

Методы возможных направлений в задачах условной оптимизации

Персиянов Дмитрий, ФИВТ, 397 группа

Май, 2015

Содержание

1	Введение	1
2	Общий подход	2
2.1	Вспомогательная задача и поиск возможного направления убывания $f(x)$	2
2.2	Выбор шага d_k в переходе к новому приближению	5
2.3	Ухудшение сходимости при приближении к оптимуму	5

1 Введение

Мы рассматриваем задачу минимизации гладкой функции $f(x)$ на заданном множестве $X \subseteq \mathbb{R}^n$. *Методы возможных направлений* – это класс итерационных методов решения такой задачи, где на каждом шаге переход к новому приближению x_k осуществляется с помощью выбора так называемого возможного направления убывания функции.

Определение 1.1. Пусть $X \subseteq \mathbb{R}^n$, $x \in X$. Вектор $e \in \mathbb{R}^n$, $e \neq 0$ называется *возможным направлением* в точке $x \in X$, если $\exists t_0 > 0$ т.ч. $\forall 0 < t \leq t_0 \rightarrow x + te \in X$.

Определение 1.2. Вектор $e \neq 0$ называется *возможным направлением убывания* функции $f(x)$ в точке x на множестве X , если e – возможное направление в точке x и $\exists 0 < \beta \leq t_0$ т.ч. $\forall 0 < \alpha < \beta \rightarrow f(x + \alpha e) < f(x)$.

Итак, идея методов проста. На каждом шаге находим возможное направление убывания e функции $f(x)$, и ищем новое приближение по формуле $x_k = x_{k-1} + \alpha_k e$, где α_k – некоторый положительный шаг.

Примеры.

1. Градиентный спуск

Пусть $X = \mathbb{R}^n$. Т.к. функция $f(x)$ непрерывно дифференцируема на X , то в любой точке $x \in X$ существует градиент $\text{grad } f(x)$ и

$$f(x + h) - f(x) = \langle \text{grad } f(x), h \rangle + o(h) \quad (1)$$

По неравенству Коши-Буняковского

$$-|\text{grad } f(x)| \cdot |h| \leq \langle \text{grad } f(x), h \rangle \leq |\text{grad } f(x)| \cdot |h|$$

Левое равенство достигается при $h = -\text{grad } f(x)$, а так как при достаточно малых $|h|$ значение $f(x + h) - f(x)$ определяется членом $\langle \text{grad } f(x), h \rangle$, то можно взять $-\text{grad } f(x)$ в качестве возможного направления убывания функции $f(x)$ в точке x . Итоговое выражение для перехода:

$$x_k = x_{k-1} - \alpha_k \cdot \text{grad } f(x) \quad (2)$$

Стоит сказать, что градиентный спуск при приближении x_k к минимуму x_* , начинает плохо работать. Это связано с тем, что в окрестности точки минимума x_* градиент $\text{grad } f(x_k)$ мало отличим от нуля, а значит квадратичная часть приращения начинает иметь большее влияние (см. выражение (1)) и метод становится чувствительным к погрешностям вычислений.

Отдельно стоит говорить о выборе длины шага α_k . Мы не будем на этом останавливаться, сказав, что в большинстве практических задач его берут фиксированным для всех итераций (подробнее см. [1, р. 276]).

2. Метод проекции градиента

Теперь не обязательно $X = \mathbb{R}^n$. Действовать, как при градиентном спуске, теперь нельзя, потому что при каком-то k x_k может не принадлежать X . Новое выражение для x_k будет следующим:

$$x_k = \text{proj}_X (x_{k-1} + \alpha_k \cdot \text{grad } f(x)), \quad (3)$$

где $\text{proj}_X(y) = \inf_{x \in X} |y - x|$, $y \in \mathbb{R}^n$ – проекция точки y на множество X . Если X – выпукло и замкнуто, то проекция определена однозначно. В этом методе нужно уметь вычислять проекцию на множество X , что вообще говоря, не всегда просто сделать. Но для некоторых структур X формулу для проекции можно выписать явно. Подробнее см. [1, р. 293].

Для эффективности метода необходимо уметь быстро вычислять возможное направление убывания и переходить от старого приближения к новому. Далее мы перейдём к общей теории и рассмотрим методы выбора возможных направлений, которые просто и удобно можно реализовать на компьютере.

2 Общий подход

Рассмотрим задачу

$$f(x) \rightarrow \inf, \quad x \in X = \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = \overline{1, m}\}, \quad (1)$$

где функции $f(x), g_i(x), i = \overline{1, m}$ определены на всём \mathbb{R}^n и непрерывно дифференцируемы на X .

