Алгоритмы типа зеркального спуска для вариационных неравенств

Владимир Викторович Семёнов

Кафедра вычислительной математики,
факультет компьютерных наук и кибернетики,
Киевский национальный университет имени Тараса Шевченко,
semenov.volodya@gmail.com

27 декабря, 2016

Научный семинар отдела методов негладкой оптимизации Института кибернетики имени В. М. Глушкова НАН Украины, Киев

Вариационные неравенства

Всюду далее E — действительное конечномерное линейное пространство с нормой $\|\cdot\|$, E^*, $\|\cdot\|_*$ — двойственное пространство и норма.

Вариационное неравенство

$$\text{найти } \ x \in C: \ (Ax, y - x) \geq 0 \quad \forall y \in C. \tag{1}$$

(1) $C \subseteq E$ — замкнутое выпуклое;

(2) $A : E \to E^*$ — монотонный и липшицевый (с константой $L > 0$) на C;[1]

(3) множество решений (1) VI не пусто.

Дуальное вариационное неравенство

$$\text{найти } \ x \in C: \ (Ay, x - y) \leq 0 \quad \forall y \in C. \tag{2}$$

Пусть VI^* — множества решений (2). В данной ситуации $VI = VI^*$. В частности, множество VI выпуклое и замкнутое.

[1] $(Ax - Ay, x - y) \geq 0 \ \forall x, y \in C, \ \|Ax - Ay\|_* \leq \|x - y\| \ \forall x, y \in C.$

Вариационные неравенства (примеры)

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая и дифференцируемая функция, $C \subseteq H$ — замкнутое выпуклое множество

$$f(x) = \min_{y \in C} f(y) \;\Leftrightarrow\; x \in C,\ (\nabla f(x), y - x) \geq 0 \;\; \forall y \in C.$$

Пусть $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$ — замкнутые выпуклые множества, $L : X \times Y \to \mathbb{R}$ — выпукло-вогнутая функция.

Седловая задача

найти $(\bar{x}, \bar{y}) \in X \times Y$:

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall (x, y) \in X \times Y.$$

Седловая задача в виде ВН

найти $(\bar{x}, \bar{y}) \in X \times Y$:

$$(\nabla_1 L(\bar{x}, \bar{y}), x - \bar{x}) + (-\nabla_2 L(\bar{x}, \bar{y}), y - \bar{y}) \geq 0 \quad \forall (x, y) \in X \times Y.$$

Пусть $K:\mathbb{R}^n\to\mathbb{R}^m$ — линейный оператор, $f:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$, $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ — собственные замкнутые выпуклые функции.
Задаче

$$\min_{x\in\mathbb{R}^n} f(Kx)+g(x) \tag{3}$$

можно придать форму

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m}(Kx,y)-f^*(y)+g(x). \tag{4}$$

Точнее (4) равносильна паре задач:

$$(primal)\ \ \min_{x\in\mathbb{R}^n} f(Kx)+g(x),\quad (dual)\ \ \max_{y\in\mathbb{R}^m} f^*(y)+g^*(-K^*y).$$

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}\|Kx-b\|_2^2+\varepsilon\|x\|_1 \ \Rightarrow\ \min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m}(Kx,y)-\frac{1}{2}\|y+b\|_2^2+\varepsilon\|x\|_1.$$

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}\|Kx-b\|_2^2+\delta_{\mathbb{R}^n_+}(x) \ \Rightarrow\ \min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m}(Kx,y)-\frac{1}{2}\|y+b\|_2^2+\delta_{\mathbb{R}^n_+}(x).$$

$$(K^*\bar{y},x-\bar{x})+g(x)-g(\bar{x})+(-K\bar{x},y-\bar{y})+f^*(y)-f^*(\bar{y})\geq 0\quad \forall x\,\forall y. \tag{5}$$

Простейший проекционный метод, 1960-е.

$$\begin{cases} x_1 \in C, \\ x_{n+1} = P_C(x_n - \lambda_n A x_n). \end{cases}$$

