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

эквивалентны

- Метод штрафной функции

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

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

ρ — штрафной параметр

⊕ задача без ограничений

⊖ $\rho \neq +\infty$ выходим из мн-ва $h_i(x)=0$

$$\lim_{\rho\to+\infty} f_\rho = \begin{cases} f(x) & x \text{ удовл. } h_i(x)=0 \\ +\infty & \text{иначе} \end{cases}$$

<u>Свойства</u> (решения штрафной задачи)

1) x^* — решение исходной, x^*_ρ — решение штрафной задачи

тогда

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

либо мы нарушаем огр
либо $f(x^*)=f(x^*_\rho)$

<u>Док-во:</u> $f(x^*) = f_\rho(x^*) \geq \min_{x\in\mathbb{R}^d} f_\rho(x) = f_\rho(x^*_\rho) \geq f(x^*_\rho)$ ■

2) С увеличением ρ решение штрафной задачи гарантированно не увеличивает степень нарушения ограничений

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

решения соотв. штрафных задач

Док-во:

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

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

$$(\rho_1 - \rho_2)\frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_1}) \leq (\rho_1 - \rho_2)\frac{1}{2}\sum_{i=1}^{m} h_i^2(x^*_{\rho_2}) \quad \Big| : (\rho_1 - \rho_2)\frac{1}{2}$$

$\rho_1 > \rho_2$

$$\sum_{i=1}^{m} h_i^2(x^*_{\rho_1}) \leq \sum_{i=1}^{m} h_i^2(x^*_{\rho_2}) \quad \blacksquare$$

3) Пусть f и h_i – непрерывные

X^* – мн-во решений исходной задачи

Тогда $\forall \varepsilon > 0 \ \exists \rho(\varepsilon) > 0$:

для $\forall \rho > \rho(\varepsilon) \hookrightarrow X^*_\rho$ – мн-во решений штрафной задачи f_ρ лежит в ε-окрестности X^*

⊕ $\rho \to +\infty$ решение штрафной приближается к исходной

⊖ Итеративный подбор ρ – долго

⊖ Увеличение ρ влечет увеличение сложности штр. задачи

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

- Лагранжиан:

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

Идея: двойственно: $g(\lambda) = \min_{x \in \mathbb{R}^d} L(x, \lambda)$

Градиентный спуск (подъем) (двойственный подъем)

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

Другой взгляд (не хочу искать $g(\lambda)$):

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

- Аугментация

$$\min_{x \in \mathbb{R}^d} f(x) + \underbrace{\frac{\rho}{2} \|Ax - b\|_2^2}_{\text{аугментация}} \quad \text{(задача не изменилась)}$$
$$\text{s.t. } Ax = b$$

⊕ практический трюк для стабилизации сходимости

Лагранжиан:

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

$$x^{k+1} = \arg\min L_\rho(x, \lambda^k)$$
$$\lambda^{k+1} = \lambda^k + \gamma (Ax^{k+1} - b)$$

$$\min_{x \in \mathbb{R}^{d_x},\, y \in \mathbb{R}^{d_y}} f(x) + g(y) \qquad A \in \mathbb{R}^{n \times d_x} \quad B \in \mathbb{R}^{n \times d_y}$$
$$\text{s.t. } Ax + By = c \qquad c \in \mathbb{R}^n$$

Аугментация

$$\min_{x \in \mathbb{R}^{d_x},\, y \in \mathbb{R}^{d_y}} f(x) + g(y) + \frac{\rho}{2} \|Ax + By - c\|_2^2$$
$$\text{s.t. } Ax + By = c$$

Лагранжиан

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

Двойственный подъем $\Rightarrow$ ADMM

($\underbrace{\text{Alternating Direction}}_{x,\,y\ \text{вместе}\ X}$ Method of $\underbrace{\text{Multipliers}}_{\lambda}$)

A D

$$\boxed{\begin{aligned} y^{k+1} &= \arg\min_{y \in \mathbb{R}^{d_y}} L_\rho(x^k, y, \lambda^k) \\ \lambda^{k+1} &= \lambda^k + \rho (Ax^k + By^k - c) \\ x^{k+1} &= \arg\min_{x \in \mathbb{R}^{d_x}} L_\rho(x, y^{k+1}, \lambda^{k+1}) \end{aligned}}$$

не ошибка

(оставляем только 1 параметр у метода)

Примеры из ML

• задача регрессии:

$$\min_{x\in\mathbb{R}^d} \ell(Ax;b) + \frac{\lambda}{2}\|x\|_2^2 \qquad A\in\mathbb{R}^{n\times d},\ \lambda>0$$