Пусть $x_0 \in X$ – некоторое начальное приближение. Мы хотим научиться делать переход от старого приближения к новому. Для этого мы сейчас рассмотрим вспомогательную задачу, являющуюся задачей линейного программирования. Решение этой задачи будет давать нам возможное направление убывания функции.

2.1 Вспомогательная задача и поиск возможного направления убывания $f(x)$

Итак, пусть дано k -е приближение x_k . Введём так называемое множество индексов *активных ограничений*:

$$I_k = \{i : 1 \leq i \leq m, g_i(x_k) = 0\}$$

Вспомогательная задача формулируется следующим образом:

$$\begin{aligned} \sigma \rightarrow \inf, \quad z = (e, \sigma) = (e^1, \dots, e^n, \sigma) \in W_k, \\ W_k = \{(e, \sigma) \in \mathbb{R}^{n+1} : \langle \text{grad } f(x_k), e \rangle \leq \sigma, \\ \langle \text{grad } g_i(x_k), e \rangle \leq \sigma, i \in I_k; |e^j| \leq 1, j = \overline{1, n}\} \end{aligned} \quad (2)$$

Из определения W_k видно, что это задача линейного программирования. Поймём, что она имеет решение. Видно, что $z = (0, 0) \in W_k \Rightarrow \inf_{W_k} \sigma = \sigma_k \leq 0$. Далее, множество W_k – полиэдр, ибо оно замкнуто, ограничено и задается линейными ограничениями. Значит, по т. Вейерштрасса, задача (2) имеет хотя бы одно решение. Искать его мы можем известными методами, например симплекс-методом.

Будем считать, что мы решили задачу (2) и нашли $(e_k, \sigma_k) \in W_k$ т.ч. $\sigma_k = \inf_{W_k} \sigma$. Возможны два случая: $\sigma_k < 0$ и $\sigma_k = 0$. Сейчас мы поймём, что второй случай означает, что x_k – решение задачи (2).

Итак, пусть $\sigma_k = 0$. Нам понадобятся следующие обозначения: $f_* = \inf_X f$, $X_* = \{x \in X : f(x) = f_*\}$. Также переформулируем задачу (2) для точки $x_* \in X_*$:

$$\begin{aligned} \sigma \rightarrow \inf, \quad z = (e, \sigma) = (e^1, \dots, e^n, \sigma) \in W_*, \\ W_* = \{(e, \sigma) \in \mathbb{R}^{n+1} : \langle \text{grad } f(x_*), e \rangle \leq \sigma, \\ \langle \text{grad } g_i(x_*), e \rangle \leq \sigma, i \in I_*; |e^j| \leq 1, j = \overline{1, n}\}, \end{aligned} \quad (3)$$

где $I_* = \{1 \leq i \leq m, g_i(x_*) = 0\}$. Для краткости далее будем вместо $\text{grad } f(x)$ писать $f'(x)$, имея ввиду n -мерный вектор частных производных.

Теорема 2.1. Пусть функции $f(x), g_1(x), \dots, g_m(x) \in C^1(X)$, а задача (1) имеет решение, т.е. $f_* > -\infty$, $X_* \neq \emptyset$. Тогда для любой точки $x_* \in X_*$ задача (3) необходимо имеет решение (e_*, σ_*) с $\sigma_* = \min_{W_*} \sigma = 0$.

Для доказательства этой теоремы сформулируем теорему из теории метода множителей Лагранжа, которую доказывать не будем.

Теорема 2.2 (Каруша-Джона). Пусть функции $f(x), g_1(x), \dots, g_m(x) \in C^1(X)$, а задача (1) имеет решение, т.е. $f_* > -\infty$, $X_* \neq \emptyset$. Тогда для любой точки $x_* \in X_*$ существуют множители Лагранжа $\lambda_0^*, \dots, \lambda_m^*$:

$$\forall i \lambda_i^* \geq 0, \exists i : \lambda_i^* \neq 0, \forall i \lambda_i^* g_i(x_*) = 0, \mathcal{L}'_x(x_*, \lambda_0^*, \dots, \lambda_m^*) = 0$$

Доказательство теоремы 2.1. По теореме 2.2 находим множители Лагранжа неотрицательные и не все равные нулю, такие, что

$$\lambda_0^* f'(x_*) + \sum_{i=1}^m \lambda_i^* g'_i(x_*) = 0, \quad \lambda_i^* g_i(x_*) = 0, \quad i = \overline{1, m} \quad (4)$$

