Convex functions

November 13, 2022

Contents

1 Convex functions

Lemma 1. *For* $f \in \mathcal{F}_L^{1,1}(X)$

$$\begin{cases} \frac{t(1-t)}{2L}\left\|\nabla f(x_1)-\nabla f(x_2)\right\|^2 \leqslant (1-t)f(x_1)+tf(x_2)-f((1-t)x_1+tx_2) & \leqslant \frac{t(1-t)L}{2}\left\|x_1-x_2\right\|^2 \\ \frac{1}{2L}\left\|\nabla f(x)-\nabla f(x_0)\right\|^2 \leqslant f(x)-f(x_0)-\langle \nabla f(x_0), x-x_0\rangle & \leqslant \frac{L}{2}\left\|x-x_0\right\|^2 \\ \frac{1}{L}\left\|\nabla f(x)-\nabla f(x_0)\right\|^2 \leqslant \langle \nabla f(x)-\nabla f(x_0), x-x_0\rangle & \leqslant L\left\|x-x_0\right\|^2 \end{cases} \tag{1}$$

Proof.

$$f(x)-f(x_0)-\langle \nabla f(x_0), x-x_0\rangle = \int_0^1 \langle \nabla f(x_0+t(x-x_0))-\nabla f(x_0), x-x_0\rangle \leqslant L\int_0^1 t\,dt\left\|x-x_0\right\|^2$$

Let

$$g(y)=f(y)-\langle \nabla f(x_0), y\rangle$$

Then $g \in \mathcal{F}_L^{1,1}(X)$ and x_0 is a minimizer of g, so with $x = x_0 + \frac{1}{L}\nabla g(x)$

$$f(x_0)-f(x)-\langle \nabla f(x_0), x_0-x\rangle \leqslant g(x_0)-g(x) \leqslant \langle \nabla g(x), x_0-x\rangle + \frac{L}{2}\left\|x-x_0\right\|^2 = -\frac{L}{2}\left\|\nabla g(x)\right\|^2$$

□

Lemma 2. *For* $f \in \mathcal{S}_{\mu,L}^{1,1}(X) := \left\{ f \in \mathcal{F}_L^{1,1}(X) \;\middle|\; f(x)-f(x_0)-\langle \nabla f(x_0), x-x_0\rangle \geqslant \frac{\mu}{2}\left\|x-x_0\right\|^2 \right\}$

$$\left\{ \frac{1}{\mu+L}\left\|\nabla f(x)-\nabla f(x_0)\right\|^2 + \frac{\mu L}{\mu+L}\left\|x-x_0\right\|^2 \leqslant \langle \nabla f(x)-\nabla f(x_0), x-x_0\rangle \right. \tag{2}$$

Let $g(x) := f(x) - \frac{\mu}{2}\|x\|^2$. *Then* $\langle \nabla g(x_1) - \nabla g(x_2), x_1 - x_2\rangle = \langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle - \mu\langle x_1 - x_2, x_1 - x_2\rangle \leqslant (L-\mu)\|x_1 - x_2\|^2$, *so* $g \in \mathcal{F}^{1,1}_{L-\mu}(X)$. *Then for* $\mu < L$

$$
\begin{aligned}
&\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle - \mu\|x_1 - x_2\|^2 = \langle \nabla g(x_1) - \nabla g(x_2), x_1 - x_2\rangle \geqslant \frac{1}{L-\mu}\|\nabla g(x_1) - \nabla g(x_2)\\
&= \frac{1}{L-\mu}\left(\|\nabla f(x_1) - \nabla f(x_2)\|^2 - 2\mu\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle + \mu^2\|x_1 - x_2\|^2\right)
\end{aligned}
$$

$\Rightarrow$

$$
(L+\mu)\langle \nabla f(x_1) - \nabla f(x_2), x_1 - x_2\rangle \geqslant \|\nabla f(x_1) - \nabla f(x_2)\|^2 + (\mu^2 + (L-\mu)\mu)\|x_1 - x_2\|^2.
$$

