О решении вариационных неравенств с оптимизмом

Семёнов Владимир Викторович

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

13 июня 2023
Международный научный семинар
«Моделирование и оптимизация в транспорте и логистике»,
посвященный памяти Дмитрия Ильича Соломона

О чем поговорим?

- Вариационные неравенства
- Обзор алгоритмов:
  - Аналог метода проекции градиента
  - Экстраградиентный метод
  - Схема P. Tseng'a
  - Субградиентный экстраградиентный метод
- Методы с оптимизмом:
  - Метод экстраполяции из прошлого
  - Метод операторной экстраполяции

Вариационные неравенства

Пусть $H = \mathbb{R}^d$, $(\cdot, \cdot)$ — стандартное скалярное произведение в H, $\|\cdot\|$ — соответствующая норма.

Вариационное неравенство (ВН):

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

где $C \subseteq H$, $A : H \to H$.

Пусть:

A1) множество $C \subseteq H$ замкнуто и выпукло;

A2) оператор $A : H \to H$ монотонный и липшицевый (с постоянной $L > 0$) на множестве C:

$$(Ax - Ay, x - y) \geq 0 \quad \forall x, y \in C,$$

$$\|Ax - Ay\| \leq L\|x - y\| \quad \forall x, y \in C;$$

A3) множество решений ВН (1) $S \neq \emptyset$.

Вариационные неравенства

Откуда берутся ВН?

Гладкая задача выпуклого программирования

$$\min_{x \in X} f(x)$$

Предположения: f — гладкая выпуклая функция, X — выпуклое замкнутое множество

Критерий оптимальности:

$$x^* \in X \quad \text{и} \quad (\nabla f(x^*), x - x^*) \geq 0 \quad \forall x \in X$$

Вариационные неравенства

Пожалуй, самый популярный сегодня источник ВН.

Седловая задача

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

Предположения: f — выпукло-вогнутая функция, X, Y — выпуклые замкнутые множества

Критерий оптимальности:

$$\left(\underbrace{\begin{pmatrix} \nabla_x f(x^*, y^*) \\ -\nabla_y f(x^*, y^*) \end{pmatrix}}_{Az^*}, \underbrace{\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}}_{z - z^*} \right) \geq 0 \quad \forall \underbrace{(x, y)}_{z} \in \underbrace{X \times Y}_{Z}$$

Вариационное неравенство

$$\text{найти} \quad z^* \in Z : \quad (Az^*, z - z^*) \geq 0 \quad \forall z \in Z$$

Факт: оператор A монотонный: $(Az_1 - Az_2, z_1 - z_2) \geq 0 \quad \forall z_1, z_2 \in Z$

Вариационные неравенства

1. Матричные игры:
$$\min_{x\in\Delta^n}\max_{y\in\Delta^m}(Px, y)$$

2. Условная оптимизация:
$$\min_x f(x) \text{ при } g(x)\le 0 \quad\longrightarrow\quad \min_x \max_{y\ge 0}(f(x)+yg(x))$$

3. Структурная минимизация:
$$f(x)+g(Lx)\to\min_x \quad\longrightarrow\quad \min_x\max_y(f(x)+(Lx,y)-g^*(y))$$

4. Структурная минимизация (дискретный максимум):
$$\max_{i=1,\dots,m} f_i(x)\to\min_x \quad\longrightarrow\quad \min_x\max_{y\in\Delta^m}\sum_{i=1}^m y_i f_i(x)$$

Вариационные неравенства

5. Machine Learning:

Adversarial training:

$$\min_x \sum_i L(g(x, a_i), b_i) \longrightarrow \min_x \max_{\omega \in \Omega} \sum_i L(g(x, a_i + \omega), b_i),$$

где $\Omega = \{\omega : \ \|\omega\| \leq e\}$.

Generative adversarial networks (GANs).

Это отдельный хороший разговор.

Вариационные неравенства

Задача о неподвижной точке[1]:

$$x = P_C\left(x - \lambda A x\right), \tag{2}$$

где P_C — оператор метрического проектирования на C, $\lambda > 0$.

ВН в форме операторного включения:

$$\text{найти } x \in H: \quad 0 \in Ax + N_C x, \tag{3}$$

где $N_C x$ — нормальный конус множества C в точке x:

$$N_C x = \begin{cases} \{z \in H: \ (z, y - x) \le 0 \ \ \forall y \in C\}, & \text{если } x \in C, \\ \emptyset, & \text{иначе.} \end{cases}$$

[1] Иная форма ВН (1).

