Ускоренный (оптимальный) метод Нестерова
гр.м.
Схемы оптимальных методов и их оценки эффективности основанные на понятии оценивающей пос-ти

<u>Опр</u> Пос-ти $\{\varphi_k(x)\}_{k=0}^{\infty}$ и $\{\lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ наз оцен. посл. фии $f(x)$, if $\lambda_k \to 0$ и $\forall x \in \mathbb{R}^n$, $k \geq 0$
$\hookrightarrow$ $\varphi_k(x) \leq (1-\lambda_k) f(x) + \lambda_k \varphi_0(x)$

<u>Лемма</u> if для нек посл-ти $\{x_k\}$ $\hookrightarrow$ $f(x_k) \leq \varphi_k^* = \min\limits_{x \in \mathbb{R}^n} \varphi_k(x)$, то $f(x_k) - f^* \leq \lambda_k [\varphi_0(x^*) - f^*] \to 0$

<u>Д-во</u>: $f(x_k) \leq \varphi_k^* = \min\limits_{x \in \mathbb{R}^n} \varphi_k(x) \leq \min\limits_{x \in \mathbb{R}^n} [(1-\lambda_k) f(x) + \lambda_k \varphi_0(x)]$

$\leq (1-\lambda_k) f(x^*) + \lambda_k \varphi_0(x^*)$

Т.е., построив $\{\lambda_k\}$ $\to$ оценим сх-ть $\{\lambda_k\}$ <u>ЧТД</u>

(**) $\Rightarrow$ Получим общую схему опт ~~втор~~ мин

<u>0</u> $x_0 \in \mathbb{R}^n$, $\gamma_0 > 0$ $\exists v_0 = x_0$

<u>1</u> $k \geq 0$ иг

а) $\alpha_k \in (0,1)$ — ? $\rightarrow$ решение $L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k \mu$

положим $\gamma_{k+1} = (1-\alpha_k)\gamma_k + \alpha_k \mu$

б) выбери $y_k = \dfrac{\alpha_k \gamma_k v_k + \gamma_{k+1} x_k}{\gamma_k + \alpha_k \mu}$

$f(y_k) = \ldots$ $\qquad f'(y_k) = \ldots$

в) Найдя x_{k+1}: $f(x_{k+1}) \leq f(y_k) - \dfrac{1}{2L}\|f'(y_k)\|^2$

г) $v_{k+1} = \dfrac{(1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k f'(y_k)}{\gamma_{k+1}}$

<u>Th 1</u> Эта схема формирует посл $\{x_k\}$:

$f(x_k) - f^* \leq \lambda_k [f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2]$; $\lambda_0 = 1$, $\lambda_k = \prod_{i=1}^{k-1}(1-\alpha_i)$ $\lambda_k \to 0$

<u>Д-во</u>

выб $\varphi_0(x) = f(x_0) + \frac{\gamma_0}{2}\|x - v_0\|^2$

<u>Лемма</u> if в $\overset{\text{схеме}}{(**)}$ $\gamma_0 \geq \mu$, то $\lambda_k \leq \min\{(1-\sqrt{\frac{\mu}{L}}), \frac{4L}{2\sqrt{L}+k\sqrt{\gamma_0}}\}$

<u>Д-во</u>

ч.т.д.

! Оптимальность схемы: <u>Th 2</u> Пусть в схеме (**) $\gamma_0 = L$

Тогда эта схема ген послед посл $\{x_k\}$, что

$f(x_k) - f^* \leq L \min\{(1-\sqrt{\frac{\mu}{L}})^k, \frac{4}{(k+2)^2}\}\|x_0 - x^*\|^2$

что ~~достигает~~ достигает min оценку (!)

и означает, что схема (**) оптимальна для

безусл мин функц $\mathcal{F}^{1,1}_{\mu,L}(\mathbb{R}^n)$, $\mu \geq 0$

Дlo: ... что

Суть град метода: релаксационная последовательность $f(x_{k+1}) \le f(x_k)$

· Сейчас мы близки к получению алгоритма

Предположим, что у нас уже имеем x_k: $\varphi_k^* \ge f(x_k)$

$\Rightarrow$ из пз $\varphi_{k+1}^* \ge (1-\alpha_k) f(x_k) + \alpha_k f(y_k) -$

$$-\frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2 + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\langle f'(y_k), v_k - y_k\rangle$$