2 FISTA smooth [BeckTeboulle09]

Let

$$
Q_\alpha(x,y) := f(y) + f'(y)(x-y) + \frac{\alpha}{2}\|x-y\|^2
$$

We have by (1)

$$
f(x) \leqslant Q_L(y,x) \quad \forall x, y \in X. \tag{3}
$$

and

$$
Q_\alpha(x,y) = f(y) - \frac{1}{2\alpha}\|\nabla f(y)\|^2 + \frac{\alpha}{2}\left\|x - y + \frac{1}{\alpha}\nabla f(y)\right\|^2 \tag{4}
$$

Let

$$
p_\alpha(y) = \operatorname{argmin}\{Q_\alpha(x,y) \mid x \in X\} \tag{5}
$$

Then

$$
p_\alpha(y) = y - \frac{1}{\alpha}\nabla f(y), \quad Q_\alpha(p_\alpha(y), y) = f(y) - \frac{1}{2\alpha}\|\nabla f(y)\|^2 = f(y) - \frac{\alpha}{2}\|p_\alpha(y) - y\|^2. \tag{6}
$$

Algorithm 1: FISTA smooth

Choose $x_0 \in X$, $y_0 = x_0$, $t_0 = 1$. Set $k = 0$.

(1) $x_{k+1} = p_L(y_k)$.

(2) $s_{k+1} = \frac{1+\sqrt{1+4s_k^2}}{2}$

(3) $y_{k+1} = x_{k+1} + \frac{s_k - 1}{s_{k+1}}(x_{k+1} - x_k)$

(4) Increment k and go to (1).

We have

$$
s_k \geqslant 1, \quad s_{k+1} \geqslant s_k, \quad s_{k+1}^2 - s_{k+1} = s_k^2, \quad s_k \geqslant (k+1)/2. \tag{7}
$$

$$
\left(\frac{s_k - 1}{s_{k+1}}\right)^2 = \frac{s_{k+1}^2 - s_{k+1} - 2s_k + 1}{s_{k+1}^2} \leqslant 1
$$

$$
\beta_k = \frac{s_k - 1}{s_{k+1}} = \frac{2s_k - 2}{1 + \sqrt{1+4s_k^2}} = \frac{(s_k - 1)(1 - \sqrt{1+4s_k^2})}{-2s_k^2}
$$

Lemma 3.

$$f(p_L(y)) - f(x) \leqslant \langle \nabla f(y), y - x\rangle - \frac{1}{2L}\|\nabla f(y)\|^2 \quad \forall x \in X. \tag{8}$$

Proof. Since by (3) Q_L is an overestimate of f, we have

$$f(p_L(y)) \leqslant Q_L(p_L(y), y)$$

Then we have, together with (6) and convexity

$$f(p_L(y)) - f(x) \leqslant Q_L(p_L(y), y) - f(y) - \langle \nabla f(y), x - y\rangle = \langle \nabla f(y), y - x\rangle - \frac{1}{2L}\|\nabla f(y)\|^2.$$

□

For $y = y_k$ and $x = x_k$, $x = x^*$ in (8), and with $\Delta f_k := f(x_k) - f(x^*)$

$$\Delta f_{k+1} - \Delta f_k \leqslant \langle \nabla f(y_k), y_k - x_k\rangle - \frac{1}{2L}\|\nabla f(y_k)\|^2$$
$$\Delta f_{k+1} \leqslant \langle \nabla f(y_k), y_k - x^*\rangle - \frac{1}{2L}\|\nabla f(y_k)\|^2$$

so multiplying the first inequality with s_k^2 and the second with $s_{k+1}^2 - s_k^2$

$$s_{k+1}^2\Delta f_{k+1} - s_k^2\Delta f_k \leqslant \langle \nabla f(y_k), s_{k+1}^2 y_k - s_k^2 x_k - (s_{k+1}^2 - s_k^2)x^*\rangle - \frac{s_{k+1}^2}{2L}\|\nabla f(y_k)\|^2$$

