Analysis of Nesterov acceleration

Sunday, March 21, 2021 8:49 AM

$x_k = y_{k-1} - \frac{1}{L}\nabla f(y_{k-1})$

$y_k = x_k + \frac{k}{k+3}(x_k - x_{k-1})$

Nesterov 1983, this analysis via course notes of Lieven Vandenberghe

For analysis, $\theta_0 = 1$, $\theta_k = \frac{2}{k+3}$ and $v_0 = x_0$, $v_k = \frac{k+2}{2}x_k - \frac{k}{2}x_{k-1}$

$y_k = \theta_k v_k + (1-\theta_k)x_k$ is a convex combination, $v_{k+1} = v_k + \frac{1}{\theta_k}(x_{k+1} - y_k)$

Analyze 1 step, so let $x = x_{k-1}$, $x^+ = x_k$, $y = y_{k-1}$, $v = v_{k-1}$, $v^+ = v_k$, $\theta = \theta_{k-1}$

$f(x^+) \le f(y) - \frac{1}{2L}\|\nabla f(y)\|^2$ (descent lemma, same analysis as grad. descent)

$$\le \overbrace{f((1-\theta)x + \theta x^*) + \langle \nabla f(y), y - (1-\theta)x - \theta x^* \rangle}^{\text{via convexity: above tangent line}} - \frac{1}{2L}\|\nabla f(y)\|^2$$

$$\le \overbrace{(1-\theta)f(x) + \theta f(x^*)}^{\text{via convexity}} + \langle \nabla f(y), y - (1-\theta)x - \theta x^* \rangle - \frac{1}{2L}\|\nabla f(y)\|^2$$

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

complete-the-square

$$= (1-\theta)f(x) + \theta f(x^*) + \frac{\theta^2 \cdot L}{2}\Big(\|v - x^*\|^2 - \underbrace{\|v - x^* - \tfrac{1}{\theta\cdot L}\nabla f(y)\|^2}_{\|v^+ - x^*\|^2}\Big)$$

putting back in subscripts...

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

divide by θ_{k-1}^2 and re-arrange:

$$\frac{1}{\theta_{k-1}^2}\left(f(x_k) - f^*\right) + \frac{L}{2}\|v_k - x^*\|^2 \le \underbrace{\frac{1-\theta_{k-1}}{\theta_{k-1}^2}}\left(f(x_{k-1}) - f^*\right) + \frac{L}{2}\|v_{k-1} - x^*\|^2$$

$\theta_k = \frac{2}{k+3}$

$\frac{1-\theta_{k-1}}{\theta_{k-1}^2} = \frac{k(k+2)}{4} \le \frac{(k+1)^2}{4} = \frac{1}{\theta_{k-2}^2}$

So

$$\frac{1}{\theta_{k-1}^2}\left(f(x_k) - f^*\right) + \frac{L}{2}\|v_k - x^*\|^2 \le \frac{1}{\theta_{k-2}^2}\left(f(x_{k-1}) - f^*\right) + \frac{L}{2}\|v_{k-1} - x^*\|^2$$

So applying this recursively ...

$$\le \overset{\theta_0 = 1}{\cancel{\frac{1-\theta_0}{\theta_0^{2}}\left(f(x_0) - f^{*} \right)}} + \frac{L}{2} \| \overset{= x_0}{v_0} - x^{*} \|^2$$

i.e., $f(x_k - f^{*}) \le \frac{L}{2}\, \theta_{k-1}^{2}\, \| x_0 - x^{*} \|^2$

$$= \frac{L}{2}\, \frac{4}{(k+2)^2}\, \| x_0 - x^{*} \|^2 = O\left(\frac{1}{k^2} \right)$$