20.11.2017 лекция XI

Композитные функции

$$F(w) = f(w) + h(w) \to \min_w$$

$f \in C^1$, вып., $h \in C, \notin C^1$, вып., простая

пример:

$$f(w) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, w^T x_i), \quad h(w) = \lambda \|w\|_2 \text{ или } \lambda \|w\|_{1,2}$$

Проксимальный градиентный метод

$$F(w) = f(w) + h(w) \le \{ f \in C_L^{1,1} \} \le m_k(w) := f(w_k) +$$

$$+ \nabla f(w_k)^T (w - w_k) + \frac{L}{2} \|w - w_k\|^2 + h(w) \to \min_w$$

$$m_k(w) = \frac{L}{2} (w^T w - 2 w^T w_k) + \nabla f(w_k)^T w + h(w) + const =$$

$$= \frac{L}{2} (w^T w - 2 w^T (w_k - \frac{1}{L} \nabla f(w_k))) + h(w) + const =$$

$$= \frac{L}{2} \|w - (w_k - \frac{1}{L} \nabla f(w_k))\|^2 + h(w) + const$$

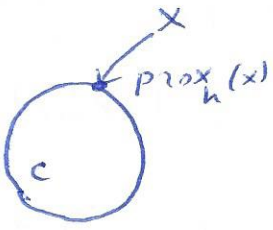
$$prox_h(x) = \arg\min_u \left(\frac{1}{2} \|u - x\|^2 + h(u) \right)$$

Prox-Grad: $w_{k+1} = \arg\min_w m_k(w) = prox_{\frac{1}{L}h}(w_k - \frac{1}{L} \nabla f(w_k))$

Примеры:

① $h(w) = 0 \Rightarrow prox_h(x) = x$

② $h(w) = I_C(w) = \begin{cases} 0, & w \in C \\ \infty, & w \notin C \end{cases}$; C – вып.



$$\min_w F(w) = \min_w (f(w) + I_C(w)) = \min_{w \in C} f(w)$$

$$prox_{I_C(\cdot)}(x) = pr_C(x)$$ метод проекции градиента

③ $h(w) = \lambda \|w\|_2$

$$prox_{\lambda h}(x) = \arg\min_u \left(\frac{1}{2} \|u - x\| + \lambda \|u\|_2 \right) =$$

$$= \{ \arg\min_{u_i} \left(\frac{1}{2} (u_i - x_i)^2 + \lambda |u_i| \right) \}_{i=1}^{d}$$

$$\varphi(u) = \frac{1}{2}(u-x)^2 + \lambda|u| \to \min_{u \in \mathbb{R}}$$

$$\partial\varphi(u) = u - x + \begin{cases} u>0, & \lambda \\ u<0, & -\lambda \\ u=0, & [-\lambda, \lambda] \end{cases}$$

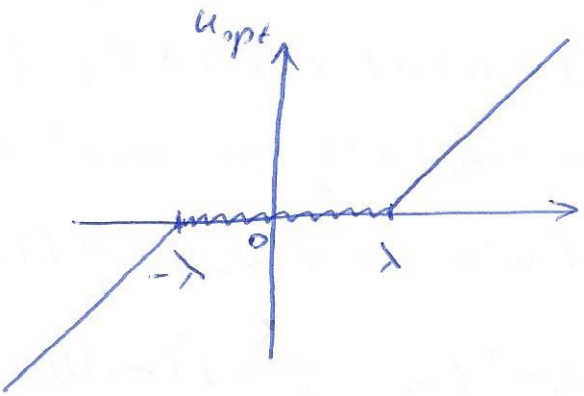
$0 \in \partial\varphi(u_{opt})$

① $u_{opt} > 0$, $u_{opt} - x + \lambda = 0 \Rightarrow u_{opt} = x - \lambda > 0$

② $u_{opt} < 0$, $u_{opt} - x - \lambda = 0 \Rightarrow u_{opt} = x + \lambda < 0$

