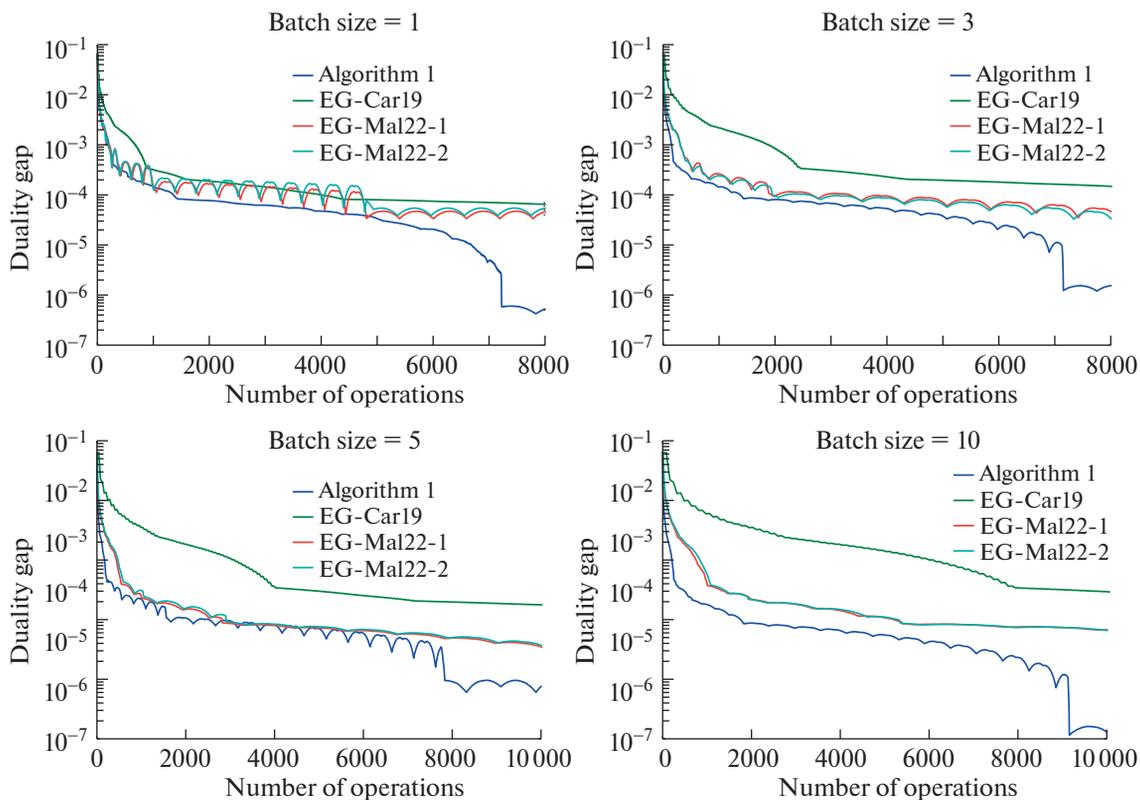
**Лемма 1** [Лемма 2.4 из [32]]. Пусть  $\mathcal{F} = (\mathcal{F}_k)_{k \geq 0}$  – фильтрация, а  $(u^k)$  – стохастический процесс, адаптированный к  $\mathcal{F}$  с  $\mathbb{E}[u^{k+1} | \mathcal{F}_k] = 0$ . Тогда для любого  $K \in \mathbb{N}$ ,  $x^0$  в  $X$  и любого компактного множества  $\mathcal{C} \subseteq X$  справедливо

$$\mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} \langle u^{k+1}, x \rangle \right] \leq \max_{x \in \mathcal{C}} \frac{1}{2} \|x^0 - x\|^2 + \frac{1}{2} \sum_{k=0}^{K-1} \mathbb{E} \|u^{k+1}\|^2.$$

**Лемма 2.** В условиях Предположения 1 для итераций Алгоритма 1 имеет место следующая оценка:



**Рис. 1.** Сравнение вычислительных сложностей для Алгоритма 3, EG-Mal22-1 (Алгоритм 1 [32]), EG-Mal22-2 (Алгоритм 2 [32]), EG-Car19 (Алгоритм 1+2 из [31]) на (7) с матрицей полицейского и грабителя из [36].



**Рис. 2.** Сравнение вычислительных сложностей для алгоритма 3, EG-Mal22-1 (Алгоритм 1 [32]), EG-Mal22-2 (Алгоритм 2 [32]), EG-Car19 (Алгоритм 1+2 из [31]) на (7) с первой тестовой матрицей из [37].

$$\begin{aligned} & \mathbb{E} \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] \leq \\ & \leq \frac{2\bar{L}^2}{b} \mathbb{E} \left[ \left\| x^k - w^{k-1} \right\|^2 + \left\| x^k - x^{k-1} \right\|^2 \right], \end{aligned} \quad (8)$$

где  $\mathbb{E}_k [\Delta^k]$  равно

$$\mathbb{E}_k [\delta_k] = 2F(x^k) - F(x^{k-1}). \quad (9)$$

*Доказательство.* Начнем со строки 6 Алгоритма 1.

$$\begin{aligned} & \mathbb{E}_k \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] = \\ & = \mathbb{E}_k \left[ \left\| \frac{1}{b} \sum_{j \in S^k} (F_j(x^k) - F_j(w^{k-1}) + (F_j(x^k) - F_j(x^{k-1}))) + \right. \right. \\ & \quad \left. \left. + F(w^{k-1}) - (2F(x^k) - F(x^{k-1})) \right\|^2 \right]. \end{aligned}$$

