Задача с ограничениями

$$\min_{x \in \mathbb{R}^d} f(x)$$
$$\text{s.t. } h_i(x) = 0 \qquad i = 1 \dots m$$

Перепишем

$$\min_{x \in \mathbb{R}^d} f(x) + \rho \cdot \frac{1}{2} \sum_{i=1}^{m} h_i^2(x) \qquad \rho > 0$$
$$\text{s.t. } h_i(x) = 0$$

решение не изменилось т.к. $h_i(x) = 0$

Упростим

$$\min_{x \in \mathbb{R}^d} \underbrace{f(x) + \rho \cdot \frac{1}{2} \sum_{i=1}^{m} h_i^2(x)}_{f_\rho(x) \text{ — штрафная функция}}$$

нет ограничений

⊕ задача с ограничениями стала без ограничений

⊖ решения другие

⊖ можно выйти за пределы ограничений

Минимизируя f_ρ, получаем решение исходной задачи при $\rho \to \infty$

$$\lim_{\rho \to +\infty} f_\rho(x) = \begin{cases} f(x), & x \text{ удовл. огран.} \\ +\infty, & \text{иначе} \end{cases}$$

Метод штрафной функции

пока x^k не достаточно хорошо (с т.з. f, h_i)

$$\rho^k = \rho^{k-1} \cdot \alpha \qquad \alpha > 1$$

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^d} f_{\rho^k}(x)$$

⊖ ρ – параметр, который подбирается в цикле

⊖ большие ρ ухудшают св-ва задачи

Более общая задача

$$\min_{x \in \mathbb{R}^d} f(x)$$

$$\text{s.t.} \quad h_i(x) = 0 \qquad i = 1 \ldots m$$

$$g_j(x) \le 0 \qquad j = 1 \ldots n$$

$$f_\rho(x) = f(x) + \rho \cdot \frac{1}{2} \sum_{i=1}^{m} h_i(x) + \rho \cdot \frac{1}{2} \sum_{j=1}^{n} \left((g_j(x))^+ \right)^2$$

$$g^+ = \max\{g; 0\}$$

Свойства штрафной функции

1) x^* – решение исходной задачи, x^*_ρ – решение штрафной задачи, тогда

$$f(x^*) \ge f(x^*_\rho)$$

Док-во:

$$h_i(x^*) = 0$$

$$f(x^*) = f_\rho(x^*) \ge \min_{x \in \mathbb{R}^d} f_\rho(x) = f_\rho(x^*_\rho) \ge f(x^*_\rho) \quad \blacksquare$$

2) С увеличением ρ решение штрафной задачи не ухудшается с точки зрения нарушения ограничений

$$\forall\ \rho_1 > \rho_2 \hookrightarrow \sum_{i=1}^{m} h_i^2(x^*_{\rho_2}) \geq \sum_{i=1}^{m} h_i^2(x^*_{\rho_1})$$

Док-во:

$x^*_{\rho_1}$ – решение f_{ρ_1}

$$+\begin{cases} f(x^*_{\rho_1}) + \rho_1 \cdot \frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_1}) \leq f(x^*_{\rho_2}) + \rho_1 \cdot \frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_2}) \\ f(x^*_{\rho_2}) + \rho_2 \cdot \frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_2}) \leq f(x^*_{\rho_1}) + \rho_2 \cdot \frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_1}) \end{cases}$$

$$\rho_1 \cdot \frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_1}) + \rho_2 \cdot \frac{1}{2}\sum h_i^2(x^*_{\rho_2}) \leq \rho_1 \cdot \frac{1}{2}\sum_{i=1} h_i^2(x^*_{\rho_2}) + \rho_2 \cdot \frac{1}{2}\sum_{i=1} h_i^2(x^*_{\rho_1})$$

$$(\rho_1 - \rho_2)\,\frac{1}{2}\sum_{i=1} h_i^2(x^*_{\rho_1}) \leq (\rho_1 - \rho_2)\,\frac{1}{2}\sum h_i^2(x^*_{\rho_2})$$

$\rho_1 > \rho_2$

$$\sum_{1} h_i^2(x^*_{\rho_1}) \leq \sum h_i^2(x^*_{\rho_2}) \ \blacksquare$$

3) f, h_i – непрер.

