Multiple-Gradient Descent Algorithm

Лемма 1. *Пусть x^0 — Парето-оптимальная точка для набора непрерывно-дифференцируемых функций $f_i(x) \in C^1(\Omega)$ $(1 \leqslant i \leqslant n)$, и пусть $u_i^0 = \nabla f_i(x^0)$ - вектора градиентов функций в точке x^0. Тогда существует выпуклая комбинация этих векторов равная нулю:*

$$\sum_{i=1}^{n} \alpha_i u_i^0 = 0, \alpha_i \geq 0 \ (\forall i = 1 \dots n), \sum_{i=1}^{n} \alpha_i = 1$$

Доказательство. Пусть r ранг набора векторов $\{u_i^0\}_{i=1}^n$

$$r = rank(\{u_i^0\}_{i=1}^n) = dimSp(\{u_i^0\}_{i=1}^n)$$

Рассмотрим различные допустимые значени ранга r

Если $r = 0$, тогда все вектора равны нулю и результат тривиален.

Если $r = 1$, тогда вектора коллинеарны:

$$u_i = \beta_i u \ (\forall i = 1 \dots n)$$

Тогда рассмотрим приращение аргумента $\delta x^0 = -\varepsilon u$. Оно вызовет приращение функции $\delta f_i = -\varepsilon \beta_i + O(\varepsilon^2)$. Если все коэффициенты одного знака, тогда новая точка $x^0 + \delta x^0$ будет парето-доминировать точку x^0, что противоречит парето оптимальности точки x^0. Следовательно среди коэффициентов можно выбрать два коэффициента, имеющие различные знаки. Рассмотрим β_1, β_2: $\beta_1 \beta_2 < 0$. Взяв, $\alpha_1 = \frac{-\beta_2}{\beta_1 - \beta_2}, \alpha_2 = \frac{\beta_1}{\beta_1 - \beta_2}$ мы получим $\alpha_1 u_1 + \alpha_2 u_2 = 0$.

Рассмотрим случай, когда $2 \leq r \leq n - 1$. Тогда:

$$u_1^0 + \sum_{k=2}^{r+1} \mu_k u_k^0 = 0$$

Покажем, что $\mu_k \geq 0$ $(k \geq 2)$. Предположим, что верно обратное высказывание и существует $\mu_j < 0$. Для определенности рассмотрим $j = 2$. Рассмотрим линейную оболочку $V = Sp(\{u_i\}_{i=3}^n)$. Тогда размерность $dimV \leq r - 1 \leq n - 2 \leq N - 2$ и, следовательно $dimV^\perp \geq 2$. Рассмотрим произвольный $\omega \in V^*$ и производные функций $f'_{i,\omega}(x)$ по направлению ω. Тогда с учетом равенства выпуклой комбинации нулю получим:

$$\forall \omega \in V^\perp, f'_{2,w} = \gamma f'_{1,w} (\gamma = \frac{-1}{\mu_2} > 0)$$

Если выполняется равенство $0 = \gamma \cdot 0 \ \forall w \in V^{\perp}$, то вектора u_1 и u_2 принадлежат V. Отсюда мы получаем противоречие, так как ранг $Sp\{u_i\}$ равен r, а размерность $dimV \leq r - 1 < r$. Следовательно, для некоторого $\omega \in V^{\perp}$ $f'_{1,w} \neq 0$. Тогда вдоль направления $-\omega$, существует точка $\tilde{x}$, в которой $f_1(\tilde{x}) < f_1(x)$, $f_2(\tilde{x}) < f_2(x)$, а остальные функции не меняются. Это противоречит тому, что x — парето оптимальная точка. По аналогии, можно показать что все $\mu_k \geq 0 (\forall k \geq 2)$. Принимая $\mu_1 = 1$ и, рассматривая $\alpha_i = \frac{\mu_i}{\sum_{k=1}^{n} \mu_k}$ получаем искомую выпуклую комбинацию:

$$\sum_{i=1}^{n} \alpha_i u_i = 0$$

Рассмотрим случай $r = n$. Определим $C_k = f_k(x^0)$. Тогда для некоторого индекса i, x^0 решение следующей задачи условной оптимизации:

$$\begin{aligned} \min_{x} \quad & f_i(x) \\ \text{s.t.} \quad & g_k(x) := f_k(x) - C_k \leq 0 \ (\forall k \neq i) \end{aligned}$$