Из неравенства Коши-Шварца имеем

$$\begin{aligned} & \mathbb{E}_k \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] \leq \\ & \leq 2\mathbb{E}_k \left[ \left\| \frac{1}{b} \sum_{j \in S^k} (F_j(x^k) - F_j(w^{k-1})) - (F(x^k) - F(w^{k-1})) \right\|^2 \right] + \\ & + 2\mathbb{E}_k \left[ \left\| \frac{1}{b} \sum_{j \in S^k} (F_j(x^k) - F_j(x^{k-1})) - (F(x^k) - F(x^{k-1})) \right\|^2 \right]. \end{aligned}$$

Выбирая  $j_1^k, \dots, j_b^k$  в  $S^k$  независимо и равномерно, можно заметить, что

$$\begin{aligned} & \mathbb{E}_k \left[ \left\langle (F_j(x^k) - F_j(w^{k-1})) - (F(x^k) - F(w^{k-1})), \right. \right. \\ & \quad \left. \left. (F_j(x^k) - F_j(w^{k-1})) - (F(x^k) - F(w^{k-1})) \right\rangle \right] = \\ & = \mathbb{E}_k \left[ \left\langle \mathbb{E}_{j_1^k} [(F_{j_1^k}(x^k) - F_{j_1^k}(w^{k-1})) - (F(x^k) - F(w^{k-1}))], \right. \right. \\ & \quad \left. \left. \mathbb{E}_{j_1^k} [(F_{j_1^k}(x^k) - F_{j_1^k}(w^{k-1})) - (F(x^k) - F(w^{k-1}))] \right\rangle \right] = 0. \end{aligned}$$

Следовательно, получаем

$$\begin{aligned} & \mathbb{E}_k \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] \leq 2\mathbb{E}_k \left[ \sum_{j \in S^k} \frac{1}{b} \left\| (F_j(x^k) - \right. \right. \\ & \quad \left. \left. - F_j(w^{k-1})) - (F(x^k) - F(w^{k-1})) \right\|^2 \right] + \\ & + 2\mathbb{E}_k \left[ \sum_{j \in S^k} \frac{1}{b} \left\| (F_j(x^k) - F_j(x^{k-1})) - \right. \right. \\ & \quad \left. \left. - (F(x^k) - F(x^{k-1})) \right\|^2 \right] = \\ & = \frac{2}{b^2} \mathbb{E}_k \left[ \sum_{j \in S^k} \left\| (F_j(x^k) - F_j(w^{k-1})) - (F(x^k) - F(w^{k-1})) \right\|^2 \right] + \end{aligned}$$

$$\begin{aligned} & + \frac{2}{b^2} \mathbb{E}_k \left[ \sum_{j \in S^k} \left\| (F_j(x^k) - F_j(x^{k-1})) - (F(x^k) - F(x^{k-1})) \right\|^2 \right] \leq \\ & \leq \frac{2}{b^2} \mathbb{E}_k \left[ \sum_{j \in S^k} \left\| F_j(x^k) - F_j(w^{k-1}) \right\|^2 \right] + \\ & + \frac{2}{b^2} \mathbb{E}_k \left[ \sum_{j \in S^k} \left\| F_j(x^k) - F_j(x^{k-1}) \right\|^2 \right]. \end{aligned}$$

В последнем переходе был использован факт, что  $\mathbb{E} \|X - \mathbb{E}X\|^2 = \mathbb{E} \|X\|^2 - \|\mathbb{E}X\|^2$ . Далее снова учитываем, что  $j_1^k, \dots, j_b^k$  в  $S^k$  выбираются равномерно

$$\begin{aligned} & \mathbb{E}_k \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] \leq \\ & \leq \frac{2}{b} \mathbb{E}_k \left[ \mathbb{E}_{j \sim \text{u.a.r.} \{1, \dots, M\}} \left[ \left\| F_j(x^k) - F_j(w^{k-1}) \right\|^2 + \right. \right. \\ & \quad \left. \left. + \left\| F_j(x^k) - F_j(x^{k-1}) \right\|^2 \right] \right] = \\ & = \frac{2}{Mb} \sum_{j=1}^M \left( \left\| F_j(x^k) - F_j(w^{k-1}) \right\|^2 + \right. \\ & \quad \left. + \left\| F_j(x^k) - F_j(x^{k-1}) \right\|^2 \right). \end{aligned}$$

Поскольку каждый оператор  $F_j$  является  $L_j$ -липшицевым, можно переписать его в виде

$$\begin{aligned} & \mathbb{E}_k \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] \leq \\ & \leq \frac{2}{mb} \sum_{j=1}^m L_j^2 \left( \|x^k - w^{k-1}\|^2 + \|x^k - x^{k-1}\|^2 \right). \end{aligned}$$

Применяя определение  $\bar{L}$ , получаем

$$\begin{aligned} & \mathbb{E}_k \left[ \left\| \Delta^k - \mathbb{E}_k [\Delta^k] \right\|^2 \right] \leq \\ & \leq \frac{2\bar{L}^2}{b} \left( \|x^k - w^{k-1}\|^2 + \|x^k - x^{k-1}\|^2 \right). \end{aligned}$$

Взятие полного математического ожидания завершает доказательство.  $\square$

**Лемма 3.** Для итераций Алгоритма 1 с  $\gamma = \rho$  на любом компакте  $\mathcal{C} \subseteq X$  выполняется следующая оценка:

$$\begin{aligned} & \mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} e_1(x, k) \right] \leq 2 \max_{x \in \mathcal{C}} \|x - x^0\|^2 + \\ & + \frac{\gamma(1-\gamma)}{2} \sum_{k=0}^{K-1} \mathbb{E} \|x^{k+1} - w^k\|^2, \end{aligned}$$

где  $e_1(k, x) = \|w^{k+1} - x\|^2 - \|w^k - x\|^2 + (1-\gamma)\|x^{k+1} - x\|^2$ .