Вариационные неравенства

Дуальное вариационное неравенство (неравенство Минти):

$$\text{найти } x \in C: \ (Ay, x - y) \le 0 \quad \forall y \in C. \tag{4}$$

Пусть S^* — множество решений (3). Множество S^* выпукло и замкнуто.

В нашей ситуации (лемма Минти)

$$S = S^*.$$

Вариационные неравенства

Как измерять качество приближенного решения ВН?

1. Если $C = H$, то ВН (1) превращается в операторное уравнение:

$$\text{найти } x \in H : \quad Ax = 0. \tag{5}$$

Тогда все ясно. Используют

$$r(x) = \|Ax\| \quad \forall x \in H.$$

2. Если $C \neq H$ и компактно, то используют неотрицательную функцию зазора (gap function):

$$\mathrm{gap}(z) = \max_{y \in C} (Ay, z - y) \quad \forall z \in C. \tag{6}$$

Для $x \in C$:

$$x \in S \quad \Leftrightarrow \quad \mathrm{gap}(x) = 0.$$

Точку $x \in C$ назовем ε-решением ВН (1), если

$$\mathrm{gap}(x) \leq \varepsilon.$$

Обзор алгоритмов: аналог метода проекции градиента

Простейший «градиентный» метод:

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

Можно гарантировать эргодическая сходимость (G. B. Passty, R. E. Bruck).

Пусть

$$\sum_{n=1}^{\infty} \lambda_n = +\infty, \ \sum_{n=1}^{\infty} \lambda_n^2 < +\infty, \tag{7}$$

$$\text{последовательность } (Ax_n) \text{ ограничена.} \tag{8}$$

Для чезаровских средних имеем

$$z_n = \frac{\sum_{k=1}^n \lambda_k x_k}{\sum_{k=1}^n \lambda_k} \to x^* \in S.$$

Обзор алгоритмов: экстраградиентный метод

Экстраградиентный метод (Г. М. Корпелевич, А. С. Антипин, 1976)

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

Динамическая система с непрерывным временем

$$\dot{x}(t) = P_C(x(t) - \lambda(t) A y(t)) - x(t),$$

$$y(t) = P_C(x(t) - \lambda(t) A x(t)),$$

$$x(0) = x_0 \in H.$$

Обзор алгоритмов: экстраградиентный метод

Лемма 1. Для $z \in S$ и (x_n), (y_n) имеет место неравенство

$$\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 - \left(1 - \lambda_n^2 L^2\right) \|x_n - y_n\|^2 .$$

Теорема 1. Если $\{\inf_n \lambda_n, \sup_n \lambda_n\} \subseteq \left(0, \frac{1}{L}\right)$, то $\|x_n - y_n\| \to 0$ и

$$x_n \to x^* \in S.$$

Теорема 2. При $\lambda = \frac{1}{L}$

$$\operatorname{gap}(z_N) = \max_{y \in C} \left(Ay, z_N - y\right) \le \frac{L \cdot \max_{y \in C} \|y - x_1\|^2}{2N},$$

где $z_N = \frac{1}{N} \sum_{n=1}^{N-1} y_n$.

Обзор алгоритмов: схема P. Tseng'а

A modified forward-backward splitting method (P. Tseng, 2000)

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

Если $C = H$, то экстраградиентный метод совпадает з методом P. Tseng'а.

Обзор алгоритмов: схема P. Tseng'a

Для задачи

$$\text{найти } x \in H: \quad Ax = 0$$

запустим метод

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

Что можно сказать о $\|Ax_n\|$?

Лемма 1

$$\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 - (1 - \lambda^2 L^2)\,\|x_n - y_n\|^2 \quad (z \in A^{-1}0)$$

дает

$$\sum_{n=1}^{N} \|Ax_n\|^2 \le \frac{\|x_1 - z\|^2}{\lambda^2 (1 - \lambda^2 L^2)}, \quad z \in A^{-1}0.$$

Получаем

$$\min_{n=1,\ldots,N} \|Ax_n\|^2 \le \frac{\|x_1 - z\|^2}{\lambda^2 (1 - \lambda^2 L^2)\, N}, \quad z \in A^{-1}0.$$

Обзор алгоритмов: схема P. Tseng'а

Теорема 3 (Э. Горбунов, 2021). При $\lambda \in \left(0, \frac{1}{\sqrt{2}L}\right)$ имеем

$$\|Ax_N\|^2 \leq \frac{\|x_1 - z\|^2}{\lambda^2 (1 - \lambda^2 L^2) N}, \quad z \in A^{-1}0.$$