Now the condition

$$s_{k+1}^2 - s_k^2 = s_{k+1} \tag{9}$$

implies

$$\begin{aligned} s_{k+1}^2\Delta f_{k+1} - s_k^2\Delta f_k \leqslant& \langle s_{k+1}\nabla f(y_k), s_{k+1}y_k + (1 - s_{k+1})x_k - x^*\rangle - \frac{L}{2}\left\|\frac{s_{k+1}}{L}\nabla f(y_k)\right\|^2 \\ =& \frac{L}{2}\left(2\langle a_k, b_k\rangle - \|a_k\|^2\right) = \frac{L}{2}\left(\|b_k\|^2 - \|b_k - a_k\|^2\right) \\ a_k :=& \frac{s_{k+1}}{L}\nabla f(y_k) = s_{k+1}(y_k - x_{k+1}), \quad b_k := s_{k+1}y_k + (1 - s_{k+1})x_k - x^* \end{aligned}$$

Since

$$b_k - a_k = s_{k+1}x_{k+1} + (1 - s_{k+1})x_k - x^*$$

$$b_k - a_k = b_{k+1} \quad \Leftrightarrow \quad s_{k+1}x_{k+1} + (1 - s_{k+1})x_k = t_{k+2}y_{k+1} + (1 - t_{k+2})x_{k+1},$$

or

$$y_{k+1} = x_{k+1} + \frac{(s_{k+1} - 1)}{t_{k+2}}(x_{k+1} - x_k).$$

Now from

$$\frac{2s_{k+1}^2}{L}\Delta f_{k+1} - \frac{2s_k^2}{L}\Delta f_k \leqslant \|b_k\|^2 - \|b_{k+1}\|^2 \tag{10}$$

it follows that for any $k \geqslant 1$

$$\frac{2s_{k+1}^2}{L}\Delta f_{k+1} \leqslant \frac{2s_{k+1}^2}{L}\Delta f_{k+1} + \|b_{k+1}\|^2 \leqslant \frac{2s_k^2}{L}\Delta f_k + \|b_k\|^2 \leqslant \frac{2t_0^2}{L}\Delta f_0 + \|b_0\|^2$$

and with (7)

$$f(x_k) - f^* \leqslant \frac{1}{s_k^2}\left(f(x_0) - f^* + \frac{L}{2}\|b_0\|^2\right). \tag{11}$$

3 FISTA with backtracking ???

4 Nesterov acceleration

Algorithm 2: AGD fixed

Choose $x_0 \in X$, $\alpha = \sqrt{\kappa_f}$, $\beta = \frac{\alpha-1}{\alpha+1}$. Set $k = 0$.

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

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

(4) Increment k and go to (1).

Let us start with, for any $x \in X$,

$$\begin{cases} f(x_{k+1}) \leqslant & f(y_k) - \frac{1}{2L}\|\nabla f(y_k)\|^2 \\ f(x) \geqslant & f(y_k) + \langle \nabla f(y_k), x - y_k\rangle + \frac{\mu}{2}\|x - y_k\|^2 \end{cases}$$
$$\Rightarrow \quad f(x_{k+1}) - f(x) \leqslant \langle \nabla f(y_k), y_k - x\rangle - \frac{1}{2L}\|\nabla f(y_k)\|^2 - \frac{\mu}{2}\|x - y_k\|^2$$

Setting $\Delta f_k := f(x_k) - f^*$ we then have with $0 < \theta < 1$ and using $2ab - a^2 = b^2 - (a-b)^2$

$$\begin{aligned}\Delta f_{k+1} - (1-\theta)\Delta f_k \leqslant & \langle \nabla f(y_k), y_k - (1-\theta)x_k - \theta x^*\rangle - \frac{1}{2L}\|\nabla f(y_k)\|^2 - \frac{\theta\mu}{2}\|x^* - y_k\|^2 \\ = & \frac{L}{2}\left(\|y_k - (1-\theta)x_k - \theta x^*\|^2 - \|x_{k+1} - (1-\theta)x_k - \theta x^*\|^2\right) - \frac{\theta\mu}{2}\|x^* - y_k\|^2\end{aligned}$$

