

Оптимальный метод решения седловых билинейных задач

Владимиров Игорь Викторович

Гаврилов Олег Алексеевич

Дёмина Анна Андреевна

Филиал МГУ в Сарове

2022



Постановка задачи

Найти седловую точку функции

$$F(x, y) = f(x) + y^T Ax - g(y) \rightarrow \min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}},$$

где $f(x) : \mathbb{R}^{d_x} \rightarrow \mathbb{R}$, $g(y) : \mathbb{R}^{d_y} \rightarrow \mathbb{R}$ — выпуклые функции;
 $A \in \mathbb{R}^{d_y \times d_x}$ — матрица.

Основная работа:

Dmitry Kovalev, Alexander Gasnikov, and Peter Richtárik.
Accelerated primal-dual gradient method for smooth and
convex-concave saddle-point problems with bilinear coupling.

Конкретная задача

В качестве $f(x)$, $g(y)$ возьмём квадратичные функции:

$$f(x) = \frac{\mu_x}{2} \|x\|_2^2, \quad g(y) = \frac{\mu_y}{2} \|y\|_2^2,$$

$\mu_x, \mu_y > 0$ — константы сильной выпуклости. Можно также заметить, что $L_x = \mu_x$, $L_y = \mu_y$. Тогда функция

$$F(x, y) = \frac{\mu_x}{2} \|x\|_2^2 + y^T Ax - \frac{\mu_y}{2} \|y\|_2^2$$

является сильно выпуклой по x и сильно вогнутой по y .

Пусть (x^*, y^*) — искомая седловая точка $F(x, y)$. Тогда

$$\nabla_x F(x^*, y^*) = \nabla_y F(x^*, y^*) = 0,$$

$$\begin{cases} \mu_x x^* + A^T y^* = 0, \\ Ax^* - \mu_y y^* = 0, \end{cases}$$

$$\left(\mu_x I + \frac{1}{\mu_y} A^T A \right) x^* = 0.$$

Отсюда следует, что $x^* = 0$, $y^* = 0$ для любой матрицы A .

Алгоритм 1 (APDG)

Задача эквивалентна вычислению корня $G_1(x, y) + G_2(x, y) = 0$, где

$$G_1 : (x, y) \mapsto (\nabla f(x), \nabla g(y)),$$

$$G_2 : (x, y) \mapsto (A^T y, -Ax).$$

Итерационный процесс:

$$(x^+, y^+) = J_{G_2}((x, y) - G_1(x, y)), \quad J_{G_2} = (I + G_2)^{-1}.$$

Здесь (x, y) — значения на текущей итерации, (x^+, y^+) — на следующей. Это эквивалентно решению линейной системы

$$\begin{cases} x^+ = x - \nabla f(x) - A^T y^+, \\ y^+ = y - \nabla g(y) + Ax^+. \end{cases}$$

В первом уравнении y^+ заменяется на $y_m = y + \theta(y - y^-)$, где $\theta \in (0, 1]$, y^- — значение y с предыдущей итерации (линейная экстраполяция).

Далее система модифицируется с помощью ускоренного метода Нестерова.

Специальные предположения

Предположение 1.1. Функция $f(x)$ является L_x -гладкой и μ_x -сильно выпуклой, $L_x \geq \mu_x \geq 0$.

Предположение 1.2. Функция $g(y)$ является L_y -гладкой и μ_y -сильно выпуклой, $L_y \geq \mu_y \geq 0$.

Предположение 1.3. Существуют константы $L_{xy} > \mu_{xy}, \mu_{yx} \geq 0$, такие что

$$\mu_{xy}^2 \leq \begin{cases} \lambda_{\min}^+(AA^T), & \nabla g(y) \in \text{Im } A, \quad \forall y \in \mathbb{R}^{d_y}, \\ \lambda_{\min}(AA^T), & \text{иначе,} \end{cases}$$
$$\mu_{yx}^2 \leq \begin{cases} \lambda_{\min}^+(A^T A), & \nabla f(x) \in \text{Im } A^T, \quad \forall x \in \mathbb{R}^{d_x}, \\ \lambda_{\min}(A^T A), & \text{иначе,} \end{cases}$$

