Стохастические методы решения седловых задач и вариационных неравенств

Александр Безносиков

МФТИ

14 апреля 2023

Седловые задачи

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\left\{h_1(x)+f(x,y)-h_2(y)\right\},$$

где $h_1(x):\mathbb{R}^{n_x}\to\overline{\mathbb{R}}$, $h_2(y):\mathbb{R}^{n_y}\to\overline{\mathbb{R}}$ выпуклые функции и $f(x,y):\mathrm{dom}\,h_1\times\mathrm{dom}\,h_2\to\mathbb{R}$, where $\mathrm{dom}\,h_i(x)=\{x:h_i(x)<+\infty\}$ for $i=1,2$.

Как измерять сходимость и качество решения?

- По аргументу $\|x^k - x^*\|^2 + \|y^k - y^*\|^2$ – все аналогично выпуклой оптимизации, самый надежный критерий.
- По функции. Напомню, что в оптимизации измеряли по $f(x^k) - f(x^*)$. Попробуем сконструировать что-то аналогичное для ВН и седловых задач.
- Попробуем: $g(x^k, y^k) - g(x^*, y^*)$. Хорош ли такой критерий?
  Рассмотрим самую простую седловую задачу $\min_x \max_y (x-1)\cdot(y+1)$.
  Решение этой задачи $x = 1, y = -1, g(1,-1) = 0$.
  Пусть начальная точка $x^0 = 0, y^0 = 0, g(0,0) = -1$. Тогда
  $g(x^0, y^0) - g(x^*, y^*)$ отрицательно. Такой критерий не подходит.
- Другой вариант: $g(x^k, y^*) - g(x^*, y^k)$.
  Рассмотрим самую простую седловую задачу $\min_x \max_y (x-1)\cdot(y+1)$.
  Решение этой задачи $x = 1, y = -1, g(1,-1) = 0$. Тогда
  $g(x, y^*) - g(x^*, y) = 0$. Такой критерий не подходит для выпукло-вогнутых седел, но подходит для сильно-выпукло–сильно-вогнутых.

Как измерять сходимость и качество решения?

- Лучший вариант: $\max_y g(x^k, y) - \min_x g(x, y^k)$.
  $\max_y g(x^k, y) \geq g(x^k, y^*) \geq g(x^*, y^*)$
  $\min_x g(x, y^k) \leq g(x^*, y^k) \leq g(x^*, y^*)$
  Тогда $\max_y g(x^k, y) - \min_x g(x, y^k) \geq g(x^k, y^*) - g(x^*, y^k) \geq 0$.
  Если $\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} (x - 1) \cdot (y + 1)$ $\mathcal{X} = \mathbb{R}, \mathcal{Y} = \mathbb{R}$, то
  $\max_y g(x^k, y) - \min_x g(x, y^k) = +\infty$, поэтому вводят еще одно предположение:

Ограниченность множества

$\mathcal{Z}$ – ограниченное, т.е. для любых $z, z' \in \mathcal{Z}$

$$\|z - z'\| \leq D_z.$$

Такое предположение нам понадобится для выпукло-вогнутых седел. В сильно-монотонном случае от него можно отказаться.

Особенности доказательства зеркального спуска в монотонном/выпукло-вогнутом случае

Зеркальный спуск:

$$z^{k+1} = \arg\min_{z \in \mathcal{Z}} \left(\langle \gamma_k G(z^k, \xi^k), z \rangle + V(z, z^k) + \gamma_k h(z) \right)$$

Как оценивать?

$$\mathbb{E}\left[\max_{z \in \mathcal{C}} \left\{ \sum_{k=0}^{K-1} \left[w_k \langle G(z^k, \xi^k) - G(z^k), z \rangle \right] \right\} \right]$$

Особенности доказательства зеркального спуска в монотонном/выпукло-вогнутом случае

Зеркальный спуск:

$$z^{k+1} = \arg\min_{z \in \mathcal{Z}} \left(\langle \gamma_k G(z^k, \xi^k), z \rangle + V(z, z^k) + \gamma_k h(z) \right)$$

Как оценивать?

$$\mathrm{E}\left[\max_{z \in \mathcal{C}} \left\{ \sum_{k=0}^{K-1} \left[w_k \langle G(z^k, \xi^k) - G(z^k), z \rangle \right] \right\} \right]$$

Алгоритм Григориадиса–Хачияна

Для симметричной игры $A = -A^T$, но может быть обобщено на несимметричные игры (трюк Данцига).