Запишем условие ККТ для этой задачи

$$u_1 + \sum_{k=2}^{n} \lambda_k u_k = 0$$

Таким образом, мы получаем противоречие с тем, что ранг $\{u_i\}_{i=1}^{n}$ равен r. Следовательно, $r \leq n - 1$ и мы рассмотрели всевозможные случаи. □

Лемма 2. *Пусть H - гильбертово пространство конечной или бесконечной размерности N, и $\{u_i\}_{i=1} (1 \leq i \leq n \leq N)$ семейство из n векторов в H. Пусть U выпуклая оболочка этих векторов*

$$U = \{w \in H | w = \sum_{i=1}^{n} \alpha_i u_i ; \alpha_i > 0 (\forall i); \sum_{i=1}^{n} \alpha_1 = 1\}$$

Тогда существует уникальный элемент $\omega \in U$, имеющий минимальную евклидову норму, и:

$$\forall u \in U : (u, \omega) \geq (\omega, \omega) = \|w\|^2 = C_{\omega}$$

Доказательство. Предположим, что существует два элемента с минимальной нормой $\|\omega_1\| = \|\omega_2\|$. Так как U — выпуклое множество, то $\forall \varepsilon \in [0, 1]$, $u = (1 - \varepsilon)\omega_1 + \varepsilon\omega_2 \in U$. Следовательно, $\|u\| \geqslant \|\omega_1\|$ и $(\omega_1 + \varepsilon\omega_{12}, \omega_1 + \varepsilon\omega_{12}) \geqslant (\omega_1, \omega_1)$, где $\omega_{12} = \omega_1 - \omega_2$. Вычитая из левой части, правую получим, что $2\varepsilon(\omega_1, \omega_{12}) + \varepsilon^2(\omega_{12}, \omega_{12}) \geqslant 0$. Рассматривая достаточно малый ε, получаем, что $(\omega_1, \omega_{12}) \geqslant 0$. Беря $\varepsilon = 1$, получаем строгое неравенство, если $\omega_{12} \neq = 0$, но в этом случае $u = \omega_2$ и равенство должно выполняться. Поэтому существует единственный элемент с минимальной нормой

Пусть $\bar{u}$ произвольный элемент в U; $r = u - \omega$. Так как U выпукла, то:

$$\forall \varepsilon \in [0,1], \omega + \varepsilon r \in U$$

Так как ω элемент с минимальной нормой, то $\|\omega + \varepsilon r\| \geq \|\omega\|$

$$\|\omega + \varepsilon r\|^2 - \|\omega\|^2 = (w + \varepsilon r, \omega + \varepsilon r) - (\omega, \omega) = 2\varepsilon(r, \omega) + \varepsilon^2(r, r) \geq 0.$$

Так как ε может быть произвольно маленьким, то

$$(r, \omega) = (\bar{u} - \omega, \omega) \geq 0$$

□

Теорема 3. *Пусть ω элемент выпуклой оболочки U с минимальной нормой. Тогда:*

1. *Либо $\omega = 0$, и точка $x = x^0$ — парето оптимальная.*
2. *Либо $\omega \neq 0$ and $-\omega$ и направление убывания для всех функиций и скалярное произведение $(u, \omega) = \|\omega\|^2$ $(\forall u \in U)$.*

Доказательство. Так как ω элемент U с минимальной нормой, то

$$w = u = \sum_{i=1}^{n} a_i u_i^0, \alpha = \arg\min j(u), j(u) = (u, u), \sum_{i=1}^{n} \alpha_i = 1$$

Выпишем функцию Лагранжа для задачи оптимизации u.

$$L(\alpha, \lambda) = j + \lambda(\sum_{i=1}^{n} \alpha_i - 1)$$

Запишем условие ККТ

$$\frac{\partial j}{\partial \alpha_i} + \lambda = 0$$

$$\frac{\partial j}{\partial \alpha_i} = 2(\frac{\partial u}{\partial \alpha_i}, u) = 2(u_i^0, w) = -\lambda.$$

Следовательно, производные по направлению ω для всех функций равны $-\lambda/2$. Наконец, для любого $u \in U, u = \sum_{i=1}^{n} \mu_i u_i$, где $\mu_i \geq 0$ $(\forall i = 1 \ldots n), \sum_{i=1}^{n} \mu_i = 1$ выполнено

$$(u, \omega) = \sum_{i=1}^{n} \mu_i (u_i^0, \omega) = -\lambda/2 = \|\omega\|^2.$$

□