30.10.17 МОМО IX

Прямо-двойственные методы оптимизации

С-ма нелинейных ур-й

$r(x) = 0, \quad x \in \mathbb{R}^n, \quad r: \mathbb{R}^n \to \mathbb{R}^n, \quad r \in C^1$

$r(x_k + d) = r(x_k) + J(x_k)d + \bar{o}(\|d\|), \quad J \in \mathbb{R}^{n \times n}, \quad J_{ij} = \frac{\partial r_i}{\partial x_j}$

Приближение ф-ии в точке линейной моделью

$r(x_k + d) \simeq m_k(d) = r(x_k) + J(x_k)d = 0, \quad d_k = -J^{-1}(x_k)\, r(x_k)$

Мера прогресса: $f(x_k) = \frac{1}{2}\|r_k\|^2$

Критерий останова: $f(x_k) \le \varepsilon$

$\nabla f(x_k) = \nabla\left(\frac{1}{2}\sum_{i=1}^{n} r_i^2(x_k)\right) = \frac{1}{2}\sum_{i=1}^{n} 2\, r_i(x_k)\nabla r_i(x_k) = J^T(x_k)\, r(x_k)$

$\nabla f^T(x_k) d_k = -r_k^T(x_k)\, J(x_k)\, J^{-1}(x_k)\, r(x_k) = -\|r(x_k)\|^2 = -2f(x_k) < 0$

$x_{k+1} = x_k + \alpha_k d_k$

$\alpha_k: \frac{1}{2}\|r(x_k + \alpha_k d_k)\|^2 \le \frac{1}{2}\|r_k\|^2 + c_1\alpha_k(-\|r(x_k)\|^2) =$

$= \frac{1}{2}\|r_k\|^2(1 - 2c_1\alpha_k)$

Схема метода Ньютона для $r(x) = 0$

x_0, ε

Вычислить r_0, J_0

для $k = 0, 1, 2, \dots$

- Найти d_k: $J_k d_k = -r_k$
- Найти α_k: $f(x_k + \alpha_k d_k) \le f(x_k)(1 - 2c_1\alpha_k)$
- $x_{k+1} = x_k + \alpha_k d_k$
- Вычислить r_{k+1}, J_{k+1}
- если $\frac{1}{2}\|r_{k+1}\|^2 \le \varepsilon$, то стоп

$f(x) = \frac{1}{2}\|r(x)\|^2 \to \min_x$

$\nabla f(x) = J^T(x)\, \cancel{J(x)}\, r(x), \quad \nabla^2_{GN} f(x) = J(x)^T J(x)$

$J^T(x) J(x) d = -J^T(x)\, r(x)$

Неточный метод Ньютона

$res_k = J_k d_k + r_k$, критерий останова: $\|res_k\| \le \eta_k \|r_k\|$

1) $\eta_k \le \eta < 1 \to$ линейная

2) $\eta_k = \min\left(\frac{1}{2}, \sqrt{\|r_k\|}\right) \to$ суперлинейная

3) $\eta_k = \min\left(\frac{1}{2}, \|r_k\|\right) \to$ квадратичная

$$\nabla f(x_k)^T d_k = r(x_k)^T \underbrace{J(x_k) d_k}_{res_k - r_k} = r_k^T res_k - \|r_k\|^2 \le \eta_k \|r_k\|^2 - \|r_k\|^2 =$$
$$= (\eta_k - 1)\|r_k\|^2 < 0$$

$$\begin{cases} f(x) \to \min\limits_x, & x \in \mathcal{D} \subseteq \mathbb{R}^n, \ f \text{ - вып.} \in C^2 \\ Ax = b, & A \in \mathbb{R}^{p\times n}, \ \operatorname{rank} A = p \end{cases}$$

допустимые точки: $x_0 \in F \to x_1 \in F \to \dots$

$$x_k \in F, \quad f(x_k + d) \simeq m_k(d) = \begin{cases} f_k + \nabla f_k^T d + \frac{1}{2} d^T \nabla f_k d \to \min\limits_d \\ A(x_k + d) = b \end{cases}$$

ККТ

$$\begin{cases} \nabla f_k + \nabla^2 f_k d + A^T \mu = 0 \\ Ad = 0 \end{cases}, \quad \mathcal{L}(x, \mu) = f(x) + \mu^T (Ax - b)$$

$$\nabla_x \mathcal{L}(x, \mu) = \begin{cases} \nabla f(x) + A^T \mu = 0 \\ Ax = b \\ x \in \mathcal{D} \end{cases}$$