Оригинальный алгоритм:

- $p_0 = (\frac{1}{n}, \ldots, \frac{1}{n})$, счетчики $x = (0, \ldots, 0)$, $u = (0, \ldots, 0)$, шаг $\gamma > 0$
- **for** $i = 0, \ldots I - 1$
    - Сэмплируем индекс k_i исходя из p_i
    - Добавляем счетчик выбора/частоты выбора $x_{(k_i)} = x_{(k_i)} + 1$
    - Добавляем счетчик выигрыша $u_j = u_j + A_{jk}$ для всех j
    - Пересчет вероятностей

$$p_j = \frac{p_j \exp(\gamma A_{jk})}{\sum_m p_m \exp(\gamma A_{mk})}.$$

- Цель: $u/I \to 0$
- Вывод: x/I.

Алгоритм Григориадиса–Хачияна

Зеркальный спуск с KL-дивергенцией:

- $p_0 = (\frac{1}{n}, \ldots, \frac{1}{n})$, шаг $\gamma > 0$
- **for** $i = 0, \ldots I - 1$
    - Пересчет вероятностей

$$p_j = \frac{p_j \exp(\gamma \langle A_j, p \rangle)}{\sum p_m \exp(\gamma \langle A_m, p \rangle)}.$$

- Вывод: p.

Алгоритм Григориадиса–Хачияна

Обсуждение:

- Рандомизированный зеркальный спуск
- Координатный зеркальный спуск?
- Importance sampling?

Stampacchia и Minty ВН

Stampacchia Variational Inequality

$$\text{Find } z^* \in \mathbb{R}^n \ : \ \langle F(z^*), z - z^* \rangle + h(z) - h(z^*) \geq 0, \ \text{for all } z \in \mathbb{R}^n.$$

Minty Variational Inequality

$$\text{Find } z^* \in \mathbb{R}^n \ : \ \langle F(z), z - z^* \rangle + h(z) - h(z^*) \geq 0, \ \text{for all } z \in \mathbb{R}^n.$$

Липшицевость

Оператор F L-Липшицев, если для любых $z_1, z_2 \in \operatorname{dom} h$

$$\|F(z_1) - F(z_2)\|_* \leq L\|z_1 - z_2\|.$$

Монотонность

Оператор F монотонный, если для любых $z_1, z_2 \in \operatorname{dom} h$

$$\langle F(z_1) - F(z_2), z_1 - z_2 \rangle \geq 0.$$

Minty = Stampacchia для монотонных и Липшецевых

- Stampacchia $\to$ Minty : $\langle F(z) - F(z^*), z - z^* \rangle \geq 0$

$$\langle F(z), z - z^* \rangle + h(z) - h(z^*) \geq \langle F(z^*), z - z^* \rangle + h(z) - h(z^*) \geq 0.$$

Minty = Stampacchia для монотонных и Липшецевых

- Minty $\to$ Stampacchia : от противного

$$\langle F(z), z - z^* \rangle + h(z) - h(z^*) \geq 0, \text{ for all } z \in \mathbb{R}^n,$$

но существует $\bar{z}$ такой что

$$\langle F(z^*), \bar{z} - z^* \rangle + h(\bar{z}) - h(z^*) = \delta < 0.$$

Minty = Stampacchia для монотонных и Липшецевых

$$\hat{z} = \alpha \bar{z} + (1-\alpha) z^* \text{ with } \alpha = \min\left\{1; -\frac{\delta}{2L\|\bar{z}-z^*\|^2}\right\} \in (0;1]$$