③ $u_{opt} = 0$, $-x + [-\lambda, \lambda] \ni 0 \Leftrightarrow \begin{cases} -x - \lambda \le 0 \\ -x + \lambda \ge 0 \end{cases} \Leftrightarrow |x| \le \lambda$

$$u_{opt} = \begin{cases} x - \lambda, & x > \lambda \\ x + \lambda, & x < \lambda \\ 0, & |x| \le \lambda \end{cases}$$



Градиентный спуск, сходимость

$f(x) \to \min_x$, $f \in C^{1,1}_L$ и сильно вып. с $\mu > 0$

$$y = x - \frac{1}{L}\nabla f(x)$$

$$f(y) \overset{f \in C^{1,1}_L}{\le} f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|^2 = \left\{y = x - \frac{1}{L}\nabla f(x)\right\} =$$

$$= f(x) - \frac{1}{L}\|\nabla f(x)\|^2 + \frac{1}{2}\frac{1}{L^2}\|\nabla f(x)\|^2 = f(x) - \frac{1}{2L}\|\nabla f(x)\|^2 \le$$

$$\le \left\{ f(x_{opt}) \ge f(x) + \nabla f(x)^T(x_{opt} - x) + \frac{\mu}{2}\|x_{opt} - x\|^2 \right\} \le$$

$$\le f(x_{opt}) + \nabla f(x)^T(x - x_{opt}) - \frac{1}{2L}\|\nabla f(x)\|^2 - \frac{\mu}{2}\|x - x_{opt}\|^2 =$$

$$= f(x_{opt}) + \frac{L}{2}\left(\|x - x_{opt}\|^2 - \|x - \frac{1}{L}\nabla f(x) - x_{opt}\|^2\right) - \frac{\mu}{2}\|x - x_{opt}\|^2 =$$

$$= f(x_{opt}) + \frac{1}{2}\left(\left(1 - \frac{\mu}{L}\right)\|x - x_{opt}\|^2 - \|y - x_{opt}\|^2\right)$$

$$0 \le f(y) - f(x_{opt}) \le \frac{1}{2}\left(\left(1 - \frac{\mu}{L}\right)\|x - x_{opt}\|^2 - \|y - x_{opt}\|^2\right)$$

$\|y - x_{opt}\|^2 \le \left(1 - \frac{\mu}{L}\right)\|x - x_{opt}\|^2$ линейная скорость

$L \leftarrow L/2$

повторять

$y = x - \frac{1}{L} \nabla f(x)$

если $f(y) \le f(x) - \frac{1}{2L} \|\nabla f(x)\|^2$, то выход

$L \leftarrow L \cdot 2$

Проксимальный градиентный метод

$F(x) = f(x) + h(x) \to \min_x$, $f \in C^{1,1}_L$ и сильно вып. $\Rightarrow$

$h \in C \notin C^1$, вып., простая

$y = prox_{\alpha h}(x - \alpha \nabla f(x)) = x - \alpha G_\alpha(x)$

$G_\alpha(x) = \frac{1}{\alpha}(x - y) = \frac{1}{\alpha}(x - prox_{\alpha h}(x - \alpha \nabla f(x)))$

градиентное отображение

$y = prox_{\alpha h}(x - \alpha \nabla f(x)) \Rightarrow 0 \in y - (x - \alpha \nabla f(x)) + \alpha \partial h(y) \mid :\alpha$

$G_\alpha(x) - \nabla f(x) \in \partial h(y)$

утв $x = \arg\min_x F(x) \Leftrightarrow G_\alpha(x) = 0$

□ $G_\alpha(x) = 0 \Leftrightarrow y = x \Leftrightarrow G_\alpha(x) - \nabla f(x) \in \partial h(x) \Leftrightarrow$

$\Rightarrow 0 \in \nabla f(x) + \partial h(x) \Leftrightarrow x = \arg\min_x F(x)$

Критерий останова: $\|G_\alpha(x)\|^2 \le \varepsilon$

$F(y) = f(y) + h(y) \le \{ f \in C^{1,1}_L \} \le f(x) + \nabla f(x)^T (y - x) +$