*Доказательство.* Для краткости введем

$$u^{k+1} = \gamma x^{k+1} + (1-\gamma)w^k - w^{k+1}.$$

Используя новые обозначения, перепишем  $e_1(k, x)$  в виде

$$e_1(k, x) = 2\langle u^{k+1}, x \rangle - \gamma \|x^{k+1}\|^2 - (1 - \gamma) \|w^k\|^2 + \|w^k\|^2.$$

Из строки 8 Алгоритма 1, учитывая, что  $\gamma = p$ , получим

$$\mathbb{E} \left[ \mathbb{E}_k [\|w^{k+1}\|^2 - \gamma \|x^{k+1}\|^2 - (1 - \gamma) \|w^k\|^2] \right] = 0.$$

Используя оба вышеприведенных свойства, получаем, что для любого компакта  $\mathcal{C} \subseteq X$

$$\begin{aligned} \mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} e_1(k, x) \right] &= 2\mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} \langle u^{k+1}, x \rangle \right] + \\ &+ \mathbb{E} \left[ -\gamma \|x^{k+1}\|^2 - (1 - \gamma) \|w^k\|^2 + \|w^k\|^2 \right] = \\ &= 2\mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} \langle u^{k+1}, x \rangle \right]. \end{aligned}$$

При  $\mathcal{F}_k = \sigma(\xi_0, \dots, \xi_k, x^k)$  имеем  $\mathbb{E}[u^{k+1} | \mathcal{F}_k] = 0$ . Это означает, что можно применить лемму 1. Таким образом, получаем

$$\begin{aligned} \mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} e_1(k, x) \right] &\leq \\ &\leq 2 \max_{x \in \mathcal{C}} \|x_0 - x\|^2 + \frac{1}{2} \sum_{k=0}^{K-1} \mathbb{E} \|u^{k+1}\|^2. \end{aligned} \quad (10)$$

Оценим  $\|u_{k+1}\|^2$ , используя факт, что  $\mathbb{E}\|X - \mathbb{E}X\|^2 = \mathbb{E}\|X\|^2 - \|\mathbb{E}X\|^2$  и строку 8 Алгоритма 1:

$$\begin{aligned} \mathbb{E} \|u_{k+1}\|^2 &= \mathbb{E} \left[ \mathbb{E}_k \|u_{k+1}\|^2 \right] = \mathbb{E} \left[ \mathbb{E}_k \left\| \mathbb{E}_k [\omega_{k+1}] - \omega_{k+1} \right\|^2 \right] = \\ &= \mathbb{E} \left[ \mathbb{E}_k \left\| \omega_{k+1} \right\|^2 - \left\| \mathbb{E}_k [\omega_{k+1}] \right\|^2 \right] = \\ &= \mathbb{E} \left[ \gamma \|x_{k+1}\|^2 + (1 - \gamma) \|\omega_k\|^2 - \|\gamma x_{k+1} + (1 - \gamma) \omega_k\|^2 \right] = \\ &= \gamma(1 - \gamma) \mathbb{E} \|x_{k+1} - \omega_k\|^2. \end{aligned}$$

Применяя этот результат к (10), получаем

$$\begin{aligned} \mathbb{E} \left[ \max_{x \in \mathcal{C}} \sum_{k=0}^{K-1} e_1(k, x) \right] &= 2 \max_{x \in \mathcal{C}} \|x_0 - x\|^2 + \\ &+ \frac{\gamma(1 - \gamma)}{2} \sum_{k=0}^{K-1} \mathbb{E} \|x_{k+1} - \omega_k\|^2. \end{aligned}$$

*Доказательство теоремы 1.* Для начала,

$$\begin{aligned} \|x^{k+1} - x\|^2 &= \|x^k - x\|^2 + 2\langle x^{k+1} - x^k, x^{k+1} - x \rangle - \\ - \|x^{k+1} - x^k\|^2 &= \|x^k - x\|^2 + 2\gamma \langle w^k - x^k, x^{k+1} - x \rangle - \\ &- 2\eta \langle \Delta^k, x^{k+1} - x \rangle - \|x^{k+1} - x^k\|^2 - \\ &- 2[\langle x^k + \gamma(w^k - x^k) - \eta \Delta^k - x^{k+1}, x^{k+1} - x \rangle]. \end{aligned}$$

Из строки 7 Алгоритма 1 и в соответствии со свойством (5) проксимального оператора следует, что

$$x^k + \gamma(w^k - x^k) - \eta \Delta^k - x^{k+1} \in \partial(\eta g)(x^{k+1}).$$

Из выпуклости  $g(\cdot)$  получаем

$$\begin{aligned} \|x^{k+1} - x\|^2 &\leq \|x^k - x\|^2 + 2\gamma \langle w^k - x^k, x^{k+1} - x \rangle - \\ - 2\eta \langle \Delta^k, x^{k+1} - x \rangle &- \|x^{k+1} - x^k\|^2 + 2\eta g(x) - 2\eta g(x^{k+1}). \end{aligned}$$

Используя  $2\gamma \langle w^k - x^k, x^{k+1} - x \rangle = 2\gamma \langle w^k - x, x^{k+1} - x \rangle - 2\gamma \langle x^k - x, x^{k+1} - x \rangle$  и следующее свойство скалярного произведения:  $2\langle a, b \rangle = \|a + b\|^2 - \|a\|^2 - \|b\|^2$ , получаем

$$\begin{aligned} \|x^{k+1} - x\|^2 &\leq \|x^k - x\|^2 + \\ &+ \gamma (\|w^k - x\|^2 + \|x^{k+1} - x\|^2 - \|x^{k+1} - w^k\|^2) - \\ &- 2\eta \langle \Delta^k, x^{k+1} - x \rangle - \gamma \|x^{k+1} - x\|^2 - \\ &- \gamma \|x^k - x\|^2 + \gamma \|x^{k+1} - x^k\|^2 - \\ &- \|x^{k+1} - x^k\|^2 + 2\eta g(x) - 2\eta g(x^{k+1}) = \|x^k - x\|^2 + \\ &+ \gamma \|w^k - x\|^2 - \gamma \|x^k - x\|^2 - \gamma \|x^{k+1} - w^k\|^2 - \\ &- 2\eta \langle \Delta^k, x^{k+1} - x \rangle - (1 - \gamma) \|x^{k+1} - x^k\|^2 + \\ &+ 2\eta g(x) - 2\eta g(x^{k+1}). \end{aligned}$$

