Универсальные методы для стохастических вариационных неравенств

Барсуков Сергей Евгеньевич

Московский физико-технический институт

Курс: Инновационный практикум/Группа 125

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

2024

Задача

In this paper, we study the following optimization problem:

$$F^* := \min_{x \in \operatorname{dom} \psi} \left[F(x) := f(x) + \psi(x) \right], \tag{2.4}$$

where $\psi \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a sufficiently *simple* proper closed convex function, and $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a closed convex function which is finite and subdifferentiable over an open set containing $\operatorname{dom} \psi$.

- ψ - "легко вычисляемая"часть

Базовое решение: градиентный спуск

$$x_{k+1} = \operatorname*{argmin}_{x \in \operatorname{dom} \psi} \left\{ \langle f'(x_k), x \rangle + \psi(x) + \frac{H_k}{2} \|x - x_k\|^2 \right\},$$

- Можно заменить f на ее разложение по Тейлору, а затем запустить обычный градиентный спуск
- Как подбирать H_k?

Line-Search Approach

$H_k \geq \bar{H}_\nu$, where

$$\bar{H}_\nu := L_\nu^{2/(1+\nu)} \left[\frac{1-\nu}{(1+\nu)\epsilon}\right]^{(1-\nu)/(1+\nu)}.$$

- Мы не знаем наилучшего значения L_ν для всех ν, поэтому не можем просто положить $H_k = \overline{H}_* = inf_{\nu \in [0,1]} H_\nu$
- Сделаем аналогично стандартному решению
- На каждом шаге запускаем линейный поиск: стартуем с $H'_{k+1} = \frac{H_k}{2}$, затем удваиваем, пока не выполнится неравенство

$$F(x_{k+1}) - F^* + \frac{H_k}{2} d_{k+1}^2 \leq \frac{H_k}{2} d_k^2 + \beta_{k+1} - \frac{H_k}{2} r_{k+1}^2, \tag{3.2}$$

where $\beta_{k+1} := f(x_{k+1}) - f(x_k) - \langle f'(x_k), x_{k+1} - x_k \rangle$.

Задача стохастической оптимизации

Now we assume that f in problem $\boxed{2.4}$ is accessible only via the *stochastic gradient oracle* $\hat{g}$. Formally, this is a pair (g, ξ) consisting of a random variable ξ and a mapping $g\colon \operatorname{dom} f \times \operatorname{Im} \xi \to \mathbb{R}^n$ (with $\operatorname{Im} \xi$ being the image of ξ). When queried at a point $x \in \operatorname{dom} \psi$, the oracle automatically generates an independent copy ξ of its randomness, and then returns $s = g(x, \xi)$ (notation: $s \sim \hat{g}(x)$)—a random estimate of a subgradient of f at x.

We make the following standard assumption on the oracle:

Assumption 4.1. *The function f in problem $\boxed{2.4}$ is accessible only via an unbiased stochastic gradient oracle $\hat{g} = (g, \xi)$ with bounded variance:*

$$f'(x) := \mathbb{E}_\xi[g(x, \xi)] \in \partial f(x), \tag{4.1}$$

$$\sigma^2 := \sup_{x \in \operatorname{dom} \psi} \mathbb{E}_\xi[\|g(x, \xi) - f'(x)\|_*^2] < +\infty. \tag{4.2}$$

- Градиент заменяется на стохастический градиент, от которого требуется несмещенность и равномерная ограниченность дисперсии

Avoid Line-Search

- Проблема линейного поиска в том, что x_k и H_k в нем скоррелированы, и это мешает обобщить наш подход на стохастические задачи
- Попробуем обойтись без него:

Our main idea now is to choose the next coefficient H_{k+1} so that the two error terms are balanced:

$$\frac{1}{2}(H_{k+1} - H_k)D^2 = \left[\beta_{k+1} - \frac{H_k}{2} r_{k+1}^2\right]_+, \tag{3.9}$$

where we additionally put the positive part $[\cdot]_+$ to respect the monotonicity relation $H_k \leq H_{k+1}$. Recall that β_{k+1} and r_{k+1} depend only on x_k and x_{k+1} (which themselves depend on H_{k-1} and H_k, see (3.1)). Thus, (3.9) is a simple linear equation for H_{k+1} which does not require any line search for solving it.

