Негладкий метод Ньютона и его применение в линейном программировании

Долгополик Максим Владимирович

Институт Проблем Машиноведения Российской Академии Наук,
Санкт-Петербург

19 октября 2023 г.

Метод Мангасаряна

- Managasrian, O.L. A Newton method for linear programming // Journal of Optimization Theory and Applications. — 2004. — Vol. 121. — No. 1. — pp. 1–18.
- Некоторые задачи машинного обучения (линейный метод опорных векторов) сводятся к задам линейного программирования, в которых число ограничений намного больше числа переменных.
- Метод Мангасаряна для таких задач: квадратичная штрафная функция + негладкий (полугладкий) метод Ньютона + особенности задач линейного программирования.

Содержание доклада

1. Метод Ньютона для негладких уравнений
2. Метод Ньютона для минимизации полугладких функций
3. Приложение в линейном программировании

Содержание доклада

Метод Ньютона

Рассмотрим систему нелинейных уравнений

$$F(x) = 0, \tag{1}$$

где $F\colon \mathbb{R}^n \to \mathbb{R}^n$. Если F непрерывно дифференцируема, то

$$x_{k+1} = x_k - [F'(x_k)]^{-1} F(x_k). \tag{2}$$

На каждой итерации необходимо решить систему линейных уравнений:

$$F'(x_k)h = F(x_k), \quad h \in \mathbb{R}^n. \tag{3}$$

Что делать, если F — негладкая функция?
Желательно, чтобы на каждой итерации нужно было решать лишь одну систему линейных уравнений (ср.[1,2]).

Ответ: Qi, L., Sun, J. A nonsmooth version of Newton's method // Mathematical programming. — 1993. — Vol. 58. — pp. 353–367.

[1] Pang, J.S. Newton's method for B-differentiable equations // Mathematics of Operations Research. — 1990. — Vol. 15.—311–341.

[2] Demyanov, V.F. Fixed point theorem in nonsmooth analysis and its applications // Numerical functional analysis and optimization. — 1995. — Vol. 16. — No. 1–2. — pp. 53–109.

Обобщённый Якобиан[3]

Пусть F — локально липшицева. Обобщённый Якобиан функции F в точке x:

$$\partial F(x) = \mathrm{co}\left\{ \lim_{x_i \to x,\, x_i \in D_F} F'(x_i) \right\},$$

где D_F — множество точек дифференцируемости функции F. $\partial F(x)$ — непустое компактное выпуклое множество матриц.

[3] Кларк, Ф. Оптимизация и негладкий анализ. — М.: Наука, 1988.

Обобщённый Якобиан

Пример: пусть задана функция $f(x) = 0.5[\langle a, x\rangle + b]_+^2$. Её градиент $F(x) = \nabla f(x) = [\langle a, x\rangle + b]_+ a$. Тогда

$$\partial F(x) = \begin{cases} aa^T, & \text{если } \langle a, x\rangle + b > 0, \\ \mathbb{O}_{n\times n}, & \text{если } \langle a, x\rangle + b < 0, \\ \mathrm{co}\{\mathbb{O}_{n\times n}, aa^T\}, & \text{если } \langle a, x\rangle + b = 0. \end{cases}$$

Обобщённый Якобиан

Обобщённый Якобиан **нельзя** вычислять построчно. Пусть

$$F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix} = \begin{pmatrix} [x]_+ \\ [-x]_+ \end{pmatrix}, \quad x \in \mathbb{R}.$$

Тогда

$$\partial F_1(0) = [0, 1], \quad \partial F_2(0) = [-1, 0],$$

но

$$\partial F(0) \neq \begin{pmatrix} \partial F_1(0) \\ \partial F_2(0) \end{pmatrix} = \left\{ \begin{pmatrix} t \\ -s \end{pmatrix} \,\middle|\, t, s \in [0, 1] \right\}, \tag{4}$$

поскольку

$$\partial F(0) = \operatorname{co} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\}.$$

Негладкий метод Ньютона

Классический метод Ньютона:

$$x_{k+1} = x_k - [F'(x_k)]^{-1} F(x_k).$$

Негладкий метод Ньютона:

$$x_{k+1} = x_k - [V_k]^{-1} F(x_k)$$

для произвольной матрицы $V_k \in \partial F(x_k)$. На каждой итерации требуется решить систему линейных уравнений:

$$V_k h = F(x_k), \quad h \in \mathbb{R}^n. \tag{5}$$

Сходимость: $C^1 \subset ? \subset Lip$.