Применяя свойства  $\mathbb{E}_k[\Delta^k]$ , указанные в лемме 2, получаем

$$\begin{aligned} \|x^{k+1} - x\|^2 &\leq \|x^k - x\|^2 + \gamma \|w^k - x\|^2 - \gamma \|x^k - x\|^2 - \\ &- \gamma \|w^k - x^{k+1}\|^2 - 2\eta \langle \mathbb{E}_k[\Delta^k], x^{k+1} - x \rangle - \\ &- (1 - \gamma) \|x^{k+1} - x^k\|^2 + 2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle + \\ &+ 2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle + 2\eta g(x) - 2\eta g(x^{k+1}) = \\ &= \|x^k - x\|^2 + \gamma \|w^k - x\|^2 - \\ &- \gamma \|x^k - x\|^2 - \gamma \|w^k - x^{k+1}\|^2 - \\ &- 2\eta \langle F(x^k) + F(x^k) - F(x^{k-1}), x^{k+1} - x \rangle - \\ &- (1 - \gamma) \|x^{k+1} - x^k\|^2 + 2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle + \\ &+ 2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle + 2\eta g(x) - 2\eta g(x^{k+1}) = \\ &= \|x^k - x\|^2 + \gamma \|w^k - x\|^2 - \\ &- \gamma \|w^k - x^{k+1}\|^2 - 2\eta \langle F(x^k) - F(x^{k+1}) + F(x^k) - \\ &- F(x^{k-1}), x^{k+1} - x \rangle - 2\eta \langle F(x^{k+1}), x^{k+1} - x \rangle - \\ &- (1 - \gamma) \|x^{k+1} - x^k\|^2 + 2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle + \\ &+ 2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle + 2\eta g(x) - 2\eta g(x^{k+1}). \end{aligned}$$

Путем несложных перестановок получаем

$$\begin{aligned}
 & 2\eta(g(x^{k+1}) - g(x)) + 2\eta\langle F(x^{k+1}), x^{k+1} - x \rangle \leq \\
 & \leq \|x^k - x\|^2 + \gamma\|w^k - x\|^2 - \gamma\|x^k - x\|^2 - \\
 & - \gamma\|w^k - x^{k+1}\|^2 - \|x^{k+1} - x\|^2 - 2\eta\langle F(x^k) - \\
 & - F(x^{k+1}) + F(x^k) - F(x^{k-1}), x^{k+1} - x \rangle - \\
 & - (1 - \gamma)\|x^{k+1} - x^k\|^2 + 2\eta\langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle + \\
 & + 2\eta\langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle = (1 - \gamma)\|x^k - x\|^2 + \\
 & + \|w^k - x\|^2 - (1 - \gamma)\|x^{k+1} - x\|^2 - \|w^{k+1} - x\|^2 + \\
 & + \|w^{k+1} - x\|^2 - \gamma\|x^{k+1} - x\|^2 - (1 - \gamma)\|w^k - x\|^2 - \\
 & - \gamma\|w^k - x^{k+1}\|^2 - 2\eta\langle F(x^k) - F(x^{k+1}), x^{k+1} - x \rangle + \\
 & + 2\eta\langle F(x^{k-1}) - F(x^k), x^{k+1} - x \rangle - \\
 & - (1 - \gamma)\|x^{k+1} - x^k\|^2 + 2\eta\langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle + \\
 & + 2\eta\langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle = (1 - \gamma)\|x^k - x\|^2 + \\
 & + \|w^k - x\|^2 - (1 - \gamma)\|x^{k+1} - x\|^2 - \|w^{k+1} - x\|^2 + \\
 & + \|w^{k+1} - x\|^2 - \gamma\|x^{k+1} - x\|^2 - (1 - \gamma)\|w^k - x\|^2 - \\
 & - \gamma\|w^k - x^{k+1}\|^2 - 2\eta\langle F(x^k) - F(x^{k+1}), x^{k+1} - x \rangle + \\
 & + 2\eta\langle F(x^{k-1}) - F(x^k), x^k - x \rangle + \\
 & + 2\eta\langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle - \\
 & - (1 - \gamma)\|x^{k+1} - x^k\|^2 + 2\eta\langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle + \\
 & + 2\eta\langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle.
 \end{aligned}$$

После взятия суммы и последующего усреднения, приходим к

$$\begin{aligned}
 & 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} [\langle F(x^{k+1}), x^{k+1} - x \rangle + g(x^{k+1}) - g(x)] \leq \\
 & \leq \frac{1}{K} \sum_{k=0}^{K-1} [(1 - \gamma)\|x^k - x\|^2 + \|w^k - x\|^2] - \\
 & - \frac{1}{K} \sum_{k=0}^{K-1} [(1 - \gamma)\|x^{k+1} - x\|^2 + \|w^{k+1} - x\|^2] - \\
 & - 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^k) - F(x^{k+1}), x^{k+1} - x \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^k - x \rangle + \\
 & + \frac{1}{K} \sum_{k=0}^{K-1} [\|w^{k+1} - x\|^2 - \gamma\|x^{k+1} - x\|^2 - \\
 & - \|(1 - \gamma)w^k - x\|^2] + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle -
 \end{aligned}$$

$$\begin{aligned}
 & - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \frac{1 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle = \\
 & = \frac{2 - \gamma}{K} \|x^0 - x\|^2 - \frac{1 - \gamma}{K} \|x^K - x\|^2 - \frac{1}{K} \|w^K - x\|^2 - \\
 & - 2\eta \cdot \frac{1}{K} \langle F(x^{K-1}) - F(x^K), x^K - x \rangle + \\
 & + \frac{1}{K} \sum_{k=0}^{K-1} [\|w^{k+1} - x\|^2 - \gamma\|x^{k+1} - x\|^2 - \\
 & - (1 - \gamma)\|w^k - x\|^2] + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle - \\
 & - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \frac{1 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle.
 \end{aligned}$$