ККТ

$$y = (x, \mu) \in \mathbb{R}^{n+p}$$

$$r(y) = \begin{bmatrix} r_1(y) \\ r_2(y) \end{bmatrix} = \begin{bmatrix} \nabla f(x) + A^T \mu \\ Ax - b \end{bmatrix}, \quad r(y) \in \mathbb{R}^{n+p}$$

$$J(y_k) d = \begin{bmatrix} \nabla^2 f(x_k), & A^T \\ A, & 0 \end{bmatrix} \begin{bmatrix} d^x \\ d^\mu \end{bmatrix} = -r(y_k)$$

α_k: 1) $\|r(y_k + \alpha_k d_k)\|^2 \le (1 - 2c_1\alpha_k)\|r(y_k)\|^2$

2) $x_k + \alpha_k d_k^x \in \mathcal{D}$

$$\begin{cases} f(x) \to \min\limits_{x, s}, & x \in \mathcal{D} \subseteq \mathbb{R}^n, \ f, g_i \text{ - вып.} \in C^2 \\ g_i(x) + s_i = 0, & i = \overline{1,m} \\ Ax = b, & A \in \mathbb{R}^{p\times n}, \ \operatorname{rank} A = p \\ s_i \geq 0, & i = \overline{1,m} \end{cases}$$

$$\mathcal{L}(x, s, \mu_1, \mu_2, \lambda) = f(x) + \mu_1^T(g(x) + s) +$$
$$+ \mu_2^T(Ax - b) - \lambda^T s$$

1) $\nabla_x \mathcal{L} = \nabla f(x) + \left(\frac{\partial g(x)}{\partial x}\right)^T \mu_1 + A^T \mu_2 = 0$

2) $\nabla_s \mathcal{L} = \mu_1 - \lambda$

3) $Ax = b, \quad g(x) + s = 0, \quad s \geq 0 \quad [g(x) \leq 0, \ g(x) \leq -s]$

4) $\lambda \geq 0$

5) $\lambda_i s_i = 0, \quad i = \overline{1, m}$

Возмущённая ККТ:
$$\begin{cases} \nabla f(x) + \left(\frac{\partial g(x)}{\partial x}\right)^T \lambda + A^T \mu = 0 \\ diag(\lambda) diag(s) \cdot 1 = t \cdot 1 \\ Ax = b \\ g(x) + s = 0 \\ \lambda, s \geq 0 \end{cases}$$

$y = (x, s, \mu, \lambda)$

$r(y) = [r_1(y), r_2(y), r_3(y), r_4(y)]^T$

$$r(y) = \begin{bmatrix} \nabla f(x) + \left(\frac{\partial g(x)}{\partial x}\right)^T \lambda + A^T \mu \\ \Lambda \cdot S \cdot 1 - t \cdot 1 \\ Ax - b \\ g(x) + s \end{bmatrix}$$

$$\underline{J(y) \cdot d = - r(y)}$$

$$\begin{array}{c} n \\ m \\ p \\ m \end{array} \begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x), & 0, & A^T, & \left(\frac{\partial g(x)}{\partial x}\right)^T \\ 0, & \Lambda, & 0, & S \\ A, & 0, & 0, & 0 \\ \frac{\partial g(x)}{\partial x}, & I, & 0, & 0 \end{bmatrix} \begin{bmatrix} d^x \\ d^s \\ d^\mu \\ d^\lambda \end{bmatrix}$$

$$= - r(y)$$

Шаг: α_k

1) $\| r(y_k + \alpha_k d_k) \|^2 \le \| r(y_k) \| (1 - 2c_1 \alpha_k)$

2) $x_k + \alpha_k d_k \in \mathcal{D} \to \alpha_{max}^1$

3) $\lambda_k + \alpha_k d_k^\lambda > 0 \to \alpha_{max}^2$ $\quad \alpha_{max}^2 = \min\limits_{i: d_{k,i}^\lambda} \left(-\frac{\lambda_{k,i}}{d_{k,i}^\lambda} \right)$