т.к. $f(x_k) \ge f(y_k) + \langle f'(y_k), x_k - y_k\rangle$, получим след оценку:

$$\varphi_{k+1}^* \ge f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2 + (1-\alpha_k)\langle f', \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) +$$

$$+ x_k - y_k\rangle.$$

Напомним, хотим, чтобы $\varphi_{k+1}^* \ge f(x_{k+1})$

Мы можем обеспечить выполнение н-ва $f(y_k) - \frac{1}{2L}\|f'(y_k)\|^2 \ge f(x_{k+1})$ разными способ., напр град метода $x_{k+1} = y_k - h_k f'(x_k)$ $h_k = 1/L$. Опр $\alpha_k \in (0,1)$ из ур-и $L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu$ $(= \gamma_{k+1})$

Тогда $\alpha_k^2/2\gamma_k = \frac{1}{2L}$ и можно заменить неравенства $\varphi_{k+1}^* \ge f(x_{k+1}) + (1-\alpha_k)\langle f'(y_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) +$

$+ x_k - y_k\rangle$

· y_k можно выбрать свободно. Найдем его из ур

$$\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(v_k - y_k) + x_k - y_k = 0$$

$$\Rightarrow y_k = \frac{\alpha_k\gamma_k v_k + \gamma_{k+1}x_k}{\gamma_k + \alpha_k\mu}$$

т.о

В построении нужных $\{\lambda_k\}$ помогает

Лемма 2

Пусть

1) $f \in \mathcal{F}^{1,1}_{\mu,L}(\mathbb{R}^n)$

2) $\varphi_0(x)$ = произвол. ф-я на $\mathbb{R}^n$

3) $\{y_k\}$ — ∀ п-п. в $\mathbb{R}^n$

4) $\{\alpha_k\}_{k=0}^{\infty}$: $\alpha_k \in (0,1)$, $\sum \alpha_k = \infty$

5) $\lambda_0 = 1$

Тогда посл-ти $\{\varphi_k(x)\}$, $\{\lambda_k\}$ пост-ты опр. соотн

$$\lambda_{k+1} = (1-\alpha_k)\lambda_k$$

$$\varphi_{k+1} = (1-\alpha_k)\varphi_k(x) + \alpha_k[f(y_k) + \langle f'(y_k), x-y_k\rangle + \frac{\mu}{2}\|x-y_k\|^2] \quad (*)$$

Явл оцен. последоват-и

Д-во:

...

• Т.о. получили правила построения. Есть две ...

Заметим, что $\varphi_0(x)$ можно выбирать произвольно, напр, простую кв. ф-ю, и получим точное опис динамики φ_k^*

Лемма 3 Пусть $\varphi_0(x) = \varphi_0^* + \frac{\gamma_0}{2}\|x-v_0\|^2$. Тогда процесс $(*)$ сохраняет канонич форму $\{\varphi_k(x)\}$

$$\varphi_k(x) = \varphi_k^* + \frac{\gamma_k}{2}\|x-v_k\|^2, \text{ где}$$

$$\gamma_{k+1} \overset{def}{=} (1-\alpha_k)\gamma_k + \alpha_k\mu$$

$$v_{k+1} \overset{def}{=} [(1-\alpha_k)\gamma_k v_k + \alpha_k \mu y_k - \alpha_k f'(y_k)]$$

$$\varphi^*_{k+1} = (1-\alpha_k)\varphi_k + \alpha_k f(y_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(y_k)\|^2 +$$

$$+ \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|y_k - v_k\|^2 + \langle f'(y_k), v_k - y_k\rangle\right)$$

схема (AAA) становится проще, если выбираем $\alpha_0 = \sqrt{\mu/L}$ $(\gamma_0 = \mu)$

$\Rightarrow \alpha_k = \sqrt{\frac{\mu}{L}}$, $\beta_k = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$ $(k \geq 0)$

$\Rightarrow$

Схема спост. шагов III: (AAAA)

0. выберем $y_0 = x_0 \in \mathbb{R}^n$

1. $k \geq 0$

$$x_{k+1} = y_k - \frac{1}{L} f'(y_k)$$

$$y_{k+1} = x_{k+1} + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}} (x_{k+1} - x_k)$$

! не работает при $\mu = 0$ (лучше брать $\gamma_0 = L$ в этом случае)

Наши оценки с сейчас нижней оценки мет. нет

для $F^{1,1}_{\mu,L}$ $k \sim O\left(\sqrt{\frac{L}{\mu}} \ln\left(\frac{LR^2}{\varepsilon}\right)\right) = \cancel{O}$

$q_0 \geq \sqrt{\frac{\mu}{L}}$

$$f(x_k) - f^* \leq \min\left\{\left(1-\sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{2\sqrt{L}+k\sqrt{\gamma_0}}\right\} \cdot \left[f(x_0) - f^* + \frac{\gamma_0}{2}\|x_0 - x^*\|^2\right]$$