$+ \frac{L}{2}\|y - x\|^2 + h(y) = \{ y = x - \alpha G_\alpha(x) \} = f(x) - \alpha \nabla f(x)^T G_\alpha(x) +$

$+ \frac{L}{2}\alpha^2 \|G_\alpha(x)\|^2 + h(y) \le \{ \alpha \le \frac{1}{L} \} \le f(x) - \nabla f(x)^T G_\alpha(x) +$

$+ \frac{\alpha}{2}\|G_\alpha(x)\|^2 + h(y) \le \left\{ \begin{array}{l} f(x_{opt}) \ge f(x) + \nabla f(x)^T (x_{opt} - x) + \frac{\mu}{2}\|x_{opt} - x\|^2 \\ h(x_{opt}) \ge h(y) + (G_\alpha(x) - \nabla f(x))^T (x_{opt} - y) \end{array} \right. \le$

$\le f(x_{opt}) + h(x_{opt}) + \cancel{\nabla f(x)^T (x - x_{opt})} - \alpha \cancel{\nabla f(x)^T} G_\alpha(x) + \frac{\alpha}{2}\|G_\alpha(x)\|^2 -$

$- \frac{\mu}{2}\|x - x_{opt}\|^2 + (G_\alpha(x) - \cancel{\nabla f(x)})^T (x - \alpha G_\alpha(x) - x_{opt}) \Rightarrow$

$\Leftarrow F(x_{opt}) - \frac{\alpha}{2}\|G_\alpha(x)\|^2 + G_\alpha(x)^T(x - x_{opt}) - \frac{\mu}{2}\|x - x_{opt}\|^2 =$

$= F(x_{opt}) + \frac{1}{2\alpha}(\|x - x_{opt}\|^2 - \|x - \alpha G_\alpha(x) - x_{opt}\|^2) -$

$- \frac{\mu}{2}\|x - x_{opt}\|^2 = F(x_{opt}) + \frac{1}{2\alpha}((1 - \alpha\mu)\|x - x_{opt}\|^2 -$

$- \|y - x_{opt}\|^2)$

$0 \le F(y) - F(x_{opt}) \le \frac{1}{2\alpha}((1 - \alpha\mu)\|x - x_{opt}\|^2 - \|y - x_{opt}\|^2)$

$\|y - x_{opt}\|^2 \le (1 - \alpha\mu)\|x - x_{opt}\|^2$

линейная скорость

$\alpha = \frac{1}{L}$, $L \leftarrow L/2$

повторять

$y = x - \frac{1}{L} G_{\frac{1}{L}}(x)$

если $f(y) \le f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|^2$, то выход

$L \leftarrow L \cdot 2$

Проксимальный метод Ньютона

$F(x) = f(x) + h(x) \approx m_u(x) = f(x_u) + \nabla f(x_u)^T(x - x_u) +$

$+ \frac{1}{2}(x - x_u)^T \nabla^2 f(x_u)(x - x_u) + h(x) \to \min\limits_x$

масштабированный проксимальный оператор

$prox_h^H(x) = \arg\min\limits_u \left(\frac{1}{2}(u - x)^T H (u - x) + h(u)\right)$

$\begin{cases} \hat{x}_{u+1} = \arg\min m_u(x) = prox_h^{H_u}(x_u - H_u^{-1}\nabla f(x_u)), \; H_u = \nabla^2 f(x_u) \\ x_{u+1} = x_u + \alpha_u(\hat{x}_{u+1} - x_u), \; H_u = \nabla^2 f(x_u) \end{cases}$

α_u : $F(x_{u+1}) \le F(x_u) + c_2\alpha_u(\nabla f(x_u)^T(\hat{x}_{u+1} - x_u) +$

$+ h(\hat{x}_{u+1}) - h(x_u))$, · дробление шага

Критерий останова: $\| G_{\gamma,L}(x_k) \|^2 \le \varepsilon$

$G_{\gamma,L}(x) = L\left(x - \operatorname{prox}_{\frac{\gamma}{2}h}\left(x - \frac{1}{L}\nabla f(x)\right)\right)$