$$\begin{aligned}
\langle F(\hat{z}), \hat{z} - z^* \rangle \quad &+ \quad h(\hat{z}) - h(z^*) \\
&= \alpha \langle F(\hat{z}), \bar{z} - z^* \rangle + h(\alpha \bar{z} + (1-\alpha) z^*) - h(z^*) \\
&\leq \alpha \langle F(\hat{z}), \bar{z} - z^* \rangle + \alpha h(\bar{z}) + (1-\alpha) h(z^*) - h(z^*) \\
&= \alpha \left[\langle F(z^*), \bar{z} - z^* \rangle + h(\bar{z}) - h(z^*) \right] \\
&\quad + \alpha \langle F(\hat{z}) - F(z^*), \bar{z} - z^* \rangle \\
&\leq \alpha \delta + \alpha \| F(\hat{z}) - F(z^*) \| \cdot \| \bar{z} - z^* \| \\
&\leq \alpha \delta + \alpha L \| \hat{z} - z^* \| \cdot \| \bar{z} - z^* \| \\
&= \alpha \delta + \alpha^2 L \| \bar{z} - z^* \|^2 \\
&\leq \frac{\alpha \delta}{2} < 0.
\end{aligned}$$

GDA

- Для негладких задач используется.
- Для гладких задач от него можно отказаться.

GDA

- Дает плохую скорость сходимости в сильно-монотонном случае

GDA

- Пример "расходимости": $\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} xy$

GDA

- Пример не очень частный: 1) седловая задача выпукло-вогнутая - есть сходимость только по функции, 2) выпукло-вогнутые задачи рассматриваются на ограниченных множествах

Alt-GDA

- $x^{k+1} = x^k - \gamma \nabla f(x^k, y^k)$
- $y^{k+1} = y^k + \gamma \nabla f(x^{k+1}, y^k)$

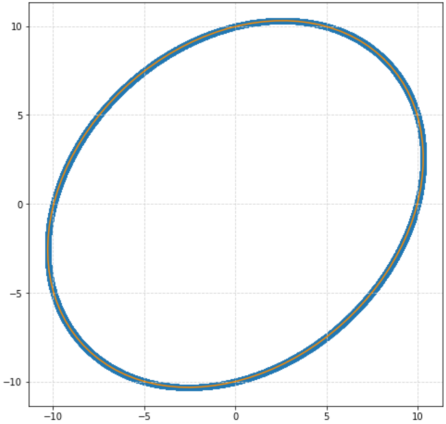
Траектория поведения для $\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} xy$:

$$\frac{(x+y)^2}{2a^2} + \frac{(x-y)^2}{2b^2} = 1, \quad \text{где}$$

$$a^2 = \frac{(x^0-y^0)^2 \cdot (x^0+y^0+\gamma x^0-\gamma y^0-\gamma^2 y^0)^2-(x^0+y^0)^2 \cdot (x^0-y^0-\gamma x^0-\gamma y^0+\gamma^2 y^0)^2}{2(x^0-y^0)^2-2(x^0-y^0-\gamma x^0-\gamma y^0+\gamma^2 y^0)^2},$$

$$b^2 = \frac{(x^0+y^0)^2 \cdot (x^0-y^0-\gamma x^0-\gamma y^0+\gamma^2 y^0)^2-(x^0-y^0)^2 \cdot (x^0+y^0+\gamma x^0-\gamma y^0-\gamma^2 y^0)^2}{2(x^0+y^0)^2-2(x^0+y^0+\gamma x^0-\gamma y^0-\gamma^2 y^0)^2}$$

Alt-GDA



GDA и интуиция EG

Задача:

$$\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} xy$$

Метод GDA:

- $x^{k+1} = x^k - \gamma \nabla f(x^k, y^k)$
- $y^{k+1} = y^k + \gamma \nabla f(x^k, y^k)$

Метод ExtraGradient:

- $x^{k+1/2} = x^k - \gamma \nabla f(x^k, y^k)$
- $y^{k+1/2} = y^k + \gamma \nabla f(x^k, y^k)$
- $x^{k+1} = x^k - \gamma \nabla f(x^{k+1/2}, y^{k+1/2})$
- $y^{k+1} = y^k + \gamma \nabla f(x^{k+1/2}, y^{k+1/2})$