Полугладкие функции[4]

Локально липшицевая функция F называется **полугладкой** в точке x, если для любого $h \in \mathbb{R}^n$ и для любых последовательностей $h_k \to h$, $t_k \to +0$ и $V_k \in \partial F(x + t_k h_k)$ существует предел

$$\lim_{k \to \infty} V_k h_k,$$

не зависящий от выбора соответствующих последовательностей.

Класс полугладких функций — SC.

[4] Mifflin, R. Semismooth and semiconvex functions in constrained optimization // SIAM Journal on Control and Optimization. — 1977. — Vol. 15. — pp. 957–972.

Полугладкие функции

Теорема 1

Пусть $f\colon \mathbb{R}^n \to \mathbb{R}$. Справедливы следующие утверждения:

1. *если $f \in C^1$, то $f \in SC$;*
2. *если f — выпуклая/вогнутая функция, то $f \in SC$;*
3. *если f — максимум/минимум конечного числа непрерывно дифференцируемых функций, то $f \in SC$;*
4. *суперпозиция полугладких функций является полугладкой функцией.*

Теорема 2

Если каждая компонента функции $F\colon \mathbb{R}^n \to \mathbb{R}^n$ является полугладкой, то и сама функция F является полугладкой.

Полугладкие функции

Класс полугладких функций SC: $C^1 \subset SC \subset Lip$.

Пример локально липшицевой функции, не являющейся полугладкой:

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{если } x \neq 0, \\ 0, & \text{если } x = 0. \end{cases}$$

Класс SC^1 функций, градиент которых является полугладким, определяется схожим образом и $C^2 \subset SC^1 \subset C^1$.

Негладкий (полугладкий) метод Ньютона

Теорема 3 (локальная сходимость)

Пусть $F(x^*) = 0$ *и функция* F *является полугладкой в окрестности точки* x^*. *Предположим также, что*

$$\det V \neq 0 \quad \forall V \in \partial F(x^*).$$

Тогда существует окрестность $\mathcal{U}_{x^*}$ *точки* x^* *такая, что для любого начального приближения* $x_0 \in \mathcal{U}_{x^*}$ *последовательность, построенная по негладкому методу Ньютона, корректно определена и сходится к точке* x^* *со сверхлинейной скоростью.*

При некоторых дополнительных предположениях можно гарантировать, что скорость сходимости квадратичная.

Негладкий (полугладкий) метод Ньютона

Теорема 4 (аналог теоремы Канторовича)

Пусть F является полугладкой на множестве $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le r\}$. Предположим, что для всех $x, y \in S$ и $V \in \partial F(x)$ матрицы V не вырождены и справедливы оценки:

$$\|V^{-1}\| \le \beta, \quad \|V(y-x) - F'(x; y-x)\| \le \gamma \|y - x\|, \tag{6}$$

$$\|F(y) - F(x) - F'(x; y-x)\| \le \delta \|y - x\|, \tag{7}$$

для некоторых $\beta, \gamma, \delta > 0$, где $\alpha = \beta(\gamma + \delta) < 1$ и $\beta \|F(x_0)\| \le r(1 - \alpha)$. Тогда последовательность, построенная по негладкому методу Ньютона не выходит из множества S и сходится к единственному решению $x^* \in S$ уравнения $F(x) = 0$. Кроме того, справедлива оценка:

$$\|x_k - x^*\| \le \frac{\alpha}{1 - \alpha} \|x_k - x_{k-1}\| \quad \forall k \in \mathbb{N}.$$

