Выпуклый L-гладкая

$$\|x_{k+1}-x^*\|^2 \le \|x_k - x^*\|^2 - 2\gamma_k\left(\frac{\mu}{2}\|x_k - x^*\|^2 + f(x_k) - f(x^*)\right) + \gamma_k^2 \cdot 2L\left(f(x_k) - f(x^*)\right)$$

$\mu = 0$

$$\underbrace{\|x_{k+1}-x^*\|^2}_{= R_{k+1}^2} \le \underbrace{\|x_k - x^*\|^2}_{= R_k^2} - 2\gamma_k\left(\qquad + f(x_k) - f(x^*)\right) + \gamma_k^2 \cdot 2L\left(f(x_k) - f(x^*)\right)$$

$(1-\mu\gamma)\|x^k - x^*\|^2$?

$$R_{k+1}^2 \le R_k^2 - 2\gamma_k(1-\gamma_k L)\left(f(x_k) - f(x^*)\right)$$

$\gamma_k = \frac{1}{2L}$ $\left(\frac{1}{L}\ \text{Nesterov}\right)$

$$R_{k+1}^2 \le R_k^2 - \frac{1}{2L}\left(f(x_k) - f(x^*)\right)$$

$$f(x_k) - f(x^*) \le 2L\left(R_k^2 - R_{k+1}^2\right)$$

$\sum_{k=0}^{K}$

$$\boxed{\sum_{k=0}^{K} f(x_k) - f(x^*)} \le 2L\left(R_0^2 - R_{K+1}^2\right) \le 2L R_0^2$$

$$\frac{1}{K+1}\sum_{k=0}^{K} f(x_k) - f(x^*) \le \frac{2L R_0^2}{K+1}$$

сублинейная

1) $\tilde{x}_K$ — случайно из $x_0 \dots x_K$ (равномерно)

$$\mathbb{E}\left[f(\tilde{x}_K) - f(x^*)\right] = \sum_{k=0}^{K} \frac{1}{K+1}\left(f(x_k) - f(x^*)\right)$$

$$\mathbb{E}\left[f(\tilde{x}_K) - f(x^*)\right] \le \frac{2LR_0^2}{K+1}$$

2) $f((1-\alpha)x + \alpha y) \le (1-\alpha)f(x) + \alpha f(y)$ — Йенсен

$$f\left(\frac{1}{K+1}\sum_{k=0}^{K} x_k\right) \le \frac{1}{K+1} f(x_0) + \dots + \frac{1}{K+1} f(x_K)$$

$$\tilde{x}^K = \frac{1}{K+1}\sum_{k=0}^{K} x_k$$

$$f(\tilde{x}^K) - f(x^*) \le \frac{2LR_0^2}{K+1}$$

3) $$f(x^{k+1}) \le f(x^k) + \langle \nabla f(x^k); x^{k+1} - x^k \rangle + \frac{L_k}{2}\|x^{k+1} - x^k\|^2$$

$L = 0.001$

$\frac{1}{L_k}$

$$= f(x^k) - \gamma \langle \nabla f(x^k); \nabla f(x^k) \rangle + \frac{L\gamma^2}{2}\|\nabla f(x^k)\|^2$$

$$= f(x^k) - \gamma\left(1 - \frac{L\gamma}{2}\right)\|\nabla f(x^k)\|^2$$

$\gamma = \frac{1}{2L}$

$$= f(x^k) - \frac{3}{8L}\|\nabla f(x^k)\|^2$$

$$f(x^{k+1}) - f(x^*) \le f(x^k) - f(x^*) - \frac{3}{8L}\|\nabla f(x^k)\|^2$$

$$f(x^{k+1}) - f(x^*) \le f(x^k) - f(x^*)$$

$$\frac{1}{K+1}\sum_{k=0}^{K} f(x^{K+1}) - f(x^*) \le \frac{1}{K+1}\sum_{k=0}^{K} f(x^k) - f(x^*)$$

Невыпуклый L-гладкий $\exists x^*$ $\forall x$ $f(x^*) \le f(x)$

$$f(x^{k+1}) \le f(x^k) - \gamma\left(1 - \frac{L\gamma}{2}\right)\|\nabla f(x^k)\|^2$$

$$\gamma = \frac{1}{L}$$

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

$$\|\nabla f(x^k)\|^2 \le \left(f(x^k) - f(x^{k+1})\right)\cdot 2L$$

$$\sum_{k=0}^{K}$$

$$\sum_{k=0}^{K}\|\nabla f(x^k)\|^2 \le 2L\left(f(x^0) - f(x^{K+1})\right)$$

$$= 2L\left(f(x^0) - f(x^*) + \underbrace{f(x^*) - f(x^{K+1})}_{\le 0}\right)$$

$$\le 2L\left(f(x^0) - f(x^*)\right)$$

1) стохастический выбор

2) последняя итерация

3)
$$\frac{1}{K+1}\sum_{k=0}^{K}\|\nabla f(x^k)\|^2 \geq \min_{k=0,..K}\|\nabla f(x^k)\|^2$$