Заметим, что первое равенство можно переписать в виде

$$\lambda_0^* f'(x_*) + \sum_{i \in I_*} \lambda_i^* g'_i(x_*) = 0 \quad (5)$$

Далее, возьмём произвольную точку $(e, \sigma) \in W_*$. Тогда $\langle f'(x_*), e \rangle \leq \sigma$, $\langle g'_i(x_*), e \rangle \leq \sigma, i \in I_*$. Домножим неравенство с $f'(x_*)$ на λ_0^* , а неравенства с $g_i(x_*)$ на λ_i^* соответственно, а затем сложим. Получим $\langle \lambda_0^* f'(x_*) + \sum_{i \in I_*} \lambda_i^* g'_i(x_*) \rangle = 0 \leq \sigma(\lambda_0^* + \dots + \lambda_m^*)$. Следовательно, $\sigma \geq 0$. Но ранее мы показали, что $\sigma_* \leq 0$. Значит $\sigma_* = 0$. \square

Оказывается, что если наша задача выпукла, то x_k является её решением. На этот счёт справедлива следующая теорема.

Теорема 2.3. *В условиях теоремы 2.1 пусть дополнительно выполнены следующие условия:*

1. $f(x), g_i(x)$ выпуклы на \mathbb{R}^n .
2. Выполнено условие Слейтера, т.е. существует точка $x' \in X$ т.ч.
 $g_i(x') < 0 \quad \forall i = 1, \dots, m$.

Тогда любая точка $x_* \in X_*$, для которой задача (3) находит $\sigma_* = 0$, является решением задачи (1).

Таким образом, если задача выпукла, и решая вспомогательную задачу на k -ой итерации мы получили $\sigma_k = 0$, то это значит, что мы нашли решение x_k задачи (1).

Если задача невыпукла, то условие $\sigma_* = 0$ не является достаточным для оптимальности x_* .

Пример 2.1. $f(x, y) = x + \cos y$, $(x, y) \in X = \{(x, y) \in \mathbb{R}^2 : g(x, y) = -x \leq 0\}$. Рассмотрим $(x_*, y_*) = (0, 0)$. Очевидно, она не является точкой минимума на множестве X , ибо $f(0, \pi) = -1 < 0$. Однако, $f'(0, 0) = (1, 0)$, $g'(0, 0) = (-1, 0)$. Тогда $W_* = \{(e^1, e^2, \sigma) : e^1 \leq \sigma, -e^1 \leq \sigma, |e^1| \leq 1, |e^2| \leq 1\}$. Тогда $|e^1| \leq \sigma \quad \forall (e^1, e^2, \sigma) \in W_* \Rightarrow \sigma_* = \inf_{W_*} \sigma = 0$.

Теперь рассмотрим случай, когда $\sigma_k < 0$. Покажем, что тогда e_k является возможным направлением убывания функции $f(x)$. Имеем

$$\langle f'(x_k), e_k \rangle \leq \sigma_k < 0, \quad \langle g'_i(x_k), e_k \rangle \leq \sigma_k < 0, \quad i \in I_k.$$

Значит, во-первых, $e_k \neq 0$, а во-вторых, для любого $i \in I_k \rightarrow g_i(x_k) = 0$ и

$$\begin{aligned} g_i(x_k + \alpha e_k) &= g_i(x_k + \alpha e_k) - g_i(x_k) = \langle g_i(x_k), e_k \rangle \alpha + o(\alpha) \leq \\ &\leq \alpha \sigma_k + o(\alpha) = \alpha(\sigma_k + o(1)) < 0 \text{ при } 0 < \alpha < \alpha_0. \end{aligned} \quad (6)$$

Если $i \notin I_k$, то $g_i(x_k) < 0$. Из непрерывности $g_i(x)$ следует, что это $g_i(x_k + \alpha e_k) < 0$ при $0 < \alpha < \alpha_i$. Значит,

$$g_i(x_k + \alpha e_k) < 0, \quad i = 1, \dots, m, \quad 0 < \alpha < \alpha_0 \quad (7)$$

Получили, что e_k является возможным направлением множества X . Аналогично рассуждениям выше можем получить (при необходимости уменьшив α_0):

$$f(x_k + \alpha e_k) - f(x_k) < 0, \quad 0 < \alpha < \alpha_0 \quad (8)$$

Таким образом, e_k является возможным направлением убывания функции $f(x)$ в точке x_k на множестве X . Теперь мы можем перейти к новому приближению $x_{k+1} = x_k + d_k e_k$, $0 < d_k < \beta_k = \sup\{t : x_k + t e_k \in X\}$.