4) $s_k + \alpha_k d_k^s > 0 \to \alpha_{max}^3$

$$\alpha_{start} = \min(1, \alpha_{max}^1, 0.95\alpha_{max}^2, 0.95\alpha_{max}^3)$$

] $r^t(y) = 0$

$r_1^t(y) = 0 \iff \nabla f(x) + \left(\frac{\partial g(x)}{\partial x}\right)^T \lambda + A^T \mu = 0$

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

$\Rightarrow q(\mu, \lambda) = L(x_*, s_*, \mu, \lambda)$ ← звёздочка

$f(x) - f_{opt} \le f(x) - q(\lambda, \mu) = f(x) - L(x_*, s_*, \lambda, \mu) =$

$= \cancel{f(x)} - \cancel{f(x)} - \underbrace{\lambda^T g(x)}_{-s,\ \text{т.к. } r_4^t(y)=0} - \underbrace{\mu^T (Ax - b)}_{0,\ \text{т.к. } r_3^t(y)=0} = \lambda^T s \overset{r_2^t(y)=0}{=} m \cdot t$

средняя комплиментарность: $t = \frac{\lambda^T s}{m}$

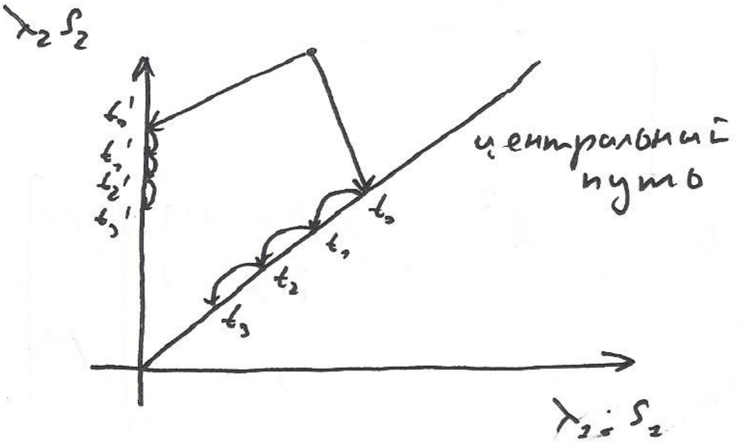


Схема PD-IPM

$x_0 \in \mathcal{D}, \lambda_0 > 0, s_0 > 0, \mu_0,$

$\varepsilon_{feas}, \varepsilon, \sigma \in (0, 1]$

для $k = 0, 1, 2, \ldots$

$t_k = \min \left(\frac{m}{\varepsilon}, \sigma \frac{\lambda_k^T s_k}{m} \right)$

d_k: $J^t(y_k) d_k = -r^t(y_k)$
неточное решение с-мы

α_k #: 1) $\| r(y_k + \alpha_k d_k) \|^2 \le \ldots$
2) $x_k + \alpha_k d_k^x \in \mathcal{D}$, 3) $\lambda_k + \alpha_k d_k^\lambda > 0$ 4) $s_k + \alpha_k d_k^s > 0$

если $\max(\| r_1^t(y_{k+1}) \|, \| r_3^t(y_{k+1}) \|, \| r_4^t(y_{k+1}) \|) \le \varepsilon_{feas}$, $\lambda_{k+1}^T s_{k+1} \le \varepsilon$
то стоп

Из большой системы:

$$\Lambda d^s + S d^\lambda = -z_2(y), \quad d^s = -\Lambda^{-1} S d^\lambda - \Lambda^{-1} z_2(y)$$

$$\begin{bmatrix} \nabla^2 f(x) + \sum_i \lambda_i \nabla^2 g_i(x), & A^T, & \left(\frac{\partial g(x)}{\partial x}\right)^T \\ A & , 0 & , 0 \\ \frac{\partial g(x)}{\partial x} & , 0 & , -\Lambda^{-1} S \end{bmatrix} \begin{bmatrix} d^x \\ d^\mu \\ d^\lambda \end{bmatrix} = \ldots$$

$$-\Lambda^{-1} S d^\lambda + \frac{\partial g(x)}{\partial x} d^x = \ldots \Rightarrow d^\lambda = S^{-1}\Lambda \frac{\partial g(x)}{\partial x} d^x + \ldots$$

$$\begin{bmatrix} \nabla^2 f(x) + \sum_i \lambda_i \nabla^2 g_i(x) + \left(\frac{\partial g(x)}{\partial x}\right)^T S^{-1} \Lambda \frac{\partial g(x)}{\partial x} & , A^T \\ A & , 0 \end{bmatrix} \begin{bmatrix} d^x \\ d^\mu \end{bmatrix} = \ldots$$