Экстраградиентный метод (А. Антипин, Г. Корпелевич), конец 1970-х.

$$\begin{cases} x_1 \in C, \\ y_n = P_C(x_n - \lambda_n A x_n), \\ x_{n+1} = P_C(x_n - \lambda_n A y_n). \end{cases}$$

Метод Л. Попова, 1980.

$$\begin{cases} x_1,\ y_1 \in C, \\ x_{n+1} = P_C(x_n - \lambda_n A y_n), \\ y_{n+1} = P_C(x_{n+1} - \lambda_n A y_n). \end{cases}$$

Метод P. Tseng, 2000.

$$\begin{cases} x_1 \in C, \\ y_n = x_n - \lambda_n A x_n, \\ z_n = P_C y_n, \\ r_n = z_n - \lambda_n A z_n, \\ x_{n+1} = P_C(x_n - y_n + r_n). \end{cases}$$

Расстояние Брэгмана (Bregman divergence)

Пусть функция $\varphi : E \to \mathbb{R} \cup \{+\infty\}$ удовлетворяет условия:

- φ — непрерывна и выпукла на C. В частности,
$$C^o = \{x \in C : \partial\varphi(x) \neq \emptyset\} \neq \emptyset;$$
- φ — регулярна на C^o, то есть у субдифференциала $\partial\varphi$ на множестве C^o есть непрерывный селектор $\nabla\varphi$;
- функция φ сильно выпукла относительно выбранной нормы $\|\cdot\|$ с константой сильной выпуклости $\sigma > 0$:
$$\varphi(a) - \varphi(b) \geq (\nabla\varphi(b), a - b) + \frac{\sigma}{2}\|a - b\|^2 \quad \forall a \in C,\ b \in C^o.$$

Соответствующее φ расстояние Брэгмана на множестве C задается формулой

$$d(a, b) = \varphi(a) - \varphi(b) - (\nabla\varphi(b), a - b) \quad \forall a \in C,\ b \in C^o.$$

Имеет место полезное 3-точечное тождество

$$d(a, c) = d(a, b) + d(b, c) + (\nabla\varphi(b) - \nabla\varphi(c), a - b).$$

Из сильной выпуклости φ следует оценка

$$d(a, b) \geq \frac{\sigma}{2}\|a - b\|^2.$$

Расстояние Брэгмана

Два основных примера. При $\varphi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, где $\|\cdot\|_2$ — евклидова норма, имеем

$$d(x,y) = \frac{1}{2}\|x-y\|_2^2.$$

Для стандартного симплекса

$$S_m = \left\{ x \in \mathbb{R}^m : \ x_i \geq 0, \ \sum_i x_i = 1 \right\}$$

и отрицательной энтропии Больцмана-Шеннона

$$\varphi(x) = \sum_i x_i \ln x_i$$

(она 1-сильно выпукла относительно ℓ_1-нормы на S_m) получаем расстояние Кульбака-Лейблера

$$d(x,y) = \sum_i x_i \ln\left(\frac{x_i}{y_i}\right) \quad \forall x \in S_m,\ y \in \mathrm{ri} S_m.$$

Расстояние Брэгмана

Предположим, что у нас есть возможность эффективно решать сильно выпуклые задачи минимизации вида

$$\pi_x(a) = \operatorname{argmin}_{y \in C} \{-(a, y - x) + d(y, x)\} \quad \forall a \in E^*,\ x \in C^o.$$