Наконец мы можем понять смысл вспомогательной задачи. Минимизируя σ , мы хотим приблизить e_k как можно ближе к антиградиенту (мы знаем, что $\langle f'(x_k), h \rangle \geq -|f'(x_k)|^2$).

Далее, поговорим о выборе шага d_k .

2.2 Выбор шага d_k в переходе к новому приближению

Мы остановились на том, что нашли e_k – возможное направление убывания функции $f(x)$ в точке x_k на множестве X . Переход к новому приближению осуществляется по следующей формуле:

$$x_{k+1} = x_k + d_k e_k, \quad 0 < d_k < \beta_k = \sup\{t : x_k + te_k \in X\} \quad (9)$$

Перечислим способы выбора d_k .

1. Если градиент $f'(x)$ удовлетворяет условию Липшица на множестве X , т.е. $|f'(x) - f'(y)| \leq L|x - y|$, $x, y \in X$ для некоторой константы $L \geq 0$ и константа нам известна, то шагом можно взять

$$d_k = \min(\beta_k, |\sigma_k|L^{-1}) \quad (10)$$

2. Можно выбирать шаг, удовлетворяющий следующему условию:

$$f(x_k) - f(x_k + d_k e_k) \geq \varepsilon d_k |\sigma_k|, \quad 0 < d_k \leq \beta_k, \quad 0 < \varepsilon < \frac{1}{2} \quad (11)$$

Сначала положим $d_k = \beta_k$ и при необходимости будем её уменьшать.

3. Если сложно вычислить β_k , то можно принять шаг равным какой-то константе $d_k = \alpha$, проверить убывание $f(x)$ и принадлежность $x_k + d_k e_k \in X$, при необходимости снова уменьшая величину шага.

2.3 Ухудшение сходимости при приближении к оптимуму

К сожалению, вышеописанный метод преследует та же проблема, что и градиентный спуск. Когда в решении вспомогательной задачи (2) значение $\sigma_k < 0$ близко к нулю, направление e_k , хоть и является теоретически возможным направлением убывания функции в точке x_k , на практике обладает свойством $\langle g'_i(x_k), e_k \rangle \approx \sigma_k \approx 0$ при некотором $i \in I_k$, либо $\langle f'(x_k), e_k \rangle \approx \sigma_k \approx 0$. В первом случае мы получим маленькое значение для β_k ($\beta_k = \sup\{t : x_k + te_k \in X\}$), а следовательно и для длины шага тоже (см. раздел 2.2 о выборе шага), а во втором случае мы получим, что целевая функция $f(x)$ убывает крайне медленно вдоль направления e_k .

Для избежания таких явлений, мы поступим следующим образом.

Пусть $x_0 \in X, \varepsilon_0 > 0$ – некоторое начальное приближение. Допустим, что k -е приближение $(x_k, \varepsilon_k), x_k \in X, \varepsilon_k > 0$ известно при каком-то k . Определим множество

$$I_k = \{i : 1 \leq i \leq m, -\varepsilon_k \leq g_i(x_k) \leq 0\} \quad (12)$$

и решим нашу вспомогательную задачу (2) для такого I_k . Её решение (e_k, σ_k) .

1. $\sigma_k \leq -\varepsilon_k$. Тогда считаем, что e_k обладает выраженным свойством возможного направления убывания $f(x)$ в x_k и делаем переход

$$x_{k+1} = x_k + d_k e_k, \quad 0 < d_k \leq \beta_k, \varepsilon_{k+1} = \varepsilon_k, \quad (13)$$

где β_k и шаг d_k определяются так же, как и ранее.

2. $-\varepsilon_k < \sigma_k \leq 0$. Тогда делаем такой переход

$$x_{k+1} = x_k, \quad \varepsilon_{k+1} = \frac{\varepsilon_k}{2}, \quad (14)$$

и снова решаем вспомогательную задачу для соответствующего множества I_{k+1} . Смысл этого перехода в том, что мы сузили множество I_k , а значит, множество W_k во вспомогательной задаче расширилось, что могло дать нам новых кандидатов на возможное направление убывания.

Список литературы

- [1] Васильев Ф.П.: *Методы оптимизации*, ч. 1, МЦНМО, 2011
- [2] Сухарев А.Г., Тимохов А.В., Фёдоров В.В.: *Курс методов оптимизации*, Физматлит, 2005
- [3] Dianne P. O'Leary, *Feasible direction methods*: <https://www.cs.umd.edu/users/oleary/a607/607constrfdhand.pdf>