$$\min_{k=0..K}\|\nabla f(x^k)\|^2 \leq \frac{2L\left(f(x^0)-f(x^*)\right)}{K+1}$$

$$\tilde{x}^K = \operatorname*{argmin}_{k=0..K}\left\{\|\nabla f(x^k)\|^2\right\}$$

$$\|\nabla f(\tilde{x}^K)\| \leq \varepsilon$$

$$K+1 \geq \frac{2L\left(f(x^0)-f(x^*)\right)}{\varepsilon}$$

$$\Uparrow$$

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

$$\Downarrow$$

$$\min_{x\in C} f(x)$$

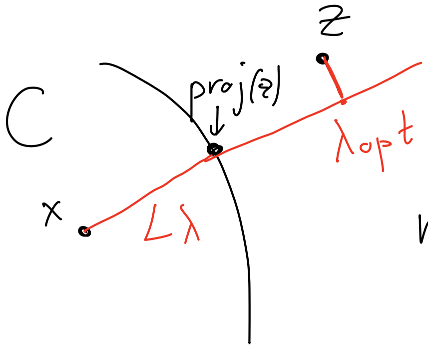
C - „простое"

проекция дешевая и точная

$$x^{k+1} = \mathrm{proj}_C\left(x^k - \gamma\nabla f(x^k)\right) \qquad \mathrm{proj}_C(x) = \operatorname*{argmin}_{y\in C}\|x-y\|$$

оператор проекции единственен!

Свойство оператора проекции (евклидова)



$L_\lambda: \ \lambda x + (1-\lambda)\,\mathrm{proj}\, z \quad \lambda \in [0;1]$

$$\min \ \mathrm{dist}(z, L_\lambda)$$

$$\Downarrow$$

$$\min_{\lambda\in[0;1]} \| z - ((1-\lambda)\,\mathrm{proj}(z) + \lambda x)\|^2 = z$$

т.к. $L_\lambda \in C$

$$\| z - \mathrm{proj}(z) + \lambda(\mathrm{proj}(z) - x)\|^2$$

$$= \| z - \mathrm{proj}(z)\|^2 + \lambda^2 \|\mathrm{proj}(z) - x\|^2$$

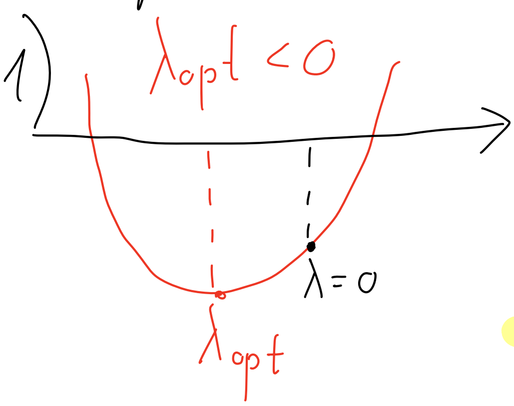
$$+ 2\lambda \langle z - \mathrm{proj}(z); \mathrm{proj}(z) - x\rangle$$

$\lambda = -\frac{b}{2a}$ (в. параболы)

$$\lambda_{opt} = \frac{-\langle z - \mathrm{proj}(z); \mathrm{proj}(z) - x\rangle}{\|\mathrm{proj}(z) - x\|^2}$$

Возможны 3 случая расположения λ_{opt} отн. $[0;1]$

1) $\lambda_{opt} < 0$



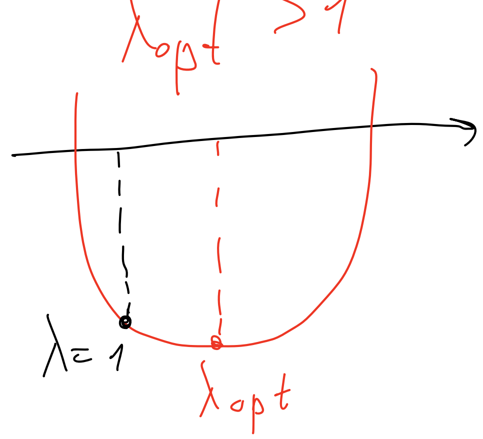
$\Rightarrow \ \lambda_{opt} < 0$

$$\Downarrow$$

$$\langle z - \mathrm{proj}(z); x - \mathrm{proj}(z)\rangle \le 0$$

$$\langle z - \mathrm{proj}(z); \mathrm{proj}(x) - \mathrm{proj}(z)\rangle \le 0$$

2)