семинар

Двойственность Фенхеля

(p) $\boxed{\begin{array}{l} \min_x \ f(x) + g(Ax) \\ s.t. \ x \in E, \ Ax \in F \end{array}}$ $\qquad f: E \to \mathbb{R}, \ f^*: E^* \to \mathbb{R}$

$\qquad g: F \to \mathbb{R}, \ g^*: F^* \to \mathbb{R}$

(D) $\boxed{\begin{array}{l} \min_\lambda \ g^*(\lambda) + f^*(-A^*\lambda) \\ s.t. \ \lambda \in E^*, \ A^*\lambda \in -F^* \end{array}}$ $\qquad f^*(\lambda) = \sup_{x \in E} \{\langle \lambda, x\rangle - f(x)\}$

$\qquad \|\lambda\|_* = \max_{\|x\|=1} \langle \lambda, x \rangle$

① $f: V \to \mathbb{R}$, $\quad f(x) = \|x\|$

$f^*: \bar{B}_{\|\cdot\|_*}(0,1) \to \mathbb{R}$, $\quad f^*(\lambda) = 0$

$\qquad = \{\lambda : \|\lambda\|_* \le 1\}$

② $f: V \to \mathbb{R} \quad f(x) = \frac{1}{2}\|x\|^2$

$f^*: V \to \mathbb{R} \quad f^*(\lambda) = \frac{1}{2}\|\lambda\|_*^2$

① Сепарабельность

$$f(x_1 \ldots x_n) = \sum_{i=1}^n g_i(x_i), \quad g_i: E_i \to \mathbb{R}, \ f: E_1 \times \ldots \times E_n \to \mathbb{R}$$

$$f^*(\lambda_1 \ldots \lambda_n) = \sum_{i=1}^n g_i^*(\lambda_i)$$

$$f^*(\lambda_1 \dots \lambda_n) = \sup_{\substack{x_1 \in E_1 \\ \dots \\ x_n \in E_n}} (\langle \lambda, x \rangle - f(x_1 \dots x_n)) =$$

$$= \sup_{\substack{x_1 \in E_1 \\ \dots \\ x_n \in E_n}} ((\lambda_1 x_1 - f(x_1)) + \dots + (\lambda_n x_n - f(x_n))) =$$

$$= \sup_{x_1 \in E_2} (\lambda_1 x_1 - f(x_1)) + \dots + \sup_{x_n \in E_n} (\lambda_n x_n - f(x_n)) = \sum_{i=1}^{m} f^*(\lambda_i)$$

② Аффинная композиция

$$g : S \to \mathbb{R}, \quad f : \underbrace{L^{-1}(S - b)}_{S'} \to \mathbb{R}, \quad f(x) = g(Lx + b)$$

$$f^*(\lambda) = \sup_{x \in S'} (\langle \lambda, x \rangle - g(\underbrace{Lx + b}_{y})) = \sup_{y \in S} (\langle \lambda, x \rangle - g(y)) =$$

$$= \sup_{y \in S} (\langle \lambda, L^{-1}(y - b) \rangle - g(y)) = \sup_{y \in S} (\langle L^{-T}\lambda, y \rangle - g(y)) -$$

$$- \langle L^{-T}\lambda, b \rangle = g^*(L^{-T}\lambda) - \langle L^{-T}\lambda, b \rangle$$

$$f^*(\lambda) = g^*(L^{-*}\lambda) - \langle L^{-*}\lambda, b \rangle, \quad L^{-*}\lambda \in S^*$$

$$f(x) = \tau \|x\| = \|\underbrace{\tau}_{L} x\|, \quad f^*(\lambda) = \|\tfrac{1}{\tau}\lambda\|_*, \quad \|\tfrac{\lambda}{\tau}\|_* \le 1$$

$$\|\lambda\|_* \le \tau, \quad f^* : \overline{B}_{\|\|_*}(0, \tau) \to \mathbb{R}$$

