

ORFE 522: Linear and Nonlinear  
Optimization  
Homework 4

Zachary Hervieux-Moore

Wednesday 7<sup>th</sup> December, 2016

**Exercise 1:** Consider the following function:

$$f(x_1, x_2) = \log(e^{x_1} + e^{x_2})$$

- 1) Show that  $f$  is a convex function
- 2) Convert the following optimization problem into a convex program (hint: use result in part (1)):

$$\begin{aligned} \min \quad & x/y \\ \text{s.t.} \quad & 2 \leq x \leq 3 \\ & x^2 + y/z \leq \sqrt{y} \\ & x/y = z^2 \\ & x, y, z \geq 0 \end{aligned}$$

**Answer:**

- 1) We first rewrite the function as

$$f(x_1, x_2) = \log(1 + e^{x_2 - x_1}) + x_1$$

Now note that  $f(x_1, x_2) = g(x_2 - x_1) + x_1$  where  $g(z) = \log(1 + e^z)$ . We have that  $g(z)$  is convex and non-decreasing since

$$\begin{aligned} g'(z) &= \frac{e^z}{1 + e^z} > 0 \\ g''(z) &= \frac{e^z}{(1 + e^z)^2} > 0 \end{aligned}$$

Now,  $g(x_2 - x_1)$  is also convex since  $x_2 - x_1$  is linear and so also convex and using the fact that composing a convex non-decreasing function with a convex function is convex. Finally,  $x_1$  is linear (and so convex) and adding convex functions is convex. So,  $f(x_1, x_2)$  is convex.

- 2) Clearly, the objective, the second constraint, and the third constraint are not convex. We begin with the second constraint since it follows the part 1) for convexity. Notice that we can rewrite it as

$$\begin{aligned} e^{2 \log x} + e^{\log y - \log z} &\leq e^{1/2 \log y} \text{ or} \\ \log(e^{2 \log x} + e^{\log y - \log z}) &\leq 1/2 \log y \end{aligned}$$

Thus, letting  $x' = 2 \log x$ ,  $y' = \log y - \log z$ , and  $z' = 1/2 \log y$  we get

$$\log(e^{x'} + e^{y'}) - z' \leq 0$$

Which is convex since it is a linear combination of convex functions. Using these new variables, the problem becomes,

$$\begin{aligned} \min \quad & e^{\frac{1}{2}x' - 2z'} \\ \text{s.t.} \quad & \frac{1}{2} \log 2 \leq x' \leq \frac{1}{2} \log 3 \\ & \log(e^{x'} + e^{y'}) - z' \leq 0 \\ & \frac{1}{2}x' - 2z' = -2y' + 4z' \\ & x', y', z' \in \mathbb{R} \end{aligned}$$

Notice that the objective is the composition of an increasing convex function with a convex function, thus it is convex. The first constraint is convex since it is linear. The second was shown to be convex (by part 1). The last equality constraint is convex since it is affine. Finally, the variables are no longer constrained to be positive since the range of  $\log$  is all real numbers. Thus, the new constraint set is

$$(x', y', z') \in [\frac{1}{2} \log 2, \frac{1}{2} \log 3] \times \mathbb{R} \times \mathbb{R}$$

which is convex.

**Exercise 2:** Let  $f = \max_{i=1,\dots,m} f_i(x)$  where each  $f_i$  is convex and everywhere differentiable. Prove that

$$f(y) \geq f(x) + \nabla f_j(x)^T(y - x), \quad \forall x, y \in \mathbb{R}^n, j \in \{i : f_i(x) = f(x)\}$$

**Answer:** Since all the  $f_i$  are convex, we have the following inequality

$$f_i(y) \geq f_i(x) + \nabla f_i(x)^T(y - x), \quad \forall x, y \in \mathbb{R}^n, i \in \{1, \dots, m\}$$

Note that  $f(y) \geq f_i(y)$ . Now, let  $j \in \{i : f_i(x) = f(x)\}$ . Thus, we have  $f_j(x) = f(x)$  for all  $j$ . Thus, the above becomes

$$f(y) \geq f_j(y) \geq f(x) + \nabla f_j(x)^T(y - x), \quad \forall x, y \in \mathbb{R}^n, j \in \{i : f_i(x) = f(x)\}$$

**Exercise 3:** Let  $f$  be twice differentiable and convex, with an optimal solution  $\mathbf{x}^*$  over a convex and closed set  $X$ . Consider the gradient projection mapping defined by

$$T(x) = [x^k - \alpha \nabla f(x^k)]^+$$

Where  $[\cdot]^+$  denotes the Euclidean projection onto  $X$ ,  $\alpha > 0$ . Prove that  $x^*$  is a global optimal solution if and only if  $x^*$  is a fixed point of the mapping  $T$ .

**Answer:** We first begin with the assumption that  $x^*$  is a fixed point of the mapping  $T$ . That is,

$$T(x^*) = x^* = [x^* - \alpha \nabla f(x^*)]^+$$

Or, equivalently, that  $x^*$  is a projection of  $x^* - \alpha \nabla f(x^*)$ . It is a well known result that a characterization of a projection is

$$\langle x^* - (x^* - \alpha \nabla f(x^*)), z - x^* \rangle \geq 0 \quad \forall z \in X$$

Now using a series of equivalences

$$\begin{aligned} & \langle x^* - (x^* - \alpha \nabla f(x^*)), z - x^* \rangle \geq 0 \quad \forall z \in X \\ \iff & \langle \alpha \nabla f(x^*), z - x^* \rangle \geq 0 \quad \forall z \in X \\ \iff & \langle \nabla f(x^*), z - x^* \rangle \geq 0 \quad \forall z \in X \\ \iff & x^* \text{ is optimal} \end{aligned}$$

Where the last equivalence follows from the fact that both  $f$  and  $X$  are convex.

**Exercise 4:** Let  $f$  be twice differentiable and convex, with an optimal solution  $x^*$  over a convex and closed set  $X$ . Consider the rescaled gradient projection method

$$x^{k+1} = [x^k - \alpha D \nabla f(x^k)]^+$