Обзор алгоритмов: схема P. Tseng'a

Покажем, что

$$\|Ax_{n+1}\| \leq \|Ax_n\| \quad \forall\, n \geq 0.$$

Формулы

$$\begin{aligned} y_n &= x_n - \lambda A x_n, \\ x_{n+1} &= x_n - \lambda A y_n. \end{aligned}$$

Монотонность и липшицевость оператора A дает

$$0 \leq (Ax_n - Ax_{n+1}, x_n - x_{n+1}) \;\Rightarrow\; 0 \leq (Ax_n - Ax_{n+1}, Ay_n),$$

$$0 \leq (Ay_n - Ax_{n+1}, y_n - x_{n+1}) \;\Rightarrow\; 0 \leq (Ay_n - Ax_{n+1}, Ay_n - Ax_n),$$

$$\|Ay_n - Ax_{n+1}\|^2 \leq L^2 \|y_n - x_{n+1}\|^2 \;\Rightarrow\; \|Ay_n - Ax_{n+1}\|^2 \leq \lambda^2 L^2 \|Ay_n - Ax_n\|^2.$$

Умножаем и складываем

$$\frac{3}{2}\|Ay_n - Ax_{n+1}\|^2 \leq 2(Ax_n - Ax_{n+1}, Ay_n) + \frac{1}{2}(Ay_n - Ax_{n+1}, Ay_n - Ax_n) + \frac{3}{2}\lambda^2 L^2 \|Ay_n - Ax_n\|^2.$$

Обзор алгоритмов: схема P. Tseng'а

Раскрываем квадраты норм та группируем

$$\frac{3}{2}\|Ax_{n+1}\|^2 \le \left(2-\frac{1}{2}-3\lambda^2L^2\right)(Ax_n, Ay_n)+\left(-2-\frac{1}{2}+3\right)(Ax_{n+1}, Ay_n)+$$
$$+\left(\frac{1}{2}-\frac{3}{2}+\frac{3\lambda^2L^2}{2}\right)\|Ay_n\|^2+\frac{1}{2}(Ax_{n+1}, Ax_n)+\frac{3\lambda^2L^2}{2}\|Ax_n\|^2 =$$
$$=\left(\frac{3}{2}-3\lambda^2L^2\right)(Ax_n, Ay_n)+\frac{1}{2}(Ax_{n+1}, Ay_n)+\frac{1}{2}(Ax_{n+1}, Ax_n)+$$
$$+\left(\frac{3\lambda^2L^2}{2}-1\right)\|Ay_n\|^2+\frac{3\lambda^2L^2}{2}\|Ax_n\|^2.$$

Применим неравенство $(a, b) \le \frac{1}{2}\|a\|^2+\frac{1}{2}\|b\|^2$

$$\frac{3}{2}\|Ax_{n+1}\|^2 \le \left(\frac{3}{4}-\frac{3\lambda^2L^2}{2}\right)\left(\|Ax_n\|^2+\|Ay_n\|^2\right)+$$
$$+\frac{1}{4}\left(\|Ax_{n+1}\|^2+\|Ay_n\|^2\right)+\frac{1}{4}\left(\|Ax_{n+1}\|^2+\|Ax_n\|^2\right)+$$
$$+\left(\frac{3\lambda^2L^2}{2}-1\right)\|Ay_n\|^2+\frac{3\lambda^2L^2}{2}\|Ax_n\|^2=\|Ax_n\|^2+\frac{1}{2}\|Ax_{n+1}\|^2.$$

Итог

$$\|Ax_{n+1}\| \le \|Ax_n\|.$$

Обзор алгоритмов: субградиентный экстраградиентный метод

Y. Censor, A. Gibali, S. Reich, 2011.

$$\begin{cases} x_1 \in H, \\ y_n = P_C\left(x_n - \lambda_n A x_n\right), \\ x_{n+1} = P_{T_n}\left(x_n - \lambda_n A y_n\right), \end{cases}$$

где

$$T_n = \left\{z \in H : \ \left(x_n - \lambda_n A x_n - y_n, z - y_n\right) \le 0\right\}.$$

Методы с оптимизмом

Минус трех рассмотренных алгоритмов:

$$\texttt{два вызова оператора } A \texttt{ за итерацию.}$$

Побороть эту неприятность можно используя «методы с оптимизмом» (Optimistic Gradient).