① Ridge Regression

$$\min_{x \in \mathbb{R}^n} \frac{\tau}{2}\|x\|^2 + \frac{1}{2}\|Ax - b\|^2$$

$$f(x) = \frac{\tau}{2}\|x\|^2, \quad f^*(\lambda) = \frac{1}{2\tau}\|\lambda\|_*^2$$

$$g(y) = \frac{1}{2}\|y - b\|^2, \quad g^*(\lambda) = \frac{1}{2}\|\lambda\|_*^2 + \langle \lambda, b \rangle$$

$$\min_{\lambda \in \mathbb{R}^m} \frac{1}{2}\|\lambda\|_*^2 + \langle \lambda, b \rangle + \frac{1}{2\tau}\|-A^*\lambda\|_*^2$$

② Lasso

$$\min_{x\in\mathbb{R}^n} \tau\|x\|_1 + \frac{1}{2}\|Ax-b\|_2^2$$

$$\min_{\|A^*\lambda\|_\infty\le\tau} \frac{1}{2}\|\lambda\|_* + \langle\lambda, b\rangle$$

③ Логистическая регрессия

$$\min_{x\in\mathbb{R}^n}\sum_{i=1}^m \ln\left(1+e^{\langle a_i,x\rangle}\right)$$

$f:\mathbb{R}^n\to\mathbb{R}$, $f(x)=0$ $\qquad f^*:\{0\}\to\mathbb{R}$ $f^*(0)$

$g:\mathbb{R}^m\to\mathbb{R}$, $g(y)=\sum_{i=1}^m\ln(1+e^{y_i})$

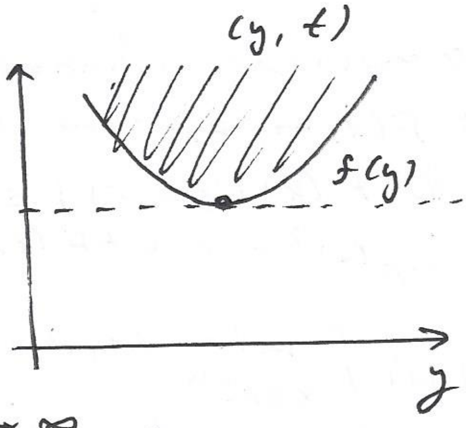
$\|$ $g(y)=\ln(1+e^y)$, $g^*(s)=s\ln s+(1-s)\ln(1-s)$

$g^*:(0,1)\to\mathbb{R}$

$$g^*(s)=\sum_{i=1}^m\left(s_i\ln s_i+(1-s_i)\ln(1-s_i)\right)$$

$As=0$ и $f^*(0)$ определяет рабочее подпространство

Методы внутренней точки.

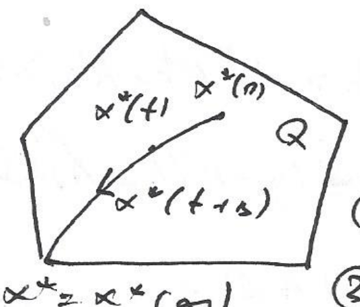


$$\min_{y\in S} f(y) \iff \min_{y\in S,\, t\in\mathbb{R}} t \quad \text{s.t. } t\ge f(y)$$

$$\boxed{\min_{x\in Q}\langle c,x\rangle}$$

$F:\operatorname{int}Q\to\mathbb{R}$, $\displaystyle F(x)\xrightarrow[x\to x_0\in\partial Q]{}\infty$

$\varphi_t(x)=t\langle c,x\rangle+F(x)$, $\displaystyle x^*(t)=\arg\min_{x\in\operatorname{int}Q}\varphi_t(x)$



Самосогласованные ф-ии

$$|D^3F(x)[v,v,v]|\le 2\left(D^2F(x)[v,v]\right)^{3/2}$$

① $F(t)=-\ln(t)$ $\qquad \varphi(x)=\frac{1}{2}\langle Ax,x\rangle+\langle b,x\rangle+c$

② $F(x)=-\ln(-\varphi(x))$ $\qquad D=\{x:\varphi(x)<0\}$

③ $F(x)=-\ln\operatorname{Det}X$

Область квадратичной сходимости, оценка для самосогласованных ф-ий

$\lambda_F(x) = \langle \nabla^2 F(x)^{-1} \nabla F(x), \nabla F(x) \rangle^{\frac{1}{2}}$

