Решение седловых задач с небольшой размерностью одной из групп переменных

Егор Гладин
Научный руководитель: д.ф.-м.н. Гасников А.В.

Московский физико-технический институт

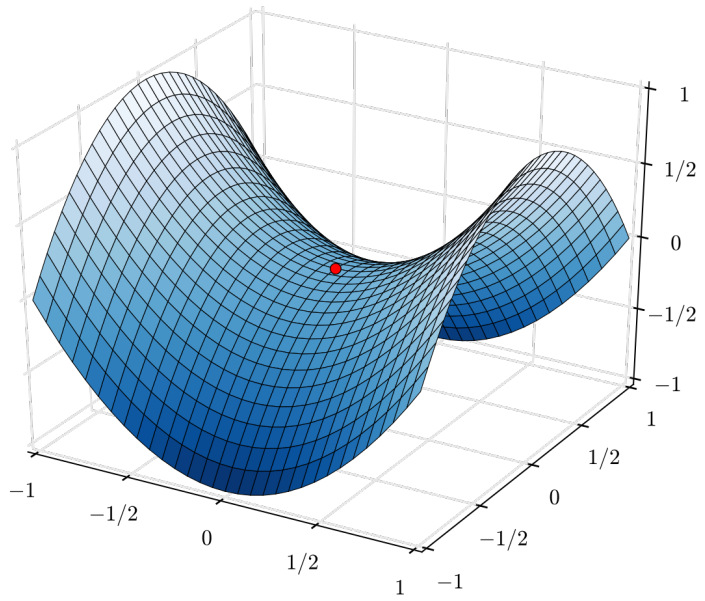
Москва
16.06.2021 г

Седловыми называются задачи вида

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y).$$

Решение **(седловая точка)**:

$$(x_*, y_*) \in \mathcal{X} \times \mathcal{Y} : f(x_*, y) \leqslant f(x_*, y_*) \leqslant f(x, y_*)$$



wikipedia.org/wiki/Saddle_point#
/media/File:Saddle_point.svg

Приложения

- машинное обучение;
- компьютерная графика;
- теория игр;
- теория оптимального транспорта.

Обзор решений

Градиентные методы

Особенности

- Обычно требуется гладкость f;
- Сильно зависят от обусловленности задачи;
- Хорошо подходят для большой размерности.

Методы секущей плоскости

Особенности

- **Не** требуется гладкость f;
- Сходятся линейно не требуя сильной выпуклости;
- **Не** подойдут для большой размерности.

$$\min_{x \in \mathcal{X}} \max_{y \in \mathbb{R}^{n_y}} f(x, y), \tag{1}$$

где

1. $\mathcal{X} \subset \mathbb{R}^{n_x}$ — выпуклое компактное множество с непустой внутренностью;
2. размерность n_x небольшая ($n_x < 100$);
3. непрерывная функция $f(x, y)$ выпукла по x и μ-сильно вогнута по y;
4. для каждого $x \in \mathcal{X}$ функция $f(x, \cdot)$ является L–гладкой, т.е.

$$\left\|\nabla_y f(x, y) - \nabla_y f(x, y')\right\|_2 \leqslant L \left\|y - y'\right\|_2 \quad \forall y, y' \in \mathbb{R}^{n_y}.$$

Оракулы первого порядка

Доступны $\nabla_x f$ и $\nabla_y f$.

Смешанный оракул

Доступны только $\nabla_x f$ и $f(x, y)$.

Цель

Предложить **эффективный** подход к решению описанного класса задач.
Критерий эффективности — сложность (число вызовов оракула).

Задачи:

1. Разработать подход, сложность которого N зависит от точности ε как

$$N \sim \log^k \frac{1}{\varepsilon}, \; k \leq 2$$

   и от числа обусловленности $\kappa = \frac{L}{\mu}$ как $N \sim \sqrt{\kappa}$;
2. Доказать оценку скорости сходимости;
3. Провести численный эксперимент.

Рассмотрим функцию

$$g(x) = \underbrace{\max_{y \in \mathbb{R}^{n_y}} f(x, y)}_{\text{“внутренняя задача”}}, \tag{2}$$

и перепишем исходную задачу (1) следующим образом:

$$\underbrace{\min_{x \in \mathcal{X}} g(x)}_{\text{“внешняя задача”}}. \tag{3}$$