Здесь также использовалась инициализация Алгоритма 1 с  $w^0 = x^{-1} = x^0$ . Применяя неравенство Юнга, используя  $L$ -липшицевость  $F$  и учитывая определение  $\eta \leq \frac{1}{8L}$  из условий теоремы для любого  $k$ , можно получить

$$\begin{aligned}
 & -2\eta\langle F(x^{K-1}) - F(x^K), x^K - x \rangle \leq \\
 & \leq 2\eta^2\|F(x^{K-1}) - F(x^K)\|^2 + \frac{1}{2}\|x^K - x\|^2 \leq \\
 & \leq 2\eta^2 L^2\|x^{K-1} - x^K\|^2 + \frac{1}{2}\|x^K - x\|^2 \leq \\
 & \leq \frac{1}{32}\|x^{K-1} - x^K\|^2 + \frac{1}{2}\|x^K - x\|^2.
 \end{aligned}$$

Объединим (11) и (12):

$$\begin{aligned}
 & a2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} [\langle F(x^{k+1}), x^{k+1} - x \rangle + g(x^{k+1}) - g(x)] \leq \\
 & \leq \frac{2 - \gamma}{K} \|x^0 - x\|^2 - \frac{1}{K} \left( \frac{1}{2} - \gamma \right) \|x^K - x\|^2 - \\
 & - \frac{1}{K} \|w^K - x\|^2 + \frac{1}{32K} \|x^{K-1} - x^K\|^2 + \\
 & + \frac{1}{K} \sum_{k=0}^{K-1} [\|w^{k+1} - x\|^2 - \gamma\|x^{k+1} - x\|^2 -
 \end{aligned}$$

$$\begin{aligned}
 & - (1 - \gamma)\|w^k - x\|^2 + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle - \\
 & - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle \leq \\
 & \leq \frac{2-\gamma}{K} \|x^0 - x\|^2 + \frac{1}{K} \sum_{k=0}^{K-1} (\|w^{k+1} - x\|^2 - \\
 & - \gamma \|x^{k+1} - x\|^2 - (1-\gamma)\|w^k - x\|^2) + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle - \\
 & - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \\
 & - \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + \frac{1}{32K} \|x^{K-1} - x^K\|^2 + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle.
 \end{aligned}$$

Далее используем монотонность  $F$ , применяем неравенство Йенсена для выпуклой функции  $g$ :

$$\begin{aligned}
 & 2\eta \left[ \left\langle F(x), \frac{1}{K} \sum_{k=0}^{K-1} x^{k+1} - x \right\rangle + g \left( \frac{1}{K} \sum_{k=0}^{K-1} x^{k+1} \right) - g(x) \right] \leq \\
 & \leq \frac{2-\gamma}{K} \|x^0 - x\|^2 + \frac{1}{K} \sum_{k=0}^{K-1} (\|w^{k+1} - x\|^2 - \\
 & - \gamma \|x^{k+1} - x\|^2 - (1-\gamma)\|w^k - x\|^2) + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle - \\
 & - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + \\
 & + \frac{1}{32K} \|x^{K-1} - x^K\|^2 + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle.
 \end{aligned}$$

Используя новое обозначение  $\bar{x}^K = \frac{1}{K} \sum_{k=0}^{K-1} x^{k+1}$  и взяв максимум на  $\mathcal{C}$ , получаем

$$\begin{aligned}
 & 2\eta \text{Gap}(\bar{x}^K) \leq \max_{x \in \mathcal{C}} \left\{ \frac{2-\gamma}{K} \|x^0 - x\|^2 + \right. \\
 & + \frac{1}{K} \sum_{k=0}^{K-1} (\|w^{k+1} - x\|^2 - \gamma \|x^{k+1} - x\|^2 - \\
 & - (1-\gamma)\|w^k - x\|^2) + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle \left. \right\} + \\
 & + \frac{1}{32K} \|x^{K-1} - x^K\|^2 - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \\
 & - \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle + \\
 & + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle \leq \\
 & \leq \max_{x \in \mathcal{C}} \left\{ \frac{2-\gamma}{K} \|x^0 - x\|^2 \right\} + \max_{x \in \mathcal{C}} \left\{ \frac{1}{K} \sum_{k=0}^{K-1} (\|w^{k+1} - x\|^2 - \right. \\
 & - \gamma \|x^{k+1} - x\|^2 - (1-\gamma)\|w^k - x\|^2) \left. \right\} + \\
 & + 2\eta \max_{x \in \mathcal{C}} \left\{ \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle \right\} + \\
 & + \frac{1}{32K} \|x^{K-1} - x^K\|^2 - \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 - \\
 & - \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - \\
 & - F(x^k), x^{k+1} - x^k \rangle + 2\eta \cdot \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle.
 \end{aligned}$$