Если $\lambda_F(x_0) \le \beta^{*\frac{1}{4}}$, $x_1 = x_0 - \nabla F(x_0)^{-1} \nabla F(x_0)$

$x_1 \in \mathrm{int}\, Q$, $\lambda_F(x_1) \le \lambda_F(x_0)^2$

Центральный путь: $x^*(t) = \underset{x \in \mathrm{int}\, Q}{\mathrm{argmin}} \{ \underbrace{t\langle c, x \rangle + F(x)}_{\varphi_t} \}$

$\overbrace{t c + \nabla F(x^*(t)) = 0}$

$t \to t + \Delta$, $\lambda_{\varphi_{t+\Delta}}(x^*(t)) \le \beta$, $\nabla^2_{\varphi_t}(x) = \nabla^2 F(x)$

$\nabla_{\varphi_t}(x) = ct + \nabla F(x)$

$\langle \nabla^2 F(x^*(t))^{-1} ((t+\Delta)c + \nabla F(x^*(t)), (t+\Delta)c\, \nabla F(x^*(t)) \rangle^{\frac{1}{2}}$

$\parallel$

$\langle \nabla^2 F(x^*(t))^{-1} \Delta c, \Delta c \rangle^{\frac{1}{2}} = \Delta \langle \nabla^2 F(x^*(t))^{-1} c, c \rangle^{\frac{1}{2}} = \frac{\Delta}{t} \le \beta$

$\Delta \le \frac{\beta t}{\partial}$, $\frac{\Delta}{t} \partial \le \beta$, $t + \Delta = (1 + \frac{\beta}{\partial}) t$

Нужно:

$\lambda^2_F(x) = \langle \nabla^2 F(x)^{-1} \nabla F(x), \nabla F(x) \rangle \le \partial$

Самосогласованные барьеры

⓪ $F(x) \to \infty$, $x \to \partial Q$ — барьерная ф-ия

① $|D^3 F(x)[v,v,v]| \le 2 (D^2 F(x)[v,v])^{3/2}$

② $\lambda_F(x)^2 = \langle \nabla^2 F(x)^{-1} \nabla F(x), \nabla F(x) \rangle \le \partial$

$$(LP) \begin{cases} \min\limits_{x \in \mathbb{R}^n} \langle c, x \rangle \\ \text{s.t.} \quad Ax = b \\ \quad x \ge 0 \end{cases}$$

$\varphi_t(x) = t\langle c, x \rangle - \overbrace{\sum_{i=1}^{n} \ln x_i}$

$\langle \nabla \varphi_t(x), h \rangle + \frac{1}{2} \langle \nabla^2 \varphi_t(x) h, h \rangle \to \min\limits_{Ah=0}$

ньютоновский шаг $x \to x + h$

$\nabla \varphi_t(x) = tc - \frac{1}{x}$, $\nabla^2 \varphi_t(x) = \mathrm{diag}(\frac{1}{x_1^2} \dots \frac{1}{x_n^2})$

$\langle tc - \frac{1}{x}, h \rangle + \langle \mathrm{diag}(\frac{1}{x_1^2} \dots \frac{1}{x_n^2}) h, h \rangle \to \min\limits_{Ah=0}$

$L(h, \lambda) = \langle \overbrace{tc - \frac{1}{x}}^{g}, h \rangle + \frac{1}{2} \langle \overbrace{\mathrm{diag}(\frac{1}{x_1^2} \dots \frac{1}{x_n^2})}^{\mathcal{D}} h, h \rangle + \cancel{\lambda^T A h} \langle Ah, \lambda \rangle$

$\begin{cases} \nabla L = g + \mathcal{D} h + A^T \lambda = 0 \\ Ah = 0 \end{cases}$ $\Big|$ $h = -\mathcal{D}^{-2}(A^T \lambda + g)$, $A\mathcal{D}^{-2}(g + A^T\lambda) = 0$ $\Big|$ $h = -\mathcal{D}^{-2}(g - A^T (A\mathcal{D}^{-1}A^T)^{-1} A \mathcal{D}^{-1} g)$

$O(m^2 n + m^3)$ $\quad A\mathcal{D}^{-1} g + A\mathcal{D}^{-1} A^T \lambda = d \Rightarrow \lambda = -(A\mathcal{D}^{-1}A^T)^{-1} A \mathcal{D}^{-1} g$