$\operatorname{prox}_h^H(x) = \operatorname{argmin}_u \left(\underbrace{\frac{1}{2}(u-x)^T H (u-x) + h(u)}_{S(u)} \right)$

prox-grad: $u_{k+1} = \operatorname{prox}_{\frac{\gamma}{2}h}\left(u_k - \frac{1}{L}\nabla S(u_k)\right)$

$\hat{G}_{\gamma,L}(u_k) = L\left(u_k - \operatorname{prox}_{\frac{\gamma}{2}h}\left(u_k - \frac{1}{L}\nabla S(u_k)\right)\right)$

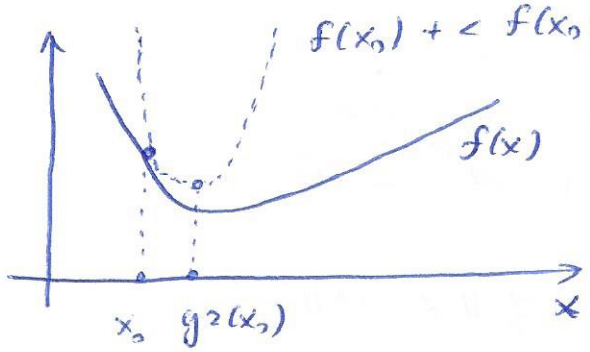
фундирующая посл-ть: $\| \hat{G}_{\gamma,L}(u_k) \| \le i_{2k} \| G_{\gamma,L}(x_k) \|$

семинар

$\min_{x \in \mathbb{R}^n} f(x)$, $f(x) \le f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2}\|x - x_0\|^2$

$\operatorname{grad}_f(x_0) = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2}\|x - x_0\|^2 \right\} = x_0 - \frac{1}{L}\nabla f(x_0)$

$\min_{x \in Q} f(x)$, $\operatorname{grad}_f(x_0) = \pi_Q\left(x_0 - \frac{1}{L}\nabla f(x_0)\right)$



$\pi_Q(v) = \operatorname{argmin}_{x \in Q} \frac{1}{2}\|x - v\|^2$

$\frac{1}{2}\|x - v\|^2 = \frac{1}{2}\|x - x_0\|^2 + \frac{2}{L}\langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2L^2}\|\nabla f(x_0)\|^2$

аргминимум верхней модели перед оценку

Липшица на градиент

$\min_{x \in Q} \{ \underbrace{f(x) + h(x)}_{\varphi(x)} \}$

$\varphi(x) \le f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{2}{2}\|x - x_0\|^2 + h(x), \quad \forall x \in Q$

$$\mathrm{Grad}_{f,h}(x_0) = \underset{x\in Q}{\operatorname{argmin}} \{ f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2}\|x - x_0\|^2 + h(x) \} = \operatorname{prox}_{\frac{1}{L}h}\left(x_0 - \frac{1}{L}\nabla f(x_0)\right)$$

$$\Pi_Q(v) = \underset{x\in Q}{\operatorname{argmin}} \frac{1}{2}\|x - v\|$$

$$\operatorname{prox}_{Q,f}(v) = \underset{x\in Q}{\operatorname{argmin}} \left\{ \frac{1}{2}\|x - v\|^2 + f(x) \right\}, \quad f: Q \to \mathbb{R}$$

$$v = x_0 - \frac{1}{L}\nabla f(x_0)$$

$f: Q \to \mathbb{R}$

$$\operatorname{prox}_f(v) = \underset{x\in Q}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2}\|x - v\|^2 \right\}$$

f - выпуклая, Q замкнутое, $\operatorname{epi} f$ замкнутое, Q вып.