Метод экстраполяции из прошлого

Метод Л. Д. Попова (Extrapolation from the Past), 1980

$$\begin{cases} x_1,\ y_0 \in H, \\ y_n = P_C(x_n - \lambda_n A y_{n-1}), \\ x_{n+1} = P_C(x_n - \lambda_n A y_n). \end{cases}$$

Лемма 2.

$$\|x_{n+1} - z\|^2 + 2\lambda_{n+1} L \|x_{n+1} - y_n\|^2 \le \|x_n - z\|^2 + 2\lambda_n L \|x_n - y_{n-1}\|^2 - \\ - (1 - \lambda_n L - 2\lambda_{n+1} L) \|x_{n+1} - y_n\|^2 - (1 - 2\lambda_n L) \|y_n - x_n\|^2,$$

где $z \in S$.

При $\{\inf_n \lambda_n, \sup_n \lambda_n\} \subseteq \left(0, \frac{1}{3L}\right)$ гарантируем сходимость.

Метод экстраполяции из прошлого

Ю. В. Малицкий, В. В. Семенов, 2014.

$$\begin{cases} x_1,\ y_0 \in C, \\ y_n = P_C(x_n - \lambda A y_{n-1}), \\ x_{n+1} = P_{T_n}(x_n - \lambda A y_n), \end{cases}$$

где

$$T_n = \{z \in H : \ (x_n - \lambda A y_{n-1} - y_n, z - y_n) \leq 0\}.$$

Метод экстраполяции из прошлого

Метод Ю. В. Малицкого (Projected Reflected Gradient), 2015

$$\begin{cases} x_1,\ x_0 \in H, \\ y_n = 2x_n - x_{n-1}, \\ x_{n+1} = P_C(x_n - \lambda_n A y_n). \end{cases}$$

Метод операторной экстраполяции

Алгоритм 1.

Init. Задаем $x_0 \in H$, $x_1 \in H$, последовательность (λ_n): $\{\inf_n \lambda_n, \sup_n \lambda_n\} \subseteq \left(0, \frac{1}{2L}\right)$. Полагаем $n = 1$.

1) Вычисляем

$$x_{n+1} = P_C\left(x_n - \lambda_n A x_n - \lambda_{n-1}\left(A x_n - A x_{n-1}\right)\right).$$

2) Если $x_{n-1} = x_n = x_{n+1}$, то stop и x_n — решение. Иначе $n := n + 1$ и переходим на 1.

Метод операторной экстраполяции

Лемма 3. Для порожденной алгоритмом 1 последовательности (x_n) имеет место неравенство

$$\begin{aligned}
&\|x_{n+1}-z\|^2 + 2\lambda_n\left(Ax_n - Ax_{n+1}, x_{n+1}-z\right) + \lambda_n L\|x_{n+1}-x_n\|^2 \leq \\
&\quad \leq \|x_n - z\|^2 + 2\lambda_{n-1}\left(Ax_{n-1} - Ax_n, x_n - z\right) + \lambda_{n-1}L\|x_n - x_{n-1}\|^2 - \\
&\qquad - \left(1 - \lambda_{n-1}L - \lambda_n L\right)\|x_{n+1}-x_n\|^2,
\end{aligned}$$

где $z \in S$.

Метод операторной экстраполяции

Алгоритм 2. Адаптивный метод операторной экстраполяции

Init. Задаем $x_0 \in H$, $x_1 \in H$, $\tau \in \left(0, \frac{1}{2}\right)$ и λ_0, $\lambda_1 > 0$. Полагаем $n = 1$.

1) Вычисляем

$$x_{n+1} = P_C\left(x_n - \lambda_n A x_n - \lambda_{n-1}\left(A x_n - A x_{n-1}\right)\right).$$

2) Если $x_{n-1} = x_n = x_{n+1}$, то stop. Иначе переходим на 3.

3) Вычисляем

$$\lambda_{n+1} = \begin{cases} \min\left\{\lambda_n, \tau \frac{\|x_{n+1} - x_n\|}{\|A x_{n+1} - A x_n\|}\right\}, & \text{если} \quad A x_{n+1} \neq A x_n, \\ \lambda_n, & \text{іиначе} . \end{cases}$$

Полагаем $n := n + 1$ и переходим на 1.