Здесь также использовалось то, что максимум суммы не превосходит сумму максимумов. После этого берем математическое ожидание и получаем

$$\begin{aligned}
 & 2\eta \mathbb{E}[\text{Gap}(\bar{x}^K)] \leq \mathbb{E} \left[ \max_{x \in \mathcal{C}} \left\{ \frac{2-\gamma}{K} \|x^0 - x\|^2 \right\} \right] + \\
 & + \mathbb{E} \left[ \max_{x \in \mathcal{C}} \left\{ \frac{1}{K} \sum_{k=0}^{K-1} (\|w^{k+1} - x\|^2 - \right. \right.
 \end{aligned}$$

$$\begin{aligned}
 & -\gamma\|x^{k+1} - x\|^2 - (1-\gamma)\|w^k - x\|^2 \Big\} + \\
 & + 2\eta \mathbb{E} \left[ \max_{x \in \mathcal{C}} \left\{ \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^k - x \rangle \right\} \right] + \\
 & \quad + \frac{1}{32K} \mathbb{E} [\|x^{K-1} - x^K\|^2] - \\
 & - \mathbb{E} \left[ \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 \right] + \\
 & + 2\eta \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle \right] + \\
 & + 2\eta \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle \right].
 \end{aligned}$$

Используя Лемму 3 для второй строки предыдущей оценки и Лемму 1 для третьей строки, имеем

$$\begin{aligned}
 2\eta \mathbb{E} [\text{Gap}(\bar{x}^K)] & \leq \mathbb{E} \left[ \max_{x \in \mathcal{C}} \left\{ \frac{2-\gamma}{K} \|x^0 - x\|^2 \right\} \right] + \\
 & + \max_{x \in \mathcal{C}} \left\{ \frac{2}{K} \|x - x^0\|^2 \right\} + \frac{\gamma(1-\gamma)}{2K} \sum_{k=0}^{K-1} \mathbb{E} [\|x^{k+1} - \omega^k\|^2] + \\
 & + \max_{x \in \mathcal{C}} \left\{ \frac{1}{K} \|x - x^0\|^2 \right\} + \frac{\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\
 & - \mathbb{E} \left[ \frac{\gamma}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 \right] + \\
 & \quad + \frac{1}{32K} \mathbb{E} [\|x^{K-1} - x^K\|^2] + \\
 & + 2\eta \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle \right] + \\
 & + 2\eta \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle \right] \leq \\
 & \leq \frac{4}{K} \mathbb{E} \left[ \max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\} \right] + \frac{\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\
 & - \mathbb{E} \left[ \frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 \right] + \\
 & \quad + \frac{1}{32K} \mathbb{E} [\|x^{K-1} - x^K\|^2] + \\
 & + 2\eta \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle \right] + \\
 & + 2\eta \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle \right].
 \end{aligned} \tag{13}$$

Согласно неравенству Юнга,

$$\begin{aligned}
 & \mathbb{E} [2\eta \langle \mathbb{E}_k[\Delta^k] - \Delta^k, x^{k+1} - x^k \rangle] \leq \\
 & \leq 4\eta^2 \mathbb{E} [\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] + \frac{1}{4} \mathbb{E} [\|x^{k+1} - x^k\|^2], \tag{14}
 \end{aligned}$$

и

$$\begin{aligned}
 & \mathbb{E} [2\eta \langle F(x^{k-1}) - F(x^k), x^{k+1} - x^k \rangle] \leq \\
 & \leq 4\eta^2 \mathbb{E} [\|F(x^{k-1}) - F(x^k)\|^2] + \frac{1}{4} \mathbb{E} [\|x^{k+1} - x^k\|^2]. \tag{15}
 \end{aligned}$$

Комбинируя (14), (15) с (13), получаем

$$\begin{aligned}
 2\eta \mathbb{E} [\text{Gap}(\bar{x}^K)] & \leq \frac{4}{K} \mathbb{E} \left[ \max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\} \right] + \\
 & \quad + \frac{\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\
 & - \mathbb{E} \left[ \frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 \right] + \\
 & \quad + \frac{1}{32K} \mathbb{E} [\|x^{K-1} - x^K\|^2] + \\
 & \quad + \frac{4\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|F(x^{k-1}) - F(x^k)\|^2] + \\
 & \quad + \frac{1}{4K} \sum_{k=0}^{K-1} \mathbb{E} [\|x^{k+1} - x^k\|^2] + \\
 & + \frac{4\eta^2}{K} \mathbb{E} \sum_{k=0}^{K-1} \|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2 + \frac{1}{4K} \sum_{k=0}^{K-1} \mathbb{E} [\|x^{k+1} - x^k\|^2] \leq \\
 & \leq \frac{4}{K} \mathbb{E} [\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] + \frac{5\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\
 & - \mathbb{E} \left[ \frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/2-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 \right] + \\
 & + \frac{1}{32K} \mathbb{E} [\|x^{K-1} - x^K\|^2] + \frac{4\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|F(x^{k-1}) - F(x^k)\|^2].
 \end{aligned}$$

$L$  – Липшицевость  $F$  (Предположение 1) и выбор  $\gamma \leq \frac{1}{L}$  дает

$$\begin{aligned}
 2\eta \mathbb{E} [\text{Gap}(\bar{x}^K)] & \leq \frac{4}{K} \mathbb{E} \left[ \max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\} \right] + \\
 & \quad + \frac{5\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\
 & - \mathbb{E} \left[ \frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/2-\gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2 \right] + \\
 & + \frac{1}{32K} \mathbb{E} [\|x^{K-1} - x^K\|^2] + \frac{4\eta^2 L^2}{K} \sum_{k=0}^{K-1} \mathbb{E} [\|x^{k+1} - x^k\|^2] \leq
 \end{aligned}$$

$$\begin{aligned} &\leq \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] + \frac{5\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E}[\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\ &- \mathbb{E}\left[\frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/2 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right] + \\ &+ \frac{1}{32K} \mathbb{E}[\|x^{K-1} - x^K\|^2] + \frac{1}{4K} \sum_{k=0}^{K-1} \mathbb{E}[\|x^{k-1} - x^k\|^2] \leq \\ &\leq \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] + \frac{5\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E}[\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\ &- \mathbb{E}\left[\frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/2 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right] + \\ &+ \frac{1}{4K} \sum_{k=0}^{K-1} \mathbb{E}[\|x^{k+1} - x^k\|^2] = \\ &= \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] + \frac{5\eta^2}{K} \sum_{k=0}^{K-1} \mathbb{E}[\|\mathbb{E}_k[\Delta^k] - \Delta^k\|^2] - \\ &- \mathbb{E}\left[\frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/4 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right]. \end{aligned}$$