$\exists!$ Q вып. замкнут. $\Pi_Q(v) = \underset{x\in Q}{\operatorname{argmin}} \|x - v\|^2$

$$\boxed{\operatorname{prox}_{\delta_Q}(v) = \Pi_Q(v)}$$

① $f(x) = \lambda\|x\|_2$, $f: \mathbb{R}^n \to \mathbb{R}$

② $f(x) = \lambda\|x\|_2$

③ $\Pi_{B_2(0,1)}(v)$

④ $f(x) = \lambda\|x\|_\infty$

⑤ $f(x) = \|x\|_* = \sum_{i=1}^{r} \sigma_i(x)$, $f: \mathbb{R}^{m\times n} \to \mathbb{R}$ — ядерная норма

$$\operatorname{prox}_{\lambda\|\cdot\|_2}(v) = \underset{x}{\operatorname{argmin}} \left\{ \lambda\|x\|_2 + \frac{1}{2}\|x - v\|^2 \right\} =$$

$$= \left\{ \underset{x_j}{\operatorname{argmin}} \{ \underbrace{\lambda|x_j| + \frac{1}{2}(x_j - v_j)^2}_{p(t) \quad \text{при } x_j = t} \} \right\}$$

$$\partial p(t) = t - v_j + \begin{cases} \lambda, & t > 0 \\ -\lambda, & t < 0 \\ [-\lambda, \lambda], & t = 0 \end{cases} = \begin{cases} \lambda + t - v_j, & t > 0 \\ -\lambda + t - v_j, & t < 0 \\ [-\lambda, \lambda] - v_j, & t = 0 \end{cases}$$

$$0 \in \partial p(t) \iff \begin{cases} t = v_j - \lambda, & t > 0 \\ t = v_j + \lambda, & t < 0 \\ v_j \in [-\lambda, \lambda], \ |v_j| \le \lambda & \text{при } t = 0 \end{cases}$$

$$prox_{\lambda\|\cdot\|_2}(v) = \begin{cases} v_i - \lambda, & v_i > \lambda \\ v_i + \lambda, & v_i < -\lambda \\ 0, & |v_i| \le \lambda \end{cases} \qquad \|\cdot\|_2 = 0$$

② $prox_{\lambda\|\cdot\|_2}(v)$?

$$0 \in \partial\left(\frac{1}{2}\|v-x\|_2^2 + \lambda\|x\|_2\right) = x - v + \lambda \begin{cases} B(0), & x = 0 \\ \frac{x}{\|x\|}, & x \neq 0 \end{cases} =$$

$$= \begin{cases} \lambda B(0) - v, & x = 0 \\ x - v + \frac{\lambda x}{\|x\|}, & x \neq 0 \end{cases}$$

$x \neq 0 \qquad x - v + \frac{\lambda x}{\|x\|} = 0, \quad x + \frac{\lambda x}{\|x\|} = v,$

$x\left(1 + \frac{\lambda}{\|x\|}\right) = v, \quad \|x\|\left(1 + \frac{\lambda}{\|x\|}\right) = \|v\|, \quad \|x\| + \lambda = \|v\| \; \|x\| = \|v\| - \lambda$

$x\left(1 + \frac{\|v\| - \lambda}{\|v\| - \lambda} \cdot \frac{\lambda}{\|v\| - \lambda}\right) = v, \quad x\left(\frac{\|v\|}{\|v\| - \lambda}\right) = v$

$$x = \frac{\|v\| - \lambda}{\|v\|} v$$

$$prox_{\lambda\|\cdot\|_2}(v) = \begin{cases} 0, & v \in \lambda B(0) \\ \frac{\|v\|_2 - \lambda}{\|v\|_2} v, & \text{иначе} \end{cases} \qquad \|v \in \lambda B(0), \|v\|_2 \le \lambda$$

③ $\Pi_{B_2(0,1)}(v) = \underset{\|x\|_2 \le 1}{\operatorname{argmin}} \frac{1}{2}\|x - v\|^2 = z$

$\mathcal{L}(x, \lambda) = \frac{1}{2}\|x - v\|^2 + \lambda(\|x\|_1 - 1)$ $\qquad z = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \mathcal{L}(x, \lambda)$