X^* – мн-во решений исходной задачи

доп обозначение: для $x^* \in X^*$

$$U(x^*) = \{x \in \mathbb{R}^d \mid f(x) \leq f(x^*)\}$$

Если $U(x^*)$ ограничено для $\forall x^* \in X^*$, тогда

$\forall \varepsilon > 0 \ \exists \rho(\varepsilon) > 0 : \ \forall \rho \geq \rho(\varepsilon) \hookrightarrow$

X_ρ^* (мн-во решений штрафной задачи)

содержится в ε-окрестности X^*

$$X_\varepsilon^* = \{ x \in \mathbb{R}^d \mid \exists \, x^* \in X^* : \|x - x^*\|_2 \leq \varepsilon \}$$

$$\min_{x \in \mathbb{R}^d} f(x) \quad \text{s.t. } Ax = b$$

<u>Лагранжиан:</u>

$$L(x, \lambda) = f(x) + \lambda^T (Ax - b)$$

<u>Двойственная функция</u> $\quad g(\lambda) = \min\limits_{x \in \mathbb{R}^d} L(x, \lambda)$

<u>Двойственный подъем:</u>

$$\lambda^{k+1} = \lambda^k + \alpha \nabla g(\lambda^k)$$
$$= \lambda^k + \alpha \nabla \left(\min_{x \in \mathbb{R}} \left(f(x) + \lambda_k^T (Ax - b) \right) \right)$$

$g(\lambda)$ сложно найти аналитически

<u>Модификация метода:</u>

$$x^{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^d} \left[f(x) + \lambda_k^T (Ax - b) \right]$$

$$\lambda^{k+1} = \lambda^k + \alpha \nabla_{\lambda_k} \left(f(x^{k+1}) + \lambda_k^T (Ax^{k+1} - b) \right)$$

$$= \lambda^k + \alpha\,(Ax^{k+1} - b)$$

(не тот самый метод, но $\approx$)

$\ominus$ нестабилен на практике

<u>Аугментация</u> (эквив. задача)

$$\min_{x\in\mathbb{R}^d} f(x) + \underbrace{\frac{\rho}{2}\|Ax-b\|_2^2}_{\text{аугментация}}$$

$$\text{s.t. } Ax = b$$

$$L_\rho(x,\lambda) = f(x) + \lambda^T(Ax-b) + \frac{\rho}{2}\|Ax-b\|_2^2$$

<u>Двойственный подъем</u>

$$x^{k+1} = \operatorname*{argmin}_{x\in\mathbb{R}^d} L_\rho(x,\lambda^k)$$

$$\lambda^{k+1} = \lambda^k + \cancel{\alpha}\,\rho\,(Ax^{k+1} - b)$$

не α, избавились от гиперпар.

не нужно $\rho \to +\infty$

$\oplus$ более стабильный метод

Чуть более общая задача:

$$\min_{x\in\mathbb{R}^{d_x},\ y\in\mathbb{R}^{d_y}} f(x) + g(y)$$

$$\text{s.t. } Ax + By = C$$

Пример

$$\min_{x\in\mathbb{R}^d} L(Ax,b) + r(x)$$

- logreg
- квадр

регул.

$$\min_{x\in\mathbb{R}^d,\ y\in\mathbb{R}^s} L(y,b) + r(x)$$

$$\text{s.t.}\quad Ax = y$$

Аугментация

$$\min_{\substack{x\in\mathbb{R}^{d_x}\\ y\in\mathbb{R}^{d_y}}} f(x) + g(y) + \frac{\rho}{2}\|Ax+By-C\|_2^2$$

$$\text{s.t.}\quad Ax + By = C$$

Лагранжиан

$$L_\rho(x,y,\lambda) = f(x) + g(y) + \lambda^T(Ax+By-C) + \frac{\rho}{2}\|Ax+By-C\|_2^2$$

ADMM (Alternating Directions Method of Multipliers)

$$x^{k+1} = \arg\min_x L_\rho(x, y^k, \lambda^k) \qquad \text{AD}$$

$$y^{k+1} = \arg\min_y L_\rho(x^{k+1}, y, \lambda^k) \qquad \text{AD}$$

$$\lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - C) \qquad \text{M}$$

(обобщение двойного подсчёта)