Точка $\pi_x(a)$ в евклидовом случае совпадает с евклидовой метрической проекцией

$$P_C(x + a) = \operatorname{argmin}_{y \in C} \|y - (x + a)\|_2 .$$

Для случая симплекса S_m и расстояния Кульбака-Лейблера имеем

$$\pi_x(a) = \operatorname{argmin} \left\{ -\sum_{i=1}^m a_i(y_i - x_i) + \sum_{i=1}^m y_i \ln\left(\frac{y_i}{x_i}\right) \ :\ \sum_{i=1}^m y_i = 1,\ y_i \geq 0,\ i = \overline{1, m} \right\},$$

$$\pi_x(a) = \left(\frac{x_1 e^{a_1}}{\sum_{j=1}^m x_j e^{a_j}}, \frac{x_2 e^{a_2}}{\sum_{j=1}^m x_j e^{a_j}}, \dots, \frac{x_m e^{a_m}}{\sum_{j=1}^m x_j e^{a_j}} \right), \quad a \in \mathbb{R}^m,\ x \in \operatorname{ri} S_m.$$

Оператор $\pi_x : E^* \to C^o$ называют прокс-отображением.

Зеркальный спуск для задачи выпуклого программирования

$$f(x) \to \min, \quad x \in C.$$

А.С. Немировский (1979)

Метод зеркального спуска (mirror descent method)

Начиная с $x_1 \in C^o$ генерируем последовательность элементов x_n при помощи итерационной схемы

$$x_{n+1} = \pi_{x_n}\left(-\lambda_n \nabla f(x_n)\right) = \operatorname{argmin}_{y \in C}\left\{(\nabla f(x_n), y - x_n) + \frac{1}{\lambda_n} d(y, x_n)\right\},$$

где $\lambda_n > 0$.

При $\varphi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, то алгоритм принимает вид:

$$x_{n+1} = P_C(x_n - \lambda_n f(x_n)).$$

Еще пара формул.

$$x_{n+1} = \operatorname{argmin}_{y \in C}\left\{(\nabla f(x_n), y - x_n) + \frac{1}{2}(H_n(y - x_n), y - x_n)\right\}, \quad H_n > 0.$$

$$x_{n+1} = x_n - H_n^{-1} \nabla f(x_n).$$

В. Семенов (2016)

2-х этапный «зеркальный спуск»

Начиная с $x_1 \in C^o$, $y_1 \in C$, генерируем последовательность элементов x_n, y_n при помощи итерационной схемы

$$\begin{cases} x_{n+1} = \pi_{x_n}\left(-\lambda A y_n\right), \\ y_{n+1} = \pi_{x_{n+1}}\left(-\lambda A y_n\right), \end{cases}$$

где $\lambda > 0$.

$$x_{n+1} = x_n = y_n \quad \Rightarrow \quad y_n \in VI.$$

При $\varphi(\cdot) = \frac{1}{2}\|\cdot\|_2^2$, то алгоритм принимает вид:

$$\begin{cases} x_{n+1} = P_C(x_n - \lambda A y_n), \\ y_{n+1} = P_C(x_{n+1} - \lambda A y_n). \end{cases}$$

Сходимость

Лемма 1

Для $z \in VI$ и последовательностей (x_n), (y_n) имеет место неравенство

$$d(z, x_{n+1}) \leq d(z, x_n) - \left(1 - \left(1 + \sqrt{2}\right) \frac{\lambda L}{\sigma}\right) d(y_n, x_n) -$$
$$- \left(1 - \sqrt{2}\frac{\lambda L}{\sigma}\right) d(x_{n+1}, y_n) + \frac{\lambda L}{\sigma} d(x_n, y_{n-1}).$$

Теорема 1

Пусть множество $C \subseteq E$ — выпуклое и замкнутое, оператор $A : E \to E^*$ — монотонный и липшицевый с константой $L > 0$, $VI \neq \emptyset$ и

$$\lambda \in \left(0, \left(\sqrt{2} - 1\right) \frac{\sigma}{L}\right).$$

Тогда последовательности (x_n) и (y_n), порожденные алгоритмом, сходятся к некоторой точке $z \in VI$.