$$L_{xy}^2 \geq \lambda_{\max}(A^T A) = \lambda_{\max}(AA^T).$$

Сходимость алгоритма 1

Теорема 1

Пусть выполнены предположения 1.1–1.3 и условие

$$\min\{\max\{\mu_x, \mu_{yx}\}, \max\{\mu_y, \mu_{xy}\}\} > 0.$$

Тогда существуют параметры алгоритма, такие что справедлива следующая оценка: если число итераций

$$k \geq \frac{1}{1 - \theta} \log \frac{C}{\varepsilon},$$

то получено ε -точное решение, то есть

$$\max\left\{\|x^k - x^*\|_2^2, \|y^k - y^*\|_2^2\right\} \leq \varepsilon.$$

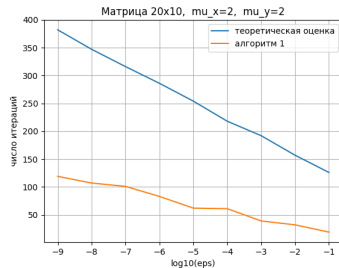
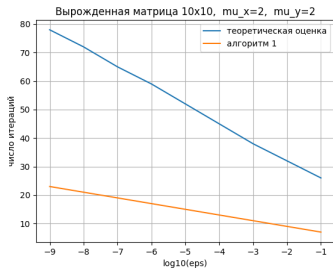
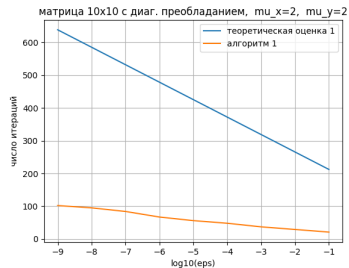
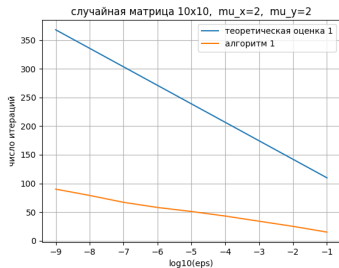
Здесь θ — константа, зависящая от $\mu_x, \mu_y, \mu_{xy}, \mu_{yx}, L_x, L_y, L_{xy}$ и от параметров алгоритма, а C зависит также от начального приближения (x^0, y^0) .

Примеры работы алгоритма 1

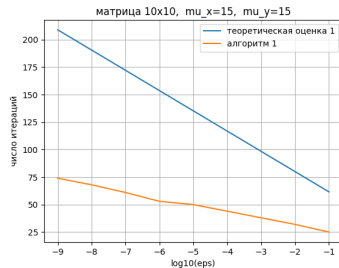
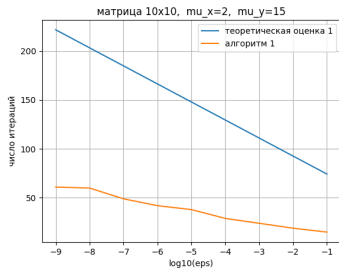
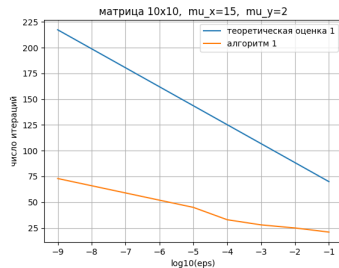
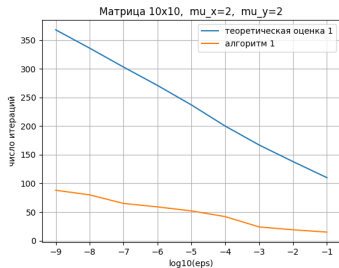
Далее будем подбирать случайным образом конкретные матрицы A разных размерностей. Для каждой такой функции $F(x, y)$ получим число итераций k алгоритма 1, необходимых для достижения точности ε . Это число k должно не превосходить теоретической оценки

$$\frac{1}{1 - \theta} \log \frac{C}{\varepsilon}.$$

Примеры для различных матриц A (здесь $\mu_x = \mu_y = 2$)



Примеры для различных μ_x, μ_y (A — фикс. квадратная)



Алгоритм 2 (GDAE)

Рассмотрим более общую задачу

$$\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} F(x, y),$$

где F выпукла по x и вогнута по y .