Здесь также использовалась инициализация Алгоритма 1 с  $x^{-1} = x^0$ . Применяя Лемму 2, получаем

$$\begin{aligned} 2\eta \mathbb{E}[\text{Gap}(\bar{x}^K)] &\leq \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] + \\ &+ \frac{10\eta^2 \bar{L}^2}{bK} \sum_{k=0}^{K-1} (\|x^k - w^{k+1}\|^2 + \|x^k - x^{k+1}\|^2) - \\ &- \mathbb{E}\left[\frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/4 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right] \leq \\ &\leq \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] + \\ &+ \frac{10\eta^2 \bar{L}^2}{bK} \sum_{k=0}^{K-1} (\|x^{k+1} - w^k\|^2 + \|x^k - x^{k+1}\|^2) - \\ &- \mathbb{E}\left[\frac{\gamma}{2K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \frac{1/4 - \gamma}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right] \leq \\ &\leq \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] - \mathbb{E}\left[\left(\frac{\gamma}{2} - \frac{10\eta^2 \bar{L}^2}{b}\right) \times \right. \\ &\quad \times \frac{1}{K} \sum_{k=0}^{K-1} \|w^k - x^{k+1}\|^2 + \\ &\quad \left. + \left(\frac{1}{4} - \gamma - \frac{10\eta^2 \bar{L}^2}{b}\right) \frac{1}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right]. \end{aligned}$$

Здесь мы снова использовали инициализацию Алгоритма 1 с  $w^{-1} = x^{-1} = x^0$ . Выбор  $\eta \leq \frac{\sqrt{\gamma b}}{8L}$  и  $0 < \gamma \leq \frac{1}{16}$  дает

$$\begin{aligned} 2\eta \mathbb{E}[\text{Gap}(\bar{x}^K)] &\leq \frac{4}{K} \mathbb{E}[\max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}] \\ &- \mathbb{E}\left[\left(\frac{1}{12} - \gamma\right) \frac{1}{K} \sum_{k=0}^{K-1} \|x^{k+1} - x^k\|^2\right] \leq \frac{4}{K} \max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}. \end{aligned}$$

А значит мы получаем

$$\mathbb{E}[\text{Gap}(\bar{x}^K)] \leq \frac{2}{\eta K} \max_{x \in \mathcal{C}} \{\|x^0 - x\|^2\}.$$

Подстановка  $\eta$  из условий теоремы и  $\gamma = p$  завершает доказательство.

### ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследования А. Пичугина и М. Печина были поддержаны Аналитическим центром при Правительстве Российской Федерации в соответствии с договором о субсидии (идентификатор договора 000000D730321P5Q0002; грант № 70-2021-00138).

### СПИСОК ЛИТЕРАТУРЫ

1. *Facchinei F., Pang J.-S.* Finite-Dimensional Variational Inequalities and Complementarity Problems (Springer Series in Operations Research). Springer, 2003.
2. *Scutari G., Palomar D.P., Facchinei F., Pang J.-S.* "Convex optimization, game theory, and variational inequality theory," IEEE Signal Processing Magazine. 2010. V. 27. № 3. P. 35–49.
3. *Neumann J.V., Morgenstern O.* Theory of games and economic behavior. Princeton University Press, 1944.
4. *Harker P.T., Pang J.-S.* "Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications," Mathematical programming, 1990.
5. *Ben-Tal A., Ghaoui L.E., Nemirovski A.* Robust Optimization. Princeton University Press, 2009.
6. *Nesterov Y.* "Dual extrapolation and its applications to solving variational inequalities and related problems," Mathematical Programming. 2007. V. 109. № 2. P. 319–344.
7. *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 Journal on Optimization. 2004. V. 15. № 1. P. 229–251.
8. *Joachims T.* "A support vector method for multivariate performance measures," Jan. 2005. P. 377–384. <https://doi.org/10.1145/1102351.1102399>
9. *Bach F., Jenatton R., Mairal J., Obozinski G.* "Optimization with sparsity-inducing penalties," arXiv preprint arXiv:1108.0775, 2011.

