Lower Complexity Bounds (Nesterov Sec 2.1.4)

Assume as before that f is strongly convex & C^2 with
$$mI \preceq \nabla^2 f(x) \preceq MI \quad \text{for all } x \in S.$$
m: μ in Nesterov; M: L in Nesterov.

Assume that at each point $x^{(k)}$, a "first-order oracle" or "black box" computes $f(x^{(k)})$ and $\nabla f(x^{(k)})$.
Assume also that for $k = 1, 2, \ldots$

(ASSUMPTION A) $\quad x_k \in \text{Linear Span}\{\nabla f(x_0), \ldots, \nabla f(x_{k-1})\}$.

For simplicity, assume $\operatorname{dom} f = \mathbb{R}^\infty \equiv l_2 =$
$$\left\{ x = (x_i)_{i=1}^\infty : \|x\|^2 = \sum_{i=1}^\infty x_i^2 < \infty \right\}.$$

Now we define a "difficult" function F by
$$F(x) = \frac{M-m}{8}\left\{ (x_1)^2 + \sum_{i=1}^\infty (x_i - x_{i+1})^2 - 2x_1 \right\} + \frac{m}{2}\|x\|^2.$$

We have
$$\frac{\partial F}{\partial x_1} = \frac{M-m}{8}\left(2x_1 + 2(x_1 - x_2) - 2\right) + m x_1$$

$j > 1$:
$$\frac{\partial F}{\partial x_j} = \frac{M-m}{8}\left(2(x_j - x_{j+1}) - 2(x_{j-1} - x_j)\right) + m x_j$$
$$= \frac{M-m}{8}\left(4x_j - 2x_{j+1} - 2x_{j-1}\right) + m x_j$$

LCB2

so $$\nabla^2 F(x) = \frac{M-m}{4}\underbrace{\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & \ddots \\ & & \ddots & \end{bmatrix}}_{\le 4I,\ \ge 0}\ T + mI$$

with $\le 4I$, ≥ 0 (because of diagonal dominance)

$$\left.\begin{aligned} \lambda_{max}(\nabla^2 F(x)) &\le M \\ \lambda_{min}(\nabla^2 F(x)) &\ge m \end{aligned}\right\} \text{ as required.}$$

and

$$\nabla F(x) = \left(\frac{M-m}{4} T + mI\right)x - \frac{M-m}{4} e_1 \qquad e_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$$

The solution x^* is given by $\nabla F(x^*) = 0$:

$$\frac{M-m}{4}(2x_1 - x_2) + m x_1 = \frac{M-m}{4}$$

$$\frac{M+m}{2} x_1 - \frac{M-m}{4} x_2 = \frac{M-m}{4}$$

$$x_2 - 2\frac{M+m}{M-m} x_1 + 1 = 0$$

and, for $j = 2, 3, \ldots$

$$x_{j+1} - 2\frac{M+m}{M-m} x_j + x_{j-1} = 0$$

This difference equation can be solved by plugging in $x_j = q^j$ and solving for q:

LCB3

$$q^{j+1} - 2\frac{M+m}{M-m} q^{j} + q^{j-1} = 0$$

$$q^2 - 2\frac{M+m}{M-m} q + 1 = 0$$

Claim: roots are $\dfrac{M+m \pm 2\sqrt{M}\sqrt{m}}{M-m}$.

check: sum of roots is then $2\dfrac{M+m}{M-m}$ ✓

product of roots is $\dfrac{(M+m)^2 - 4Mm}{(M-m)^2} = 1$ ✓

<u>Smaller</u> root is

$$q = \frac{M+m-2\sqrt{M}\sqrt{m}}{M-m} = \frac{(\sqrt{M}-\sqrt{m})^2}{(\sqrt{M}-\sqrt{m})(\sqrt{M}+\sqrt{m})}$$

$$= \frac{\sqrt{M}-\sqrt{m}}{\sqrt{M}+\sqrt{m}}$$

$$= \frac{1-\sqrt{1/K}}{1+\sqrt{1/K}} \qquad \text{where } K = \frac{M}{m}$$

NOTE THE SQRT.