Итерации строятся следующим образом:

$$\begin{cases} x^+ = x - \eta_x \nabla_x F(x, y^+), \\ y^+ = y + \eta_y \nabla_y F(x^+, y), \end{cases}$$

где $\nabla_x F(x, y^+)$ заменяется на

$$\Delta_x = \nabla_x F(x, y) + \theta (\nabla_x F(x^-, y) - \nabla_x F(x^-, y^-)), \quad \theta \in (0, 1].$$

Предположение 2.1. Функция $F(x, y)$ является L_x -гладкой и μ_x -сильно выпуклой по x и L_y -гладкой и μ_y -сильно выпуклой по y , где $L_x \geq \mu_x \geq 0, L_y \geq \mu_y \geq 0$.

Предположение 2.2. Существует $L_{xy} > 0$, такая что $\forall x, x_1, x_2 \in \mathbb{R}^{d_x}$ и $\forall y, y_1, y_2 \in \mathbb{R}^{d_y}$ выполнены неравенства

$$\begin{aligned}\|\nabla_x F(x, y_1) - \nabla_x F(x, y_2)\| &\leq L_{xy} \|y_1 - y_2\|, \\ \|\nabla_y F(x_1, y) - \nabla_y F(x_2, y)\| &\leq L_{xy} \|x_1 - x_2\|.\end{aligned}$$

Предположение 2.3. Существуют $\mu_{xy}, \mu_{yx} \geq 0$, такие что $\forall x, x_1, x_2 \in \mathbb{R}^{d_x}$ и $\forall y, y_1, y_2 \in \mathbb{R}^{d_y}$ выполнены неравенства

$$\begin{aligned}\|\nabla_x F(x, y_1) - \nabla_x F(x, y_2)\| &\geq \mu_{xy} \|y_1 - y_2\|, \\ \|\nabla_y F(x_1, y) - \nabla_y F(x_2, y)\| &\geq \mu_{yx} \|x_1 - x_2\|.\end{aligned}$$

Сходимость алгоритма 2

Справедлива аналогичная оценка достаточного числа итераций для достижения точности ε :

$$k \geq \frac{1}{1 - \theta} \log \frac{C}{\varepsilon},$$

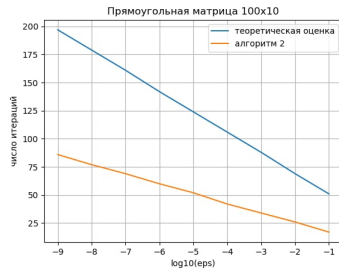
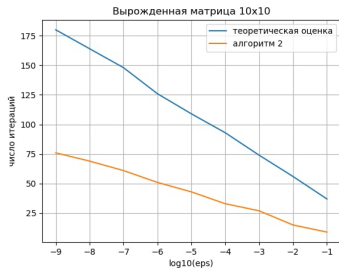
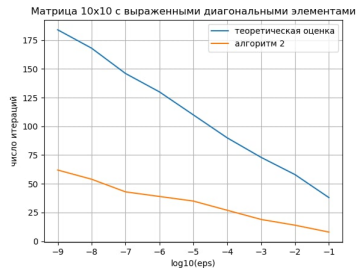
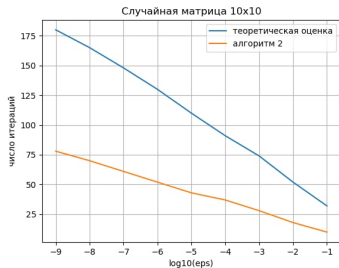
где константы θ, C зависят от начального приближения (x^0, y^0) и свойств функции $F(x, y)$.

Будем рассматривать функции того же вида

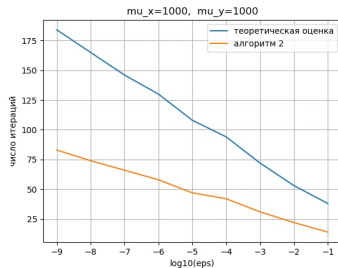
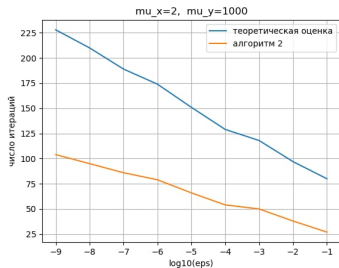
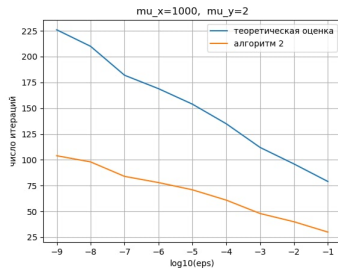
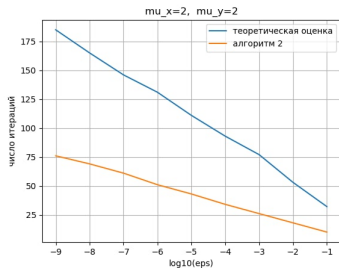
$$F(x, y) = \frac{\mu_x}{2} \|x\|_2^2 + y^T A x - \frac{\mu_y}{2} \|y\|_2^2$$

с разными матрицами A и параметрами μ_x, μ_y .