10. *Xu L., Neufeld J., Larson B., Schuurmans D.* “Maximum margin clustering,” in *Advances in Neural Information Processing Systems*. 2005. V. 17. P. 1537–1544.
11. *Bach F., Mairal J., Ponce J.* “Convex sparse matrix factorizations,” arXiv preprint arXiv:0812.1869, 2008.
12. *Esser E., Zhang X., Chan T.F.* “A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science,” *SIAM Journal on Imaging Sciences*. 2010. V. 3. № 4. P. 1015–1046.
13. *Chambolle A., Pock T.* “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of mathematical imaging and vision*. 2011. V. 40. № 1. P. 120–145.
14. *Omidshafiei S., Pazis J., Amato C., How J.P., Vian J.* “Deep decentralized multi-task multi-agent reinforcement learning under partial observability,” in *Proceedings of the 34th International Conference on Machine Learning (ICML)*. V. 70, PMLR, 2017. P. 2681–2690. [Online]. Available: <http://proceedings.mlr.press/v70/omidshafiei17a.html>.
15. *Jin Y., Sidford A.* “Efficiently solving MDPs with stochastic mirror descent,” in *Proceedings of the 37th International Conference on Machine Learning (ICML)*. V. 119, PMLR, 2020. P. 4890–4900.
16. *Madry A., Makelov A., Schmidt L., Tsipras D., Vladu A.* “Towards deep learning models resistant to adversarial attacks,” in *International Conference on Learning Representations (ICLR)*, 2018.
17. *Goodfellow I.J., Pouget-Abadie J., Mirza M. et al.*, *Generative adversarial networks*, 2014. arXiv: 1406.2661 [stat.ML].
18. *Daskalakis C., Ilyas A., Syrgkanis V., Zeng H.* “Training gans with optimism,” arXiv preprint arXiv:1711.00141, 2017.
19. *Gidel G., Berard H., Vignoud G., Vincent P., Lacoste-Julien S.* “A variational inequality perspective on generative adversarial networks,” arXiv preprint arXiv:1802.10551, 2018.
20. *Mertikopoulos P., Lecouat B., Zenati H., Foo C.-S., Chandrasekhar V., Piliouras G.* “Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile,” arXiv preprint arXiv:1807.02629, 2018.
21. *Robbins H., Monro S.* “A Stochastic Approximation Method,” *The Annals of Mathematical Statistics*. 1951. V. 22. № 3. P. 400–407.
22. *Johnson R., Zhang T.* “Accelerating stochastic gradient descent using predictive variance reduction,” vol. 26, 2013. P. 315–323.
23. *Defazio A., Bach F., Lacoste-Julien S.* “Saga: A fast incremental gradient method with support for non-strongly convex composite objectives,” in *Advances in neural information processing systems*, 2014. P. 1646–1654.
24. *Nguyen L.M., Liu J., Scheinberg K., Takáč M.* “SARAH: a novel method for machine learning problems using stochastic recursive gradient,” in *International Conference on Machine Learning*, PMLR, 2017. P. 2613–2621.
25. *Allen-Zhu Z.* “Katyusha: The first direct acceleration of stochastic gradient methods,” in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, 2017. P. 1200–1205.
26. *Korpelevich G.M.* “The extragradient method for finding saddle points and other problems,” *Matecon*. V. 12. P. 35–49, 1977; Russian original: *Economika Mat. Metody*. 1976. V. 12. № 4. P. 747–756.
27. *Mokhtari A., Ozdaglar A., Pattathil S.* “A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach,” in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2020. P. 1497–1507.
28. *Popov L.D.* “A modification of the arrow-hurwicz method for search of saddle points,” *Mathematical notes of the Academy of Sciences of the USSR*. 1980. V. 28. P. 845–848.
29. *Tseng P.* “A modified forward-backward splitting method for maximal monotone mappings,” *SIAM Journal on Control and Optimization*. 2000. V. 38. № 2. P. 431–446.
30. *Juditsky A., Nemirovski A., Tauvel C.* “Solving variational inequalities with stochastic mirror-prox algorithm,” *Stochastic Systems*. 2011. V. 1. № 1. P. 17–58.
31. *Carmon Y., Jin Y., Sidford A., Tian K.* “Variance reduction for matrix games,” arXiv preprint arXiv:1907.02056, 2019.
32. *Alacaoglu A., Malitsky Y.* “Stochastic variance reduction for variational inequality methods,” arXiv preprint arXiv:2102.08352, 2021.
33. *Alacaoglu A., Malitsky Y., Cevher V.* “Forward-reflected-backward method with variance reduction,” *Computational Optimization and Applications*. V. 80, Nov. 2021. <https://doi.org/10.1007/s10589-021-00305-3>
34. *Han Y., Xie G., Zhang Z.* “Lower complexity bounds of finite-sum optimization problems: The results and construction,” arXiv preprint arXiv:2103.08280, 2021.
35. *Kovalev D., Beznosikov A., Sadiev A., Pershiyanov M., Richt'arik P., Gasnikov A.* “Optimal algorithms for decentralized stochastic variational inequalities,” arXiv preprint arXiv:2202.02771, 2022.
36. *Nemirovski A.* *Mini-course on convex programming algorithms*, 2013.
37. *Nemirovski A., Juditsky A., Lan G., Shapiro A.* “Robust stochastic approximation approach to stochastic programming,” *SIAM Journal on optimization*. 2009. V. 19. № 4. P. 1574–1609.
38. *Condat L.* “Fast projection onto the simplex and the l1 ball,” *Mathematical Programming*. 2016. V. 158. № 1–2. P. 575–585.

## OPTIMAL ANALYSIS OF METHOD WITH BATCHING FOR MONOTONE STOCHASTIC FINITE-SUM VARIATIONAL INEQUALITIES

A. Pichugin<sup>a</sup>, M. Pechin<sup>a</sup>, A. Beznosikov<sup>a</sup>, A. Savchenko<sup>b</sup>, and A. Gasnikov<sup>a</sup>

<sup>a</sup>*Moscow Institute of Physics and Technology, Dolgoprudny, Russian Federation*

<sup>b</sup>*Sber AI Lab, Moscow, Russian Federation*

Presented by Academician of the RAS A.A. Shaninin

Variational inequalities are a universal optimization paradigm that is interesting in itself, but also incorporates classical minimization and saddle point problems. Modern realities encourage to consider stochastic formulations of optimization problems. In this paper, we present an analysis of a method that gives optimal convergence estimates for monotone stochastic finite-sum variational inequalities. In contrast to the previous works, our method supports batching and does not lose the oracle complexity optimality. The effectiveness of the algorithm, especially in the case of small but not single batches is confirmed experimentally.

*Keywords:* stochastic optimization, variational inequalities, finite-sum problems, batching