Идея

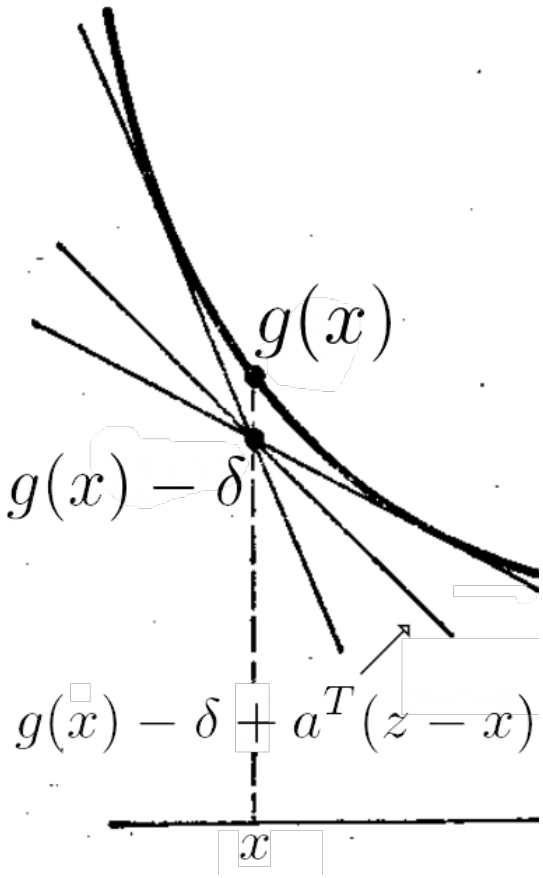
- Решать задачу (2) приближённо.
- Воспринимать (3) как задачу с неточным оракулом;

Определение

Вектор $a \in \mathbb{R}^n$ называется δ–субградиентом выпуклой функции g в x, если

$$g(z) \geq g(x) + a^T(z - x) - \delta \quad \forall z \in \operatorname{dom} f.$$

Обозначение: $a \in \partial_\delta g(x)$.



$$\min_{x \in Q} f(x). \tag{4}$$

Суть метода:

- алгоритм производит последовательность пар $\left(A^{(t)}, b^{(t)}\right) \in \mathbb{R}^{m_t \times n} \times \mathbb{R}^{m_t}$ таких, что множество $\left\{x \in \mathbb{R}^n : A^{(t)} x > b^{(t)}\right\}$ содержит решение;
- на каждом шаге к текущему множеству либо добавляется ограничение, чтобы увеличить его размер, либо убирается наименее “полезное”, если их слишком много.

Теорема[2] (Гладин)

Сходимость метода Вайды с δ–субградиентом:

$$f(x^N) - f(x_*) = O\left(n^{1.5} e^{-\frac{\gamma N}{2n}}\right) + \delta. \tag{5}$$

где $\gamma > 0$ — параметр алгоритма.

[1] Vaidya, “A new algorithm for minimizing convex functions over convex sets”

[2] Gladin и др., “Solving smooth min-min and min-max problems by mixed oracle algorithms”

$$\min_{y \in \mathbb{R}^n} f(y). \tag{6}$$

Пусть $\tau > 0$, $y \in \mathbb{R}^n$. Введём обозначение:

$$\operatorname{grad}_f(y, \tau, \mathbf{e}) = \frac{n}{\tau} \left(f(y + \tau \mathbf{e}) - f(y) \right) \mathbf{e}, \tag{7}$$

где $\mathbf{e}$ равномерно распределён на единичной сфере в $\mathbb{R}^n$.

Метод производит **три последовательности:**

- $y_{k+1} := t_k z_k + (1 - t_k) w_k$ with $t_k := \frac{2}{k+2}$
- $w_{k+1} := y_{k+1} - \frac{1}{2L} \mathbf{g}_{k+1}$, где $\mathbf{g}_{k+1} := \operatorname{grad}_f(y_{k+1}, \tau, \mathbf{e}_{k+1})$.
- $z_{k+1} := \arg\min_{z \in \mathbb{R}^n} \left\{ \alpha_{k+1} \langle \mathbf{g}_{k+1}, z - z_k \rangle + \frac{1}{2} \|z_k - z\|_2^2 \right\}$, где $\alpha_k := \frac{k+1}{96 n^2 L}$.

[3] Dvurechensky, Gorbunov и Gasnikov, “An accelerated directional derivative method for smooth stochastic convex optimization”

Предлагаемый подход:
Применять метод Вайды ко внешней задаче (3);
Решать внутреннюю задачу (2) с помощью

- Быстрого градиентного метода в случае оракула 1-го порядка;
- метода ARDD для сильно выпуклых функций в случае оракула 0-го порядка.

Теорема (Гладин)

Описанный подход находит ε–решение задачи (3) после

- $N_x = O\left(n_x \log \frac{1}{\varepsilon}\right)$ вычислений $\partial_x f$,
- $N_y = O\left(n_x \sqrt{\frac{L}{\mu}} \log^2 \frac{1}{\varepsilon}\right)$ вычислений $\nabla_y f$ **или** $n_y \cdot N_y$ вычислений $f(x, y)$.