$\lambda(\|x\|_1 - 1) = 0$ проекция на ромб

$\lambda \ge 0$

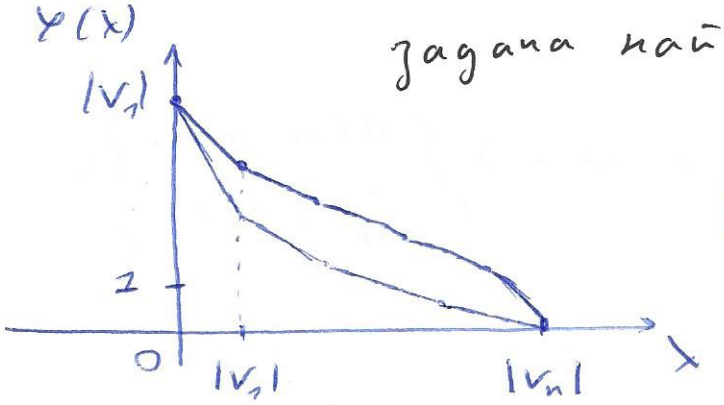
1) $\lambda = 0, \quad z \in B_2(0,1), \quad$ то $z = v$

2) $\|v\|_2 > 1, \quad z = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \mathcal{L}(x, \lambda) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{\frac{1}{2}\|x - v\|_2^2 + \lambda\|x\|_1\right\}$

$$z_i = \begin{cases} v_i - \lambda, & v_i > \lambda \\ v_i + \lambda, & v_i < -\lambda \\ 0, & |v_i| \le \lambda \end{cases} \qquad \begin{matrix} z_i = v_i - \lambda \\ z_i = |v_i| - \lambda \end{matrix} \Rightarrow \begin{matrix} z_i = |v_i| - \lambda \\ z_i = |v_i| - \lambda \end{matrix}$$

$$\|z\|_1 = \sum_{i=1}^{n} |z_i| = \sum_{i=1}^{n} \max(0, |v_i| - \lambda) = 1$$

$$\varphi(\lambda) = \sum_{i=1}^{n} \max(0, |v_i| - \lambda), \quad |v_1| \le \dots \le |v_n|$$



задача найти λ

$$\varphi(\lambda) = \sum_{i=k}^{n} |v_i| - (n-k)\lambda$$

итерационный поиск λ

перед сортировку $v_1 \dots v_n$

и бинарный поиск max

$0 \le \lambda \le |v_1|$

$$\varphi(\lambda) = \sum_{i=2}^{n} |v_i| - n\lambda$$

$|v_1| < \lambda \le |v_2|$

$$\varphi(\lambda) = \sum_{i=2}^{n} |v_i| - (n-1)\lambda$$

⑤ $\operatorname{prox}_{\lambda\|\cdot\|_2}(A) = U \operatorname{Diag}(\operatorname{prox}_{\lambda\|\cdot\|_2}(\sigma)) V^T$

$A = U \operatorname{Diag}(\sigma) V^T$ сингул. разл-е

$$\min_{X \in \mathbb{R}^{m\times n}} \left\{ \frac{1}{2}\|X - A\|_F^2 + \lambda\|X\|_* \right\} = \dots$$

$$= \min_X \left\{ \frac{1}{2}\|U(\underbrace{U^T X V}_{Y} - \Sigma)V^T\|_F^2 + \lambda\|X\|_* \right\}$$

$$\min_Y \left\{ \frac{1}{2}\|Y - \Sigma\|_F^2 + \lambda\|Y\|_* \right\}$$

$$Y \to \operatorname{Diag}(Y), \quad \sum_{i=1}^{n} |y_{ii}| \le \sum_{i=1}^{n} \sigma_i(Y)$$

$$\min_{y \in \mathbb{R}^n} \left\{ \frac{1}{2}\|y - \sigma\|^2 + \lambda\|y\|_1 \right\}$$

④ $$\min_x \left\{ \frac{1}{2}\|x - v\|^2 + \lambda\|x\|_\infty \right\} \quad (P)$$

$$\begin{cases} \min\limits_y \frac{1}{2}\|y - w\|^2 \\ \text{s.t. } \|y\|_1 \le \lambda \end{cases} \quad (D)$$