это означает, что

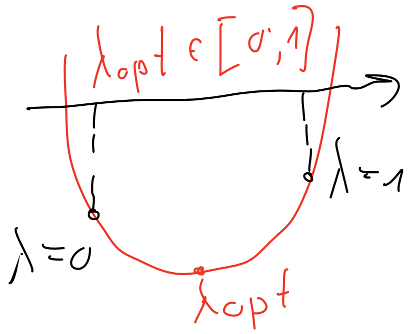
$$dist(z, L_1) \le dist(z, L_0)$$

$$\qquad\qquad\qquad\qquad \| $$

$$\qquad\qquad\qquad dist(z, proj\, z)$$

$$L_1 = proj\, z = x$$

3)



$$\Rightarrow dist(z, L_{\lambda_{opt}}) \le dist(z, L_0)$$

$$\qquad\qquad\qquad\qquad \| $$

$$\qquad\qquad\qquad dist(z, proj\, z)$$

$$L_{\lambda_{opt}} = proj\, z \Rightarrow$$ $\lambda_{opt} = 0$ или $proj(z) = x$

$$\langle z - proj(z) \underbrace{\cancel{-x + proj(x)}}_{=0}; \; proj(x) - proj(z) \rangle \le 0$$

$$\langle z - x - (proj(z) - proj(x)); \; proj(x) - proj(z) \rangle \le 0$$

$$\langle z - x; \; proj(x) - proj(z) \rangle + \| proj(x) - proj(z) \|^2 \le 0$$

$$\| proj(x) - proj(z) \|^2 \le \langle z - x; \; proj(z) - proj(x) \rangle$$

$$\le \| z - x \| \, \| proj(z) - proj(x) \|$$

$$\| proj(x) - proj(z) \|^2 \le \| z - x \|^2$$

$$\| x^{k+1} - x^* \| = \| proj(x^k - \gamma \nabla f(x^k)) - \underset{\substack{\in C \\ proj(x^*)}}{x^*} \|^2$$

$$\le \| x^k - \gamma \nabla f(x^k) - x^* \|^2$$

дальше, как раньше

$$\min_{x\in\mathbb{R}^d} \underbrace{f(x)}_{\text{осн.}} + \underbrace{r(x)}_{\text{композит}}$$

$f(x)$ – выпуклая L-гладкая

$r(x)$ – выпуклая

1) $\frac{\lambda}{2}\|x\|^2$ 2) $\delta_C = \begin{cases} 0 & x\in C \\ +\infty & x\notin C \end{cases}$

3) $\frac{\lambda}{2}\|x\|_1$

1) $g(x) = f(x) + r(x)$

2) использовать $r(x)$ – „простая"

$$\Downarrow$$

$$\min_{x\in\mathbb{R}^d}\left\{\frac{1}{2}\|x-y\|^2 + r(x)\right\}$$ – дешево и точно

1) $x^{k+1} = x^k - \gamma\left(\nabla f(x^k) + \nabla r(x^k)\right)$

2) $x^{k+1} = \mathrm{prox}_{\gamma r}\left(x^k - \gamma\nabla f(x^k)\right)$

$$\mathrm{prox}_{\gamma r}(y) = \arg\min_{x\in\mathbb{R}^d}\left\{\frac{1}{2}\|x-y\|^2 + \gamma r(x)\right\}$$

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^d}\left\{\underbrace{\frac{1}{2}\|x^k - \gamma\nabla f(x^k) - x\|^2 + \gamma r(x)}_{\nabla = 0}\right\}$$

$$-\left(x^k - \gamma\nabla f(x^k) - x_{opt}\right) + \gamma\nabla r(x_{opt}) = 0$$

$$x^{k+1} = x_{opt} = x^k - \gamma\nabla f(x^k) - \gamma\nabla r(x^{k+1})$$

$y_1 = \mathrm{prox}_r(x_1)$ $\qquad$ $y_2 = \mathrm{prox}_r(x_2)$

$$y = \mathrm{prox}_r(x) = \arg\min_z\left(\frac{1}{2}\|z-x\|^2 + r(z)\right)$$

$$y_1 - x_1 + \nabla r(y_1) = 0 \qquad (y_1 - x_1 + \partial r(y_1) \in 0)$$

$$-(y_1 - x_1) = \nabla r(y_1) \qquad -(y_2 - x_2) = \nabla r(y_2)$$

$$\langle \nabla r(y_1) - \nabla r(y_2); y_1 - y_2 \rangle \geq 0$$

$$+\begin{cases} \langle \nabla r(y_1); y_1 - y_2 \rangle \geq r(y_1) - r(y_2) \\ \langle \nabla r(y_2); y_2 - y_1 \rangle \geq r(y_2) - r(y_1) \end{cases}$$

$$\langle y_2 - x_2 - y_1 + x_1; y_1 - y_2 \rangle \geq 0$$

$$-\| y_2 - y_1 \|^2 + \langle x_1 - x_2; y_1 - y_2 \rangle \geq 0$$

$$\| y_1 - y_2 \|^2 \leq \langle x_1 - x_2; y_1 - y_2 \rangle \leq \| x_1 - x_2 \| \, \| y_1 - y_2 \|$$

$$\| y_1 - y_2 \| \leq \| x_1 - x_2 \|$$

$$\| \mathrm{prox}_r(x_1) - \mathrm{prox}_r(x_2) \|^2 \leq \| x_1 - x_2 \|^2$$