With $\theta = \frac{1-\beta}{1+\beta}$, such that $\beta = \frac{1-\theta}{1+\theta}$ and $\frac{1-\theta}{\theta\beta} - 1 = \frac{1+\theta}{\theta} - 1 = \frac{1}{\theta}$

$$\begin{aligned} z_k := & \frac{1}{\theta}x_k - \frac{1-\theta}{\theta}x_{k-1} = x_k + \frac{1-\theta}{\theta}(x_k - x_{k-1}) = y_k + \left(\frac{1-\theta}{\theta} - \beta\right)(x_k - x_{k-1}) \\ = & y_k + \left(\frac{1-\theta}{\beta\theta} - 1\right)(y_k - x_k) = y_k + \frac{1}{\theta}(y_k - x_k)\end{aligned}$$

and

$$\frac{1}{\theta}y_k - \frac{1-\theta}{\theta}x_k = x_k + \frac{1}{\theta}(y_k - x_k) = z_k + x_k - y_k$$

Putting these together, we find

$$\Delta f_{k+1} - (1-\theta)\Delta f_k \leqslant \frac{L}{2}\theta^2\left(\|z_k - x^* + x_k - y_k\|^2 - \|z_{k+1} - x^*\|^2\right) - \frac{\theta\mu}{2}\|x^* - y_k\|^2$$

Now we have

$$\begin{aligned} 2\langle z_k - x^*, x_k - y_k\rangle = & 2\theta\langle z_k - x^*, y_k - z_k\rangle = \theta\left(\|y_k - x^*\|^2 - \|z_k - x^*\|^2 - \|y_k - z_k\|^2\right) \\ = & \theta\left(\|y_k - x^*\|^2 - \|z_k - x^*\|^2 - \frac{1}{\theta^2}\|y_k - x_k\|^2\right)\end{aligned}$$

such that

$$
\begin{aligned}
\|z_k - x^* + x_k - y_k\|^2 &= \|z_k - x^*\|^2 + 2\langle z_k - x^*, x_k - y_k\rangle + \|x_k - y_k\|^2 \\
&= \|z_k - x^*\|^2 - \theta\left(\|z_k - x^*\|^2 - \|y_k - x^*\|^2\right) - \frac{1}{\theta}\|y_k - x_k\|^2 + \|x_k - y_k\|^2 \\
&=(1-\theta)\|z_k - x^*\|^2 + \theta\|y_k - x^*\|^2 - \frac{1-\theta}{\theta}\|y_k - x_k\|^2
\end{aligned}
$$

and we get

$$
\begin{aligned}
\theta\left(\Delta f_{k+1} + \frac{L}{2}\theta^2\|z_{k+1} - x^*\|^2\right) \leqslant &(1-\theta)\left(\Delta f_k - \Delta f_{k+1}\right) + \frac{L}{2}\theta^2(1-\theta)\left(\|z_k - x^*\|^2 - \|z_{k+1} - x^*\|^2\right) \\
&+ \frac{L}{2}\theta^3\|y_k - x^*\|^2 - 2L\theta^2\frac{1-\theta}{\theta}\|y_k - x_k\|^2 - \frac{\theta\mu}{2}\|x^* - y_k\|^2
\end{aligned}
$$

For

$$
\frac{L}{2}\theta^2 \leqslant \frac{\mu}{2} \quad \Leftrightarrow \quad \theta \leqslant \frac{1}{\sqrt{\kappa}}
$$

we have with $a_k := \Delta f_k + 2L\theta^2\|z_k - x^*\|^2$

$$
\sum_{k=n+1}^{\infty} a_k \leqslant \frac{1-\theta}{\theta} a_n,
$$

so with $S_n := \sum_{k=n}^{\infty} a_k$

$$
S_{n+1} \leqslant \frac{1-\theta}{\theta}\left(S_n - S_{n+1}\right) \quad \Rightarrow \quad S_{n+1} \leqslant (1-\theta)S_n \quad \Rightarrow \quad S_{n+k} \leqslant (1-\theta)^k S_n
$$

or

$$
a_{n+k} \leqslant S_{n+k} \leqslant (1-\theta)^{k-1} S_{n+1} \leqslant \frac{(1-\theta)^k}{\theta} a_n \tag{12}
$$