ЛЕКЦИЯ № 14

Численные методы нелинейного программирования

1. Градиентный метод

2. Теоремы сходимости

3. Метод Такахаши (дуализация/градиен-тный метод)

Численные методы НЛП

Задача поиска безусловного минимума:

$$f(x) \longrightarrow \min_{x \in R^n}.$$

Для решения используются численные методы, в которых текущее приближение вычисляется по формуле

$$x^{k+1} = x^k + \alpha_k p^k,$$

где p^k — направление спуска, α_k — длина шага вдоль этого направления.

Численные методы НЛП

Методы, в которых последовательность векторов $x^0, x^1, \ldots, x^k, \ldots,$ удовлетворяет условию

$$f(x^0) \geq f(x^1) \geq \ldots \geq f(x^k) \geq \ldots$$

называются релаксационными.

Пусть x^*– минимум функции $f(x)$. Скорость сходимости линейная: если для $k = 0, 1, \ldots$

$$\|x^{k+1} - x^*\| \leq q \, \|x^k - x^*\|, 0 < q < 1,$$

или говорят, что метод сходится со скоростью геометрической прогрессии, т.к.

$$\|x^k - x^*\| \le q^k \, \|x^0 - x^*\|.$$

Скорость сходимости сверхлинейна, если

$$\|x^{k+1} - x^*\| \le q_k \, \|x^k - x^*\|,$$

где $q_k \longrightarrow 0$ при $k \longrightarrow \infty,$ и квадратична, если

$$\|x^{k+1} - x^*\| \le C \, \|x^k - x^*\|^2, C \ge 0.$$

Численные методы НЛП

Метод нулевого порядка, если в процессе вычислений используются только значения целевой функции.

В методах первого порядка, помимо значений целевой функции используются её производные.

и так далее.

Численные методы НЛП

Все итерационные процессы, в которых направление движения на каждом шаге совпадает с антиградиентом (градиентом) функции, называются градиентными методами:

$$x^{k+1} = x^k - \alpha_k f'(x_k), \alpha_k \geq 0.$$

Методы отличаются способами выбора длины шага α_k.

Численные методы НЛП

Метод с постоянным шагом: $\alpha_k = \alpha$.

Метод с дроблением шага: на каждом шаге проверяется неравенство

$$f(x^k - \alpha_k f'(x^k)) - f(x^k) \leq -\epsilon\, \alpha_k \|f'(x^k)\|^2,$$

где $0 < \epsilon < 1$.

Метод наискорейшего спуска: при выборе α_k минимизируется по α функция $f(x^k - \alpha\, f'(x^k))$:

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x^k - \alpha\, f'(x^k)).$$

Теорема 17 (Первая теорема сходимости)

Пусть функция $f \in C^1(R^n)$, ограничена снизу $f(x) \geq f^* > -\infty$, выполняется условие Липшица для градиента $f'(x)$:

$$\|f'(x) - f'(y)\| \leq L\,\|x - y\|$$

и длина шага α удовлетворяет условию $0 < \alpha < 2/L$. Тогда $f'(x^k) \to 0$ при $k \to \infty$ и $f(x^{k+1}) \leq f(x^k)$, при любом выборе начального приближения x_0.

Доказательство. Воспользуемся формулой конечных приращений

$$\boldsymbol{f(x+y) = f(x) + \int\limits_0^1 \langle f'(x+\tau y), y\rangle \, d\tau,}$$

прибавим и вычтем из правой части величину $\langle \boldsymbol{f'(x), y} \rangle$:

Численные методы НЛП

$$f(x+y) = f(x)+\langle f'(x), y\rangle+\int\limits_0^1 \langle f'(x+\tau y)-$$

$$-f'(x), y\rangle \, d\tau.$$

Подставим вместо x и y, соответственно, x^k и $-\alpha f'(x^k)$. Имеем

$$f(x^{k+1}) \le f(x^k) + \langle f'(x^k), -\alpha f'(x^k)\rangle+$$

$$+\int\limits_0^1 |\langle f'(x^k - \tau\alpha f'(x^k)) - f'(x^k),$$

$$-\alpha f'(x^k)\rangle| d\tau \leq \ldots$$

Первое неравенство следствие появления модуля под интегралом. Следующее неравенство вытекает из неравенства Коши-Буняковского $|\langle a, b\rangle| \leq \|a\|$ $\|b\|$