Метод операторной экстраполяции

Лемма 4. Для порожденной алгоритмом 2 последовательности (x_n) имеет место неравенство

$$\left\|x_{n+1}-z\right\|^2+2\lambda_n\left(Ax_n-Ax_{n+1},x_{n+1}-z\right)+\tau\frac{\lambda_n}{\lambda_{n+1}}\left\|x_{n+1}-x_n\right\|^2\leq$$
$$\leq\left\|x_n-z\right\|^2+2\lambda_{n-1}\left(Ax_{n-1}-Ax_n,x_n-z\right)+\tau\frac{\lambda_{n-1}}{\lambda_n}\left\|x_n-x_{n-1}\right\|^2-$$
$$-\left(1-\tau\frac{\lambda_{n-1}}{\lambda_n}-\tau\frac{\lambda_n}{\lambda_{n+1}}\right)\left\|x_{n+1}-x_n\right\|^2, \quad (9)$$

где $z\in S$.

Метод операторной экстраполяции

Пусть $z \in VI(A, C)$. Имеем

$$\|x_{n+1} - z\|^2 \le \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + 2\left(\lambda_n A x_n + \lambda_{n-1}\left(Ax_n - Ax_{n-1}\right), z - x_{n+1}\right). \qquad (10)$$

Из монотонности оператора A следует

$$\begin{aligned}\left(\lambda_n A x_n + \lambda_{n-1}\left(Ax_n - Ax_{n-1}\right), z - x_{n+1}\right) &= \lambda_n \left(Ax_n - Ax_{n+1}, z - x_{n+1}\right) + \\ &+ \lambda_{n-1}\left(Ax_n - Ax_{n-1}, z - x_{n+1}\right) + \underbrace{\lambda_n\left(Ax_{n+1}, z - x_{n+1}\right)}_{\le 0} \le \\ &\le \lambda_n \left(Ax_n - Ax_{n+1}, z - x_{n+1}\right) + \lambda_{n-1}\left(Ax_n - Ax_{n-1}, z - x_n\right) + \\ &+ \lambda_{n-1}\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right). \qquad (11)\end{aligned}$$

Применим (11) для оценки правой части (10) и получим

$$\begin{aligned}\|x_{n+1} - z\|^2 &\le \|x_n - z\|^2 - \|x_{n+1} - x_n\|^2 + 2\lambda_n\left(Ax_n - Ax_{n+1}, z - x_{n+1}\right) + \\ &+ 2\lambda_{n-1}\left(Ax_n - Ax_{n-1}, z - x_n\right) + 2\lambda_{n-1}\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right). \qquad (12)\end{aligned}$$

Метод операторной экстраполяции

Правило пересчета λ_n дает верхнюю оценку для $2\lambda_{n-1}\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right)$ в (12). Имеем

$$
\begin{aligned}
2\lambda_{n-1}\left(Ax_n - Ax_{n-1}, x_n - x_{n+1}\right) &\le 2\lambda_{n-1}\left\|Ax_n - Ax_{n-1}\right\|\left\|x_n - x_{n+1}\right\| \le \\
&\le 2\tau\frac{\lambda_{n-1}}{\lambda_n}\left\|x_n - x_{n-1}\right\|\left\|x_{n+1} - x_n\right\| \le \\
&\le \tau\frac{\lambda_{n-1}}{\lambda_n}\left\|x_n - x_{n-1}\right\|^2 + \tau\frac{\lambda_{n-1}}{\lambda_n}\left\|x_n - x_{n+1}\right\|^2 .
\end{aligned}
$$

Получим

$$
\begin{aligned}
&\left\|x_{n+1} - z\right\|^2 + 2\lambda_n\left(Ax_n - Ax_{n+1}, x_{n+1} - z\right) + \tau\frac{\lambda_n}{\lambda_{n+1}}\left\|x_{n+1} - x_n\right\|^2 \le \\
&\le \left\|x_n - z\right\|^2 + 2\lambda_{n-1}\left(Ax_{n-1} - Ax_n, x_n - z\right) + \tau\frac{\lambda_{n-1}}{\lambda_n}\left\|x_n - x_{n-1}\right\|^2 - \\
&- \left(1 - \tau\frac{\lambda_{n-1}}{\lambda_n} - \tau\frac{\lambda_n}{\lambda_{n+1}}\right)\left\|x_{n+1} - x_n\right\|^2,
\end{aligned}
$$

что и нужно было доказать.

Метод операторной экстраполяции

Теорема 4. Порожденная алгоритмом 2 последовательность (x_n) сходится к решению ВН (1).

Теорема 5. Для алгоритма 1 с $\lambda = \frac{1}{2L}$ для

$$z_{N+1} = \frac{1}{N} \sum_{n=1}^{N} x_{n+1}$$

имеем:

$$\mathrm{gap}(z_{N+1}) \le \frac{L \max_{y \in C} \|y - x_1\|^2}{N}.$$

Благодарю за внимание!