Цель — создать вредоносный объект $y' = y_0 + y \in \mathbb{R}^{n_y}$, способный обмануть m моделей машинного обучения.

Оптимизационная задача:

$$\max_{y \in \mathbb{R}^{n_y}} \min_{w \in \mathcal{P}} \sum_{i=1}^{m} w_i \mathcal{L}_i(y) - \frac{\gamma}{2} \|y\|_2^2,$$

где $\mathcal{P}$ — вероятностный симплекс, $\mathcal{L}_i$ — функция потерь модели i.

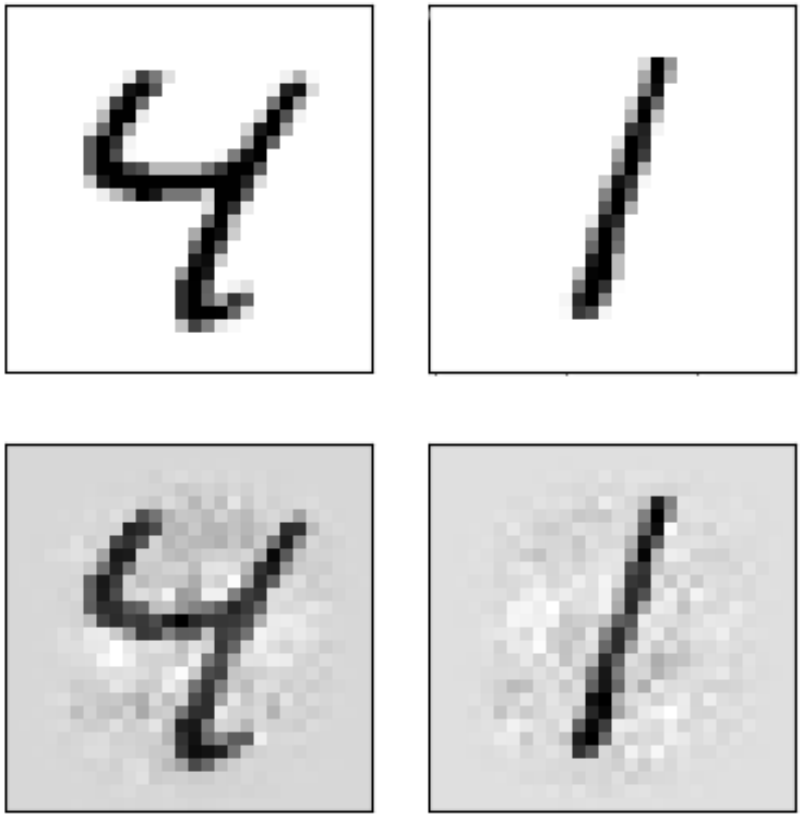
Эксперимент:

- функция потерь $\mathcal{L}$ — кросс-энтропия;
- выборка — MNIST (изображения рукописных цифр);
- атакуемые модели: две логистические регрессии, обученные на разных половинах выборки, и две SVM, обученные на разных половинах выборки.

Результаты эксперимента

Атаки привели к падению точности классификации:

Model	Acc_{orig}	$\text{Acc}_{\text{advers}}$
SVM #1	0.94	0.34
SVM #2	0.94	0.36
LogReg #1	0.96	0.44
LogReg #2	0.96	0.44



Заключение

1. Предложен подход к решению седловых задач, получены его оценки сложности;
2. Разработанный подход накладывает менее жёсткиес требования к задаче, но не уступает существующим методам по скорости сходимости;
3. Эксперименты иллюстрируют потенциал подхода для практических применений.

Выносится на защиту

Статьи:

1. Egor Gladin и Karina Zaynullina. “Ellipsoid method for convex stochastic optimization in small dimension”. в: *arXiv preprint arXiv:2011.04462* (2020)
2. Egor Gladin и др. “Solving smooth min-min and min-max problems by mixed oracle algorithms”. в: *arXiv preprint arXiv:2103.00434* (2021)
3. Egor Gladin, Mohammad Alkousa и Alexander Gasnikov. “On solving convex min-min problems with smoothness and strong convexity in one variable group and small dimension of the other”. в: *arXiv preprint arXiv:2102.00584* (2021)
4. Egor Gladin и др. “Solving strongly convex-concave composite saddle point problems with a small dimension of one of the variables”. в: *arXiv preprint arXiv:2010.02280* (2020)

Конференции:

- Moscow Conference on Combinatorics and Applications 2021;
- Theory and Numerics of Inverse and Ill-posed Problems 2021;
- Quasilinear Equations, Inverse Problems and Their Applications 2020;
- Intelligent Data Processing: Theory and Applications 2020;
- 63^{rd} MIPT Scientific Conference 2020 (award-winning talk);
- 62^{nd} MIPT Scientific Conference 2019 (award-winning talk).