Численные методы НЛП

$$\dots \le f(x^k) - \alpha\,\|f'(x^k)\|^2 + \int\limits_0^1 \|f'(x^k -$$

$$-\tau\alpha f'(x^k)) - f'(x^k)\|\|\alpha f'(x^k)\|d\tau \le \dots$$

из условия Липшица имеем

$$\le f(x^k) - \alpha\,\|f'(x^k)\|^2 +$$

$$+ \int\limits_0^1 L\|\tau\alpha f'(x^k)\|\;\|\alpha f'(x^k)\|d\tau =$$

Численные методы НЛП

упрощая, получим

$$= f(x^k) - \alpha\|f'(x^k)\|^2 + L\alpha^2\,\|f'(x^k)\|^2 \int_0^1 \tau d\tau =$$

$$= f(x^k) - \alpha(1 - L\alpha/2)\|f'(x^k)\|^2 =$$
$$= f(x^k) - \gamma\|f'(x^k)\|^2$$

$(\gamma = \alpha(1 - L\alpha/2)) \Longrightarrow$ и условия $\gamma > 0$ имеем

$$f(x^{k+1}) \leq f(x^k).$$

Индукцией легко доказать $\forall s$

$$f(x^{s+1}) \leq f(x^0) - \gamma \sum_{k=0}^{s} \|f'(x^k)\|^2.$$

Тогда, учитывая ограниченность функции $f(x) \geq f^* > -\infty$ получим

$$\sum_{k=0}^{s} \|f'(x^k)\|^2 \le (f(x^0) - f(x^{s+1}))/\gamma \le$$

$$\le (f(x^0) - f^*)/\gamma.$$

$$\Longleftrightarrow f'(x^k) \to 0 \text{ при } k \to \infty. \blacksquare$$

Определение Дифференцируемая функция $\boldsymbol{f}$ называется сильно выпуклой (с константой $\boldsymbol{l > 0}$), если для любых $\boldsymbol{x}$ и $\boldsymbol{y}$ из $\boldsymbol{R^n}$ справедливо

$$\boldsymbol{f(x+y) \geq f(x) + \langle f'(x), y \rangle + l\|y\|^2/2.} \tag{23}$$

Лемма 12. Если функция $\boldsymbol{f}$ является сильно выпуклой (с константой $\boldsymbol{l > 0}$), то она имеет глобальный минимум на $\boldsymbol{R^n}$.

Доказательство. Перепишем (23), используя

Градиентные методы

неравенство Коши - Буняковского

$$f(x+y) \geq f(x) - \|f'(x)\|\|y\| + l\|y\|^2/2.$$

Вынесем величину $l\|y\|/2$ за скобки

$$f(x+y) \geq f(x) + l\|y\|/2(\|y\| - 2\|f'(x)\|/l).$$

$\Longrightarrow \forall y : \|y\| > r = 2\|f'(x)\|/l$ имеем неравенство $f(x+y) > f(x)$, из которого следует, что минимум $\exists$ и достигается на шаре $B(x, r)$. $\blacksquare$

Лемма 13. Если функция f является сильно выпуклой (с константой $l > 0$) и x^* — ее глобальный минимум, то для любого $x \in R^n$ выполняется неравенство $\|f'(x)\|^2 \geq 2l(f(x) - f(x^*))$.

Доказательство. Подставим в неравенство (23) вместо y вектор $x^* - x$. Получим

$$f(x) - f(x^*) + \langle f'(x), x^* - x \rangle +$$
$$+ l\|x^* - x\|^2/2 \leq 0.$$

Градиентные методы

$$\|f'(x)/\sqrt{2l} + \sqrt{l/2}(x^* - x)\|^2 =$$

$$\|f'(x)\|^2/2l + \langle f'(x), x^* - x\rangle + l\|x^* - x)\|^2/2 \geq$$

$$\geq 0 \geq f(x) - f(x^*) + \langle f'(x), x^* - x\rangle + \\ + l\|x^* - x\|^2/2.$$

Градиентные методы

После приведения подобных членов получим требуемое неравенство:

$$\|f'(x)\|^2 \geq 2l(f(x) - f(x^*)).\blacksquare$$

Теорема 18 (Вторая теорема сходимости) Пусть функция f дифференцируема в R^n, является сильно выпуклой, выполняется условие Липшица для градиента $f'(x) : \|f'(x) - f'(y)\| \leq L\|x - y\|$ и длина шага α удовлетворяет условию $0 < \alpha < 2/L$.

Тогда $x^k \to x^*$ при $k \to \infty$ и $\|x^k - x^*\| \leq Cq^k$, $0 \leq q < 1$.