Примеры алгоритма 2 (здесь $\mu_x = \mu_y = 2$)



Примеры алгоритма 2 (здесь A фиксирована)



Сравнение двух алгоритмов

Из результатов статьи следует, что для сильно выпуклых-сильно вогнутых задач (в частности, для нашей конкретной задачи), алгоритм 1 является ускоренным, то есть число итераций пропорционально величинам $\sqrt{L_x/\mu_x}$ и $\sqrt{L_y/\mu_y}$, в то время как для алгоритма 2 число итераций связано с L_x/μ_x , L_y/μ_y . В нашей задаче, для которой

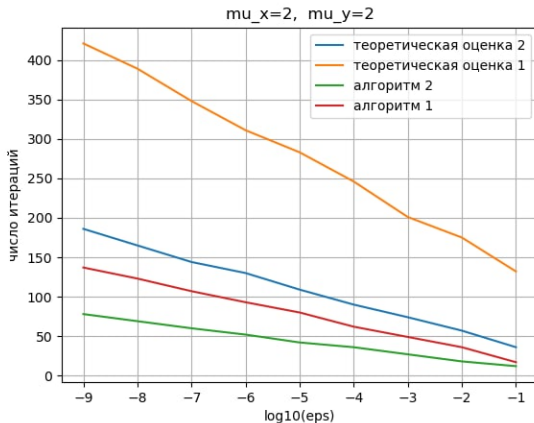
$$f(x) = \frac{\mu_x}{2} \|x\|_2^2, \quad g(y) = \frac{\mu_y}{2} \|y\|_2^2,$$

выполнены равенства

$$L_x = \mu_x, \quad L_y = \mu_y,$$

поэтому для такой задачи алгоритм 1 *не обязательно* даёт более лучшие результаты, чем алгоритм 2.

Сравнение двух алгоритмов, пример



Модифицированная задача

Рассмотрим функции

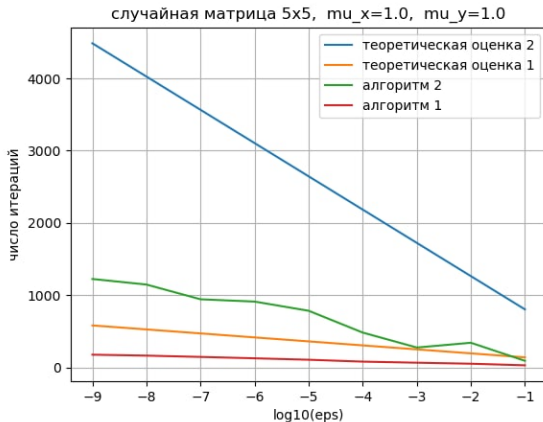
$$f(x) = \frac{1}{2} \|Bx\|_2^2, \quad g(y) = \frac{1}{2} \|By\|_2^2.$$

с некоторой квадратной матрицей B (для простоты считаем, что $d_x = d_y$). Для их констант сильной выпуклости и гладкости выполнены равенства

$$\mu_x = \mu_y = \lambda_{\min}(B^T B), \quad L_x = L_y = \lambda_{\max}(B^T B).$$

Модифицированная задача, пример

Пусть $d_x = d_y = 5$, $B = \text{diag}\{1, 2, 3, 4, 5\}$.



Заключение

1. Реализованы два алгоритма, предложенные в статье, для решения конкретной седловой задачи.
2. Исследована сходимость обоих алгоритмов для разных параметров задачи. Установлено, что число итераций для достижения требуемой точности удовлетворяет теоретической оценке.
3. Проведено сравнение двух алгоритмов на основе модифицированной задачи. Алгоритм 1 показал оптимальную сходимость.

Основная работа:

Dmitry Kovalev, Alexander Gasnikov, and Peter Richtárik.
Accelerated primal-dual gradient method for smooth and convex-concave saddle-point problems with bilinear coupling.