трюк с доп. переменной

$$\min_{x\in\mathbb{R}^d,\, z\in\mathbb{R}^n} \ell(z,b) + \frac{\lambda}{2}\|x\|_2^2 \quad \text{s.t. } z = Ax$$

⊕ разделить ℓ и A

$$\ell(Ax,b) = \tilde{\ell}(\tilde{A}x,b)$$

$$\tilde{\ell}(\cdot,b) = \ell\left(\frac{\cdot}{\beta};b\right)$$

$$\tilde{A} = \beta A$$

Док-во: Условие оптимальности для

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

$$\nabla_y L_\rho(x^k, y^{k+1}, \lambda^k) = 0:$$

1) $\nabla g(y^{k+1}) + B^T\lambda^k + \rho B^T(Ax^k + By^{k+1} - c) = 0$

Условие оптимальности для

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

$$\nabla_x L_\rho(x^{k+1}; y^{k+1}; \lambda^{k+1}) = 0$$

2) $\nabla f(x^{k+1}) + A^T\lambda^{k+1} + \rho A^T(Ax^{k+1} + By^{k+1} - c) = 0$

3) $\lambda^{k+1} - \lambda^k = \rho(Ax^k + By^k - c)$

Ремарка: $\langle \nabla_x f(x^k); x^k - x^* \rangle$

$$\Downarrow$$

$$\left\langle \begin{pmatrix} \nabla_x L_0 \\ \nabla_y L_0 \\ -\nabla_\lambda L_0 \end{pmatrix}; \ldots \right\rangle$$

Из условий оптимальности 1), 2), 3)

$$\begin{pmatrix} \nabla f(x^{k+1}) + A^T \lambda^{k+1} \\ \nabla g(y^{k+1}) + B^T \lambda^{k+1} \\ -(Ax^{k+1} + By^{k+1} - c) \end{pmatrix} = - \begin{pmatrix} \rho A^T(Ax^{k+1} + By^{k+1} - c) \\ \cancel{-B^T(\lambda^{k+1} - \lambda^k) + \rho B^T(Ax^k + By^{k+1} - c)} \\ \frac{1}{\rho}(\lambda^{k+1} - \lambda^k) + A(x^{k+1} - x^k) \end{pmatrix}$$

0 3)

$$\parallel$$

$$\begin{pmatrix} \nabla_x L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ \nabla_y L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ -\nabla_\lambda L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \end{pmatrix}$$

$$P = \begin{pmatrix} \rho A^T A & 0 & -A^T \\ 0 & 0 & 0 \\ -A & 0 & \frac{1}{\rho} I \end{pmatrix}$$

$$\begin{pmatrix} \nabla_x L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ \nabla_y L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ -\nabla_\lambda L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \end{pmatrix} = -P \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix}$$

$$\left\langle \begin{pmatrix} \nabla_x L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ \nabla_y L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ -\nabla_\lambda L_0(x^{k+1}, y^{k+1}, \lambda^{k+1}) \end{pmatrix}; \begin{pmatrix} x^{k+1} - x \\ y^{k+1} - y \\ \lambda^{k+1} - \lambda \end{pmatrix} \right\rangle$$

$$= -\left\langle P \begin{pmatrix} x^{k+1}-x^k \\ y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}; \begin{pmatrix} x^{k+1}-x \\ y^{k+1}-y \\ \lambda^{k+1}-\lambda \end{pmatrix} \right\rangle$$

$\|x\|_P =$

$= \langle x, Px \rangle$

$$= \left\| \begin{pmatrix} x^{k}-x \\ y^{k}-y \\ \lambda^{k}-\lambda \end{pmatrix} \right\|_P^2 - \left\| \begin{pmatrix} x^{k+1}-x \\ y^{k+1}-y \\ \lambda^{k+1}-\lambda \end{pmatrix} \right\|_P^2$$

$$\cancel{- \left\| \begin{pmatrix} x^{k+1}-x^k \\ y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\|_P^2}$$

$\frac{1}{K}\sum_{k=0}^{K-1}$

Сходимость ADMM:

$$L_0\left(\frac{1}{K}\sum x^k;\ \frac{1}{K}\sum y^k;\ \lambda\right) - L\left(x, y, \frac{1}{K}\sum \lambda_k\right) \le \frac{1}{2K} \left\| \begin{pmatrix} x_0-x \\ y_0-y \\ \lambda_0-\lambda \end{pmatrix} \right\|_P^2$$

$\forall\, x, y, \lambda$