Доказательство. У функции f $\exists$ глобальный минимум x^* (лемма 12). При доказательстве теоремы 17 получили:

$$f(x^{k+1}) \leq f(x^k) - \alpha(1 - L\alpha/2)\|f'(x^k)\|^2$$

отсюда, учитывая неравенство

$$\|f'(x^k)\|^2 \geq 2l(f(x^k) - f(x^*)) \text{ (лемма 13)}$$

получим

$$f(x^{k+1}) \leq f(x^k) - l\alpha(2 - L\alpha)(f(x^k) - f(x^*))$$

или,

Градиентные методы

вычитая $f(x^*)$ из обеих частей, имеем:

$$f(x^{k+1}) - f(x^*) \le (1 - l\alpha(2 - L\alpha))\times$$
$$\times(f(x^k) - f(x^*)) \qquad (24)$$

Положим $q_1 = 1 - l\alpha(2 - L\alpha)$. Из (24) $\Longrightarrow$

$$f(x^{k+1}) - f(x^*) \le q_1^{k+1}(f(x^0) - f(x^*)) \qquad (25)$$

Функция $f \ne const$ ((от противного) Пусть $f = const$.

Тогда $f'(x) = 0$ и из неравенства $f(x+y) \geq f(x) + \langle f'(x), y\rangle + l\|y\|^2/2$ имеем $l\|y\|^2/2 \leq 0 \Rightarrow$ противоречие с условием $l > 0$)

$\Longrightarrow$ найдётся начальная точка x^0 такая, что

$$f(x^0) > f(x^*).$$

Тогда при $k = 0$ имеем (неравенство (24))

$$0 \leq f(x^1) - f(x^*) \leq q_1(f(x^0) - f(x^*)).$$

Следовательно $q_1 \geq 0$. Так как $q_1 < 1$, то из (25) следует, что

$$f(x^k) \to f(x^*).$$

Подставим в неравенство

$$f(x+y) \geq f(x) + \langle f'(x), y \rangle + l\|y\|^2/2$$

вместо y вектор $(x^k - x^*)$ и x^* вместо x.

Градиентные методы

Учитывая, что x^* – глобальный минимум, и, следовательно, $f'(x^*) = 0,$ получим

$$(f(x^k) - f(x^*)) \geq l\|x^k - x^*\|^2/2.$$

Из этого неравенства и (25) имеем:

$$\|x^k - x^*\|^2 \leq 2q_1^k\,(f(x^0) - f(x^*))/l.$$

Перепишем это неравенство в следующем виде

$\|x^k - x^*\| \leq Cq^k,$ где

$$C = \sqrt{2(f(x^0) - f(x^*))/l}, q = \sqrt{q_1}.$$

Градиентные методы

Таким образом метод имеет линейную оценку скорости сходимости. Из этого неравенства также получим, что $\boldsymbol{x}^{\boldsymbol{k}} \longrightarrow \boldsymbol{x}^*$ при $\boldsymbol{k} \longrightarrow \infty$. $\blacksquare$

Метод Такахаши = (дуализация/градиентный метод)

Рассмотрим задачу P

$$\max_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x})$$

$$\boldsymbol{Ax} = \boldsymbol{b}, \boldsymbol{Bx} = \boldsymbol{d}.$$

$\boldsymbol{f}$ – вогнутая, дифференцируемая функция. Предполагается, что задача P без ограничений $\boldsymbol{Bx} = \boldsymbol{d}$ легко разрешима.

Рассмотрим задачу D двойственную к задаче P

$$\min_{\boldsymbol{\lambda}} \boldsymbol{h}(\boldsymbol{\lambda}), \text{ где}$$

Метод Такахаши $=$ (дуализация/градиентный метод)

$$\boldsymbol{h(\lambda) = \max_x \{f(x) + (Bx - d, \lambda)\}}$$
$$\boldsymbol{Ax = b.}$$

1. Выбрать $\boldsymbol{\lambda^0}$.

2. Найти $\boldsymbol{h(\lambda^0)}$. Пусть $\boldsymbol{x^0}$ – оптимальное решение, на котором достигается $\boldsymbol{h(\lambda^0)}$. Если $\boldsymbol{Bx^0 = d}$, то $\boldsymbol{x^0}$ – оптимальное решение задачи $\mathbf{P}$. И алгоритм заканчивает свою работу. Иначе переходим на шаг 3.

3. Пусть $\boldsymbol{\lambda = \lambda^0 - \alpha(Bx^0 - d)}$. На шаг 2 с $\boldsymbol{\lambda}$ вместо $\boldsymbol{\lambda^0}$.