- Можно доказать, что H_k растут с нужной скоростью, используя те же техники

Universal Line-Search-Free Gradient Method

Algorithm 3.1 Universal Line-Search-Free Gradient Method

1: **Initialize:** $x_0 \in \operatorname{dom}\psi$, diameter $D > 0$, $H_0 := 0$.
2: **for** $k = 0, 1, \dots$ **do**
3: Compute $g_k \in \partial f(x_k)$.
4: $x_{k+1} = \operatorname{argmin}_{x \in \operatorname{dom}\psi} \{\langle g_k, x\rangle + \psi(x) + \frac{H_k}{2}\|x - x_k\|^2\}$.
5: $H_{k+1} := H_k + \dfrac{[\beta_{k+1} - \frac{1}{2}H_k r_{k+1}^2]_+}{D^2 + \frac{1}{2}r_{k+1}^2}$,
where $r_{k+1} := \|x_k - x_{k+1}\|$, $\beta_{k+1} := \beta_f^{g_k}(x_k, x_{k+1})$.
6: **end for**

- По сути, это классический градиентный спуск, но с конкретным правилом выбора размера шага (формула на 5 строке - решение уравнения баланса)
- Метод имеет почти такую же эффективность, но диаметр допустимого множества уменьшается
- Но, в отличие от предыдущего, он легко обобщается на стохастические задачи

Universal Stochastic Gradient Method

Algorithm 4.1 Universal Stochastic Gradient Method

Initialize: $x_0 \in \operatorname{dom}\psi$, $D > 0$, $H_0 := 0$, $g_0 \sim \hat{g}(x_0)$.
for $k = 0, 1, \dots$ **do**
  $x_{k+1} = \operatorname{argmin}_{x \in \operatorname{dom}\psi}\{\langle g_k, x\rangle + \psi(x) + \frac{H_k}{2}\|x - x_k\|^2\}$.
  $g_{k+1} \sim \hat{g}(x_{k+1})$.
  $H_{k+1} := H_k + \dfrac{[\hat{\beta}_{k+1} - \frac{1}{2}H_k r_{k+1}^2]_+}{D^2 + \frac{1}{2}r_{k+1}^2}$,
  where $r_{k+1} = \|x_{k+1} - x_k\|$, $\hat{\beta}_{k+1} = \langle g_{k+1} - g_k, x_{k+1} - x_k\rangle$.
end for

- Просто заменяем g и β на их стохастические аналоги

Вариационные неравенства

- Аналог градиентного метода для вариационных неравенств и седловых задач является экстраградиентный метод, рассмотрим его вариант - проксимальный зеркальный метод.
- Пусть задано векторное поле $g(x)$, если существуют L, δ, т. ч.

$$\langle g(y) - g(x), y - z \rangle \leqslant LV(y, x) + LV(y, z) + \delta.$$

Тогда для проксимального зеркального метода

$$y^{k+1} = \arg\min_{x \in Q} \left\{ \langle g(x^k), x - x^k \rangle + LV(x, x^k) \right\},$$
$$x^{k+1} = \arg\min_{x \in Q} \left\{ \langle g(y^{k+1}), x - x^k \rangle + LV(x, x^k) \right\}$$

имеет место следующая оценка:

$$\frac{1}{N} \sum_{k=1}^{N} \langle g(y^k), y^k - x \rangle \leqslant \frac{LV(x, x^0) - LV(x, x^N)}{N} + \delta.$$

Универсальный вариант

Универсальный проксимальный зеркальный метод

$L^{k+1} = \frac{L^k}{2}$

While True Do

$y^{k+1} = \arg\min_{x \in Q} \left\{ \langle g(x^k), x - x^k \rangle + L^{k+1} V(x, x^k) \right\},$

$x^{k+1} = \arg\min_{x \in Q} \left\{ \langle g(y^{k+1}), x - x^k \rangle + L^{k+1} V(x, x^k) \right\}$

If $\left\{ \langle g(y^{k+1}) - g(x^k), y^{k+1} - x^{k+1} \rangle \leqslant L^{k+1} V(y^{k+1}, x^k) + L^{k+1} V(y^{k+1}, x^{k+1}) + \frac{\varepsilon}{2} \right\}$

Перейти на следующую итерацию: $k \to k+1$

Else

$L^{k+1} := 2L^{k+1}$

Цель работы

- Мы увидели, как строятся универсальные варианты методов для задач оптимизации и вариационных неравенств
- Недавно вышедшая статья обобщила универсальные методы на стохастические задачи
- Наша цель - обобщить проксимальный зеркальный метод (решающий вариационные неравенства) на стохастические вариационные неравенства