Варианты

Если $\sigma = 1$, то можно использовать схему:

$$\begin{cases} x_{n+1} = \pi_{x_n}\left(-\frac{1}{3L}Ay_n\right), \\ y_{n+1} = \pi_{x_{n+1}}\left(-\frac{1}{3L}Ay_n\right). \end{cases}$$

Рассмотрим вариационное неравенство на стандартном симплексе:

$$\text{найти} \quad x \in S_m : \ (Ax, y - x) \geq 0 \quad \forall y \in S_m.$$

Выбирая расстояние Кульбака-Лейблера, получаем следующую версию алгоритма:

$$\begin{cases} x_i^{n+1} = \dfrac{x_i^n e^{-\lambda(Ay_n)_i}}{\sum_{j=1}^m x_j^n e^{-\lambda(Ay_n)_j}}, \quad i = 1, \ldots, m, \\ y_i^{n+1} = \dfrac{x_i^{n+1} e^{-\lambda(Ay_n)_i}}{\sum_{j=1}^m x_j^{n+1} e^{-\lambda(Ay_n)_j}}, \quad i = 1, \ldots, m, \end{cases}$$

где $(Ay_n)_i \in \mathbb{R}$ — i-я координата вектора $Ay_n \in \mathbb{R}^m$, $\lambda > 0$.

Декартово произведение масштабированных симплексов

$$C = \prod_{k=1}^{p} r_k S_{m_k} \subseteq \mathbb{R}^{\sum_{k=1}^{p} m_k},$$

где $r_k S_{m_k} = \{x \in \mathbb{R}^{m_k} : \ x_i \geq 0, \ \ \sum_{i=1}^{m_k} x_i = r_k\}$, $r_k > 0$.

$$\text{найти} \ \ x \in \prod_{k=1}^{p} r_k S_{m_k} \ : \quad (Ax, y - x) \geq 0 \quad \forall\, y \in \prod_{k=1}^{p} r_k S_{m_k}.$$

$$\begin{cases} x_{k,i}^{n+1} = r_k \dfrac{x_{k,i}^{n} e^{-\lambda r_k (Ay_n)_{k,i}}}{\sum_{j=1}^{m_k} x_{k,j}^{n} e^{-\lambda r_k (Ay_n)_{k,j}}}, & k = 1, \ldots, p, \ \ i = 1, \ldots, m_k, \\ y_{k,i}^{n+1} = r_k \dfrac{x_{k,i}^{n+1} e^{-\lambda r_k (Ay_n)_{k,i}}}{\sum_{j=1}^{m_k} x_{k,j}^{n+1} e^{-\lambda r_k (Ay_n)_{k,j}}}, & k = 1, \ldots, p, \ \ i = 1, \ldots, m_k, \end{cases}$$

где $(Ay_n)_{k,i}$ – $\left(\sum_{t=1}^{k-1} m_t + i\right)$-я координата вектора $Ay_n \in \mathbb{R}^{\sum_{k=1}^{p} m_k}$, $\lambda > 0$.

Библиография

- Beck A., Teboulle M., *Mirror Descent and Nonlinear Projected Subgradient Methods for Convex Optimization // Operations Research Letters. – 2003. – Vol. 31, No. 3. – P. 167–175.*

- Lyashko S.I., Semenov V.V., *A New Two-Step Proximal Algorithm of Solving the Problem of Equilibrium Programming. In: Boris Goldengorin (ed.) Optimization and Its Applications in Control and Data Sciences. SOIA, vol. 115, pp. 315–325. Springer International Publishing Switzerland (2016)*

- Брэгман Л.М., *Релаксационный метод нахождения общей точки выпуклых множеств и его применение для решения задач выпуклого программирования // Ж. вычисл. матем. и матем. физ. – 1967, том 7. – №3. – С. 620–631.*

- Немировский А.С., Юдин Д.Б., *Сложность задач и эффективность методов оптимизации. – М.: Наука,1979.*

- Попов Л.Д., *Модификация метода Эрроу-Гурвица поиска седловых точек // Математические заметки. – 1980. – Т. 28, №5. – С. 777–784.*

- Семенов В.В., *Вариант метода зеркального спуска для вариационных неравенств // Кибернетика и системный анализ. — 2017 (принята).*

Благодарю за внимание!


С Новым годом!