LCB4

Nesterov Thm 2.1.13.

<u>We get</u> <u>Theorem</u> For any $x^{(0)} \in \mathbb{R}^\infty$ and any $m > 0$, $M > m$, $\exists$ function F (quadratic) with $mI \le \nabla^2 F \le MI$ with (under ASSUMPTION A)

$$\|x^{(k)} - x^*\|^2 \geq \left(\frac{1-\sqrt{1/K}}{1+\sqrt{1/K}}\right)^{2k} \|x^{(0)} - x^*\|^2$$

where x^* minimizes F and $K = M/m$, and hence

$$F(x^{(k)}) - F^* \geq \frac{m}{2}(\text{ same })^{2k} \|x^{(0)} - x^*\|^2$$

<u>Pf</u> WLOG take $x^{(0)} = 0$. Then we use F as defined already, getting:

$$\|x^{(0)} - x^*\|^2 = \sum_{i=1}^{\infty} (x^*)_i^2 = \sum_{i=1}^{\infty} q^{2i} = \frac{q^2}{1-q^2}$$

Since T is tridiagonal and $\nabla F(x^{(0)}) = e_1$, we can show by induction that $\nabla F(x^{(j-1)}) \in \text{span}(e_1, \dots, e_j)$ so $x^{(j)} \in \text{span}(e_1, \dots, e_j)$

so $$\|x^{(j)} - x^*\|^2 \geq \sum_{i=j+1}^{\infty} (x^*)_i^2 = \sum_{i=j+1}^{\infty} q^{2i}$$

$$= \frac{q^{2(j+1)}}{1-q^2} = q^{2j} \|x^{(0)} - x^*\|^2$$

with $q = \frac{1-\sqrt{1/K}}{1+\sqrt{1/K}}$. (Last inequality follows from our original (1), just Taylor's Thm.) (BV (9.8))

LCB4½

Putting this in terms of function values only:

Using (3) in gradient notes, with $x = x^*$, $y = x^{(0)}$,

$$\|x^{(0)} - x^*\|^2 \geq \frac{2}{M}\left(F(x^{(0)}) - F^*\right)$$

so lower bound becomes

$$F(x^{(k)}) - F^* \geq \underbrace{\frac{m}{M}}_{1/K}\left(\frac{1-\sqrt{1/K}}{1+\sqrt{1/K}}\right)^{2k}\left(F(x^{(0)}) - F^*\right).$$

Compare to Nesterov complexity of gradient method for $t = \frac{2}{m+M}$

$$f(x^{(k)}) - p^* \leq K\left(\frac{1-1/K}{1+1/K}\right)^{2k}\left(f(x^{(0)}) - f^*\right)$$

KEY POINT: NO SQUARE ROOTS.

or — continued on next page.

In comparison, the gradient method with $t = \frac{1}{M}$ gave us

$$f(x^{(k)}) - f^* \leq \left(1 - \frac{1}{K}\right)^k \left(f(x^{(0)}) - f^*\right)$$

Compare $\dfrac{1-\sqrt{1/K}}{1+\sqrt{1/K}} \approx \left(1 - \sqrt{1/K}\right)^2 \approx 0.998$

if $K = 10^6$, while $1 - \frac{1}{K} = 0.999999$.

So lower bound indicates we may be able to do much better. The rest of Nesterov's Chapter 2 derives "optimal gradient" method, but the argument is very complicated! In the end, the simplest is:

NESTEROV OPTIMAL GRADIENT ALGORITHM (p. 81)

Choose $y^{(0)} = x^{(0)} \in \mathbb{R}^n$

For $k = 0, 1, 2, \ldots$

$$\left[\begin{array}{l} \text{set } x_{k+1} = y_k - \frac{1}{M} \nabla f(y_k) \\ \text{set } y_{k+1} = x_{k+1} + q\,(x_{k+1} - x_k) \end{array}\right.$$

where $q = (1 - \sqrt{1/K})/(1 + \sqrt{1/K})$