из Th 2

o! $f(x_k) - f^* = LR^2\left(1-\sqrt{\frac{\mu}{L}}\right)^k$ $\leq LR^2 \exp\left(\frac{-k}{\sqrt{L/\mu}}\right)$

$\ln\varepsilon = \ln(R^2 L) + 0 - \frac{k}{\cancel{\sqrt{L/\mu}}}$

$$\ln\varepsilon \approx \cancel{k \ln LR^2 (1-\sqrt{\frac{\mu}{L}})} \ \ln(LR^2) + k\ln\left(1-\sqrt{\frac{\mu}{L}}\right)$$

$\approx -\sqrt{\frac{\mu}{L}}$

$\ln\varepsilon \approx \ln LR^2 - \frac{k}{\cancel{\ }} \cdot \frac{1}{\sqrt{L/\mu}}$

$\boxed{k \sim \sqrt{\frac{L}{\mu}} \ln \frac{LR^2}{\varepsilon}}$

$k \sim \sqrt{\frac{L}{\mu}} \ln\left(\frac{LR^2}{\varepsilon}\right)$

Показали оценку сверху (для верхн. границы?)

Нижняя оценка для μ сильно вып. и L-глад f

$$k \sim \Omega\left(\sqrt{\frac{L}{\mu}} \ln\left(\mu \frac{R^2}{\varepsilon}\right)\right)$$

Д-во данный результ следует из $f(x_0)-f^* \le \frac{L}{2}\|x_0-x^*\|$ а также Th1, и Л4.

Пусть $\mu > 0$. Из нижних границ сложности для μ-го класса ф-ий получим

$$f(x_k)-f^* \ge \frac{\mu}{2}\left(\frac{\sqrt{Q_f}-1}{\sqrt{Q_f}+1}\right)^{2k} R^2 \ge \frac{\mu}{2}\exp\left(-\frac{4k}{\sqrt{Q_f}-1}\right)R^2$$

$Q_f = \frac{L}{\mu}$, $R = \|x_0-x^*\|$ $\Rightarrow$ ниж. гр. для $f(x_k)-f^* \le \varepsilon$ не менее,

или $k \ge \frac{\sqrt{Q_f}-1}{4}\left[\ln\frac{1}{\varepsilon}+\ln\frac{\mu}{2}+2\ln R\right]$

Для нашей схемы имеем верх оц. $f(x_k)-f^* \le LR^2\left(1-\sqrt{\frac{\mu}{L}}\right)^k \le LR^2 \cdot \exp\left(-\frac{k}{\sqrt{Q_f}}\right)$

$\Rightarrow k \le \sqrt{Q_f}\left[\ln\frac{1}{\varepsilon}+\ln L+2\ln R\right] \Rightarrow$ основ член в оцен пропорц ниж. гр. Аналогич для $F_{0,L}^{1,1}(\mathbb{R}^n)$

ЧТД

- Схема с постоянным шагом (II) $\quad$ (☆☆☆)
  (исключим $\{\gamma_k\}$ из схемы I)

0 - $\exists\, x_0 \in \mathbb{R}^n$, $\alpha_0 \in (0,1)$. Положим $y_0 = x_0$, $q = \frac{\mu}{L}$

1 k-ая итер ($k \ge 0$)

а) вычислим $f(y_k)$ и $f'(y_k)$. Положим

$$x_{k+1} = y_k - \frac{1}{L}f'(y_k)$$

б) вычислим $\alpha_k \in (0,1)$ из ур-я

$$\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + q\alpha_{k+1}$$

и положим $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$

$$y_{k+1} = x_{k+1}+\beta_k(x_{k+1}-x_k)$$

$\Downarrow$

Теорема 3. Если в схеме (☆☆☆) выбрать $\alpha_0 \ge \sqrt{\frac{\mu}{L}}$, то

$$f(x_k)-f^* \le \min\left\{\left(1-\sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L}+k\sqrt{\gamma_0})^2}\right\} \cdot \left[f(x_0)-f^*+\frac{\gamma_0}{2}\|x_0-x^*\|^2\right],$$

где $\gamma_0 = \frac{\alpha_0(\alpha_0 L-\mu)}{1-\alpha_0}$

$\Leftarrow$ Th2 $\left((\gamma_0 \ge \mu) \Leftrightarrow \alpha_0 \ge \sqrt{\frac{\mu}{L}}\right)$ схемы.