Содержание доклада

Минимизация выпуклых SC^1 функций

Рассмотрим задачу

$$f(x) \to \min_{x \in \mathbb{R}^n},$$

где $f\colon \mathbb{R}^n \to \mathbb{R}$ — выпуклая непрерывно дифференцируемая функция. Данная задача эквивалентна задаче нахождения решения системы нелинейных уравнений:

$$F(x) = \nabla f(x) = 0.$$

Если $f \in SC^1$ (т.е. F — полугладкая функция), то можно применить негладкий метод Ньютона:

$$x_{k+1} = x_k - [V_k]^{-1} \nabla f(x_k), \quad V_k \in \partial F(x_k) = \partial^2 f(x_k).$$

Здесь $\partial^2 f(x_k)$ — обобщённый гессиан функции f.[5]

[5] Hiriart-Urruty, J.B., Strodiot, J.J., Nguyen, V.H. Generalized Hessian Matrix and Second-Order Optimality Conditions for Problems with $C^{1,1}$ Data // Applied Mathematics and Optimization. — 1984. — Vol. 11. — pp. 43–56.

Полугладкий метод Ньютона[6]

1. Выберем начальное приближение $x_0 \in \mathbb{R}^n$ и параметры $\rho, \sigma \in (0, 1)$.
2. k-я итерация:
   1. Вычислим $V_k \in \partial^2 f(x_k)$ и выберем $\delta_k > 0$. Найдём решение d_k системы линейных уравнений

      $$(V_k + \delta_k I_n)d = \nabla f(x_k),$$

      где I_n — единичная матрица.
   2. Найдём наименьшее неотрицательное целое число m_k такое, что

      $$f(x_k + \rho^{m_k} d_k) - f(x_k) \leq \sigma \rho^{m_k} \langle \nabla f(x_k), d_k \rangle.$$

      Положим $x_{k+1} = x_k + \rho^{m_k} d_k$.

[6] Pang, J.S., Qi, L. A globally convergent Newton method for convex SC^1 minimization problem // Journal of Optimization Theory and Applications. — 1995. — Vol. 85. — No. 3. — pp. 633–648.

Полугладкий метод Ньютона

Теорема 5

Пусть f — выпуклая функция класса SC^1, последовательность $\{\delta_k\}$ ограничена и существует $\mu > 0$ такое, что

$$\langle d, (V_k + \delta_k I_n) d \rangle \geq \mu \|d\|^2 \quad \forall d \in \mathbb{R}^n,\ k \in \mathbb{N}.$$

Тогда все предельные точки последовательности $\{x_k\}$, построенной по полугладкому методу Ньютона, являются точками глобального минимума функции f.

Полугладкий метод Ньютона

Теорема 6

Пусть последовательность $\{x_k\}$*, построенная по полугладкому методу Ньютона, сходится к точке* x^**. Предположим, что* $\delta_k \to 0$ *при* $k \to \infty$ *и все матрицы из* $\partial^2 f(x^*)$ *не вырождены. Тогда:*

1. x^* *— единственная точка глобального минимума функции* f*;*
2. *существует* $k_0 \in \mathbb{N}$ *такое, что* $m_k = 0$ *для всех* $k \geq k_0$*;*
3. *последовательность* $\{x_k\}$ *сходится к* x^* *со сверхлинейной скоростью.*

Содержание доклада

Негладкий метод Ньютона для линейного программирования

Рассмотрим задачу линейного программирования:

$$c^T x \to \min_{x \in \mathbb{R}^n}, \quad Ax \leq b, \tag{8}$$

где $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ и $m \gg n$.
Квадратичная штрафная функция для данной задачи:

$$f(x) = \varepsilon c^T x + \frac{1}{2}\|[Ax - b]_+\|^2, \quad \varepsilon > 0.$$

Её градиент:

$$F(x) := \nabla f(x) = \varepsilon c + A^T [Ax - b]_+. \tag{9}$$

Функция $f \in SC^1$. Поэтому можно воспользоваться полугладким методом Ньютона.

Обобщённый гессиан штрафной функции

Для любого $x \in \mathbb{R}^n$ обозначим через $x_* \in \mathbb{R}^n$:

$$x_*[i] = \partial[\cdot]_+(x[i]) = \begin{cases} 1, & \text{если } x > 0, \\ [0, 1], & \text{если } x = 0, \\ 0, & \text{если } x < 0. \end{cases}$$

Тогда[7,8]

$$\partial^2 f(x) = A^T \operatorname{diag}((Ax - b)_*)A. \tag{10}$$

---

[7] Managasrian, O.L. A Newton method for linear programming // Journal of Optimization Theory and Applications. — 2004. — Vol. 121. — No. 1. — pp. 1–18.

[8] Mangasarian, O.L. A finite Newton method for classification // Optimization Methods and Software. — 2002. — Vol. 17. — No. 5. — pp. 913–929.

Обобщённый гессиан штрафной функции

Пусть $D(x)$ — диагональная матрица, такая, что

$$D(x)[i,i] = \begin{cases} 1, & \text{если } (Ax-b)[i] > 0, \\ 0, & \text{если } (Ax-b)[i] \le 0. \end{cases}$$

Тогда[9]

$$A^T D(x) A \in \partial^2 f(x)$$

[9] Голиков, А.И., Евтушенко, Ю.Г., Кетабчи, С. О семействах гиперплоскостей разделяющих полиэдры // Ж. вычисл. матем. и матем. физ. — 2005. — Том. 45. — № 2. — с. 238–253.

Контрпример

Рассмотрим систему неравенств в $\mathbb{R}^2$:

$$x[1] \le 0, \quad -x[1] \le 0.$$

В этом случае $A = \left(\begin{smallmatrix} 1 & 0 \\ -1 & 0 \end{smallmatrix}\right)$. Пусть $c = 0$.
Имеем

$$F(x) := \nabla f(x) = A^T[Ax - b]_+ = \begin{pmatrix} [x[1]]_+ - [-x[1]]_+ \\ 0 \end{pmatrix} = \begin{pmatrix} x[1] \\ 0 \end{pmatrix}.$$

Поэтому $F'(x) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ для всех x. Таким образом,

$$A^T D(x) A = \mathbb{O}_{2\times 2} \notin \partial^2 f(x) \text{ при } x[1] = 0.$$

Обобщённый гессиан штрафной функции

Лемма 1

$A^T D(x) A \in \partial^2 f(x)$ *для любого* $x \in \mathbb{R}^n$, *если множество планов имеет непустую внутренность.*

Поэтому будем предполагать, что множество планов имеет непустую внутренность.

Полугладкий метод Ньютона

Полугладкий метод Ньютона для $f(y) = \varepsilon c^T y + \frac{1}{2}\|[Ay - b]_+\|^2$.

1. Выберем начальное приближение $y_0 \in \mathbb{R}^n$ и параметры $\varepsilon, \delta > 0$.
2. k-я итерация:
   1. Найдём решение d_k системы линейных уравнений
   $$(A^T D(y_k)A + \delta I_n)d = \varepsilon c + A^T [Ay_k - b]_+.$$
   2. Найдём наибольшее $\lambda_k \in \{1, 1/2, 1/4, \dots\}$ такое, что
   $$f\left(y_k + \lambda_k d_k\right) - f(y_k) \leq \frac{\lambda_k}{4}\langle \varepsilon c + A^T [Ay_k - b]_+, d_k\rangle.$$

   Положим $y_{k+1} = y_k + \lambda_k d_k$.

Двойственная задача

Исходная задача:

$$-c^T x \to \max_{x \in \mathbb{R}^n}, \quad Ax \leq b, \tag{11}$$

Двойственная задача:

$$b^T u \to \min_{u \in \mathbb{R}^m}, \quad A^T u + c = 0, \quad u \geq 0. \tag{12}$$

Вспомогательная задача:

$$b^T v + \frac{\varepsilon}{2} \|v\|^2 \to \min_{v \in \mathbb{R}^m}, \quad A^T v + c = 0, \quad v \geq 0. \tag{13}$$

Анализ условий Куна-Таккера

Существуют $y \in \mathbb{R}^n$ и $\mu \in \mathbb{R}^m$ такие, что

$b + \varepsilon v - Ay + \mu = 0, \quad A^T v + c = 0, \quad v \geq 0, \quad \mu \leq 0, \quad v[i]\mu[i] = 0.$

Если $v[i] \neq 0$, то $\mu[i] = 0$ и

$$v[i] = \frac{1}{\varepsilon}(Ay - b)[i] > 0.$$

Если же $v[i] = 0$, то $\mu[i] = (Ay - b)[i] \leq 0$. Таким образом,

$$v[i] = \begin{cases} \frac{1}{\varepsilon}(Ay - b)[i], & \text{если } (Ay - b)[i] > 0, \\ 0, & \text{если } (Ay - b)[i] \leq 0 \end{cases} = \frac{1}{\varepsilon}[Ay - b]_+[i].$$

При этом $\mu = [Ay - b]_-$.

Анализ условий Куна-Таккера

Условия Куна-Таккера:

$$v = \frac{1}{\varepsilon}[Ay - b]_+, \quad A^T[Ay - b]_+ + \varepsilon c = 0. \qquad (14)$$

Заметим, что

$$\nabla f(y) = \varepsilon c + A^T[Ay - b]_+.$$

Лемма 2

Вектор $y \in \mathbb{R}^n$ является точкой глобального минимума штрафной функции f тогда и только тогда, когда вектор $v = \frac{1}{\varepsilon}[Ay - b]_+$ является оптимальным планом задачи

$$b^T v + \frac{\varepsilon}{2}\|v\|^2 \to \min_{v \in \mathbb{R}^m}, \quad A^T v + c = 0, \quad v \geq 0. \qquad (15)$$

Решение с наименьшей нормой[10,11]

Теорема 7

Существует $\overline{\varepsilon} > 0$ *такое, что для любого* $\varepsilon \in [0, \overline{\varepsilon}]$ *оптимальный план* v *задачи*

$$b^T v + \frac{\varepsilon}{2}\|v\|^2 \to \min_{v \in \mathbb{R}^m}, \quad A^T v + c = 0, \quad v \geq 0. \tag{16}$$

является оптимальным планом двойственной задачи

$$b^T u \to \min_{u \in \mathbb{R}^m}, \quad A^T u + c = 0, \quad u \geq 0. \tag{17}$$

с наименьшей евклидовой нормой.

---

[10] Mangasarian, O.L., Meyer, R.R. Nonlinear perturbation of linear programs // SIAM J. on Control and Optimization. — 1979. — Vol. 17. — pp. 745–752.

[11] Голиков, А.И., Евтушенко, Ю.Г. Отыскание нормальных решений в задачах линейного программирования // Ж. вычисл. матем. и матем. физ. — 2000. — Том. 40. — № 12. — с. 1766–1786.

Решение с наименьшей нормой

Рассмотрим задачу:

$$\frac{1}{2}\|w\|^2 \to \min_{w \in \mathbb{R}^n}, \quad A^T w + c = 0, \quad w \geq 0, \quad b^T w \leq \theta, \tag{18}$$

где θ — оптимальное значение двойственной задачи. Её оптимальный план — это оптимальный план двойственной задачи с наименьшей нормой.

Доказательство теоремы 7 основано на одновременном анализе условий Куна-Таккера для задачи (18) и двойственной задачи.

Метод Мангасаряна

1. Выбрать достаточно малое $\varepsilon > 0$ (обычно $\varepsilon \sim 10^{-3}$).
2. Найти точку минимума y^* штрафной функции $f(y) = \varepsilon c^T y + \frac{1}{2}\|[Ay - b]_+\|^2$ с помощью полугладкого метода Ньютона.
3. Вычислить вектор $v^* = \frac{1}{\varepsilon}[Ay^* - b]_+$. Если $\varepsilon > 0$ достаточно мало, то v^* — оптимальный план двойственной задачи.
4. Воспользовавшись условием дополнительности

$$v^*[j](Ax^* - b)[j] = 0 \quad \forall j,$$

найти оптимальный план прямой задачи, решив систему уравнений:

$$A_j x^* = b[j], \quad j \in S := \{j \mid v^*[j] > 0\}.$$

Метод Мангасаряна

- Managasrian, O.L. A Newton method for linear programming // Journal of Optimization Theory and Applications. — 2004. — Vol. 121. — No. 1. — pp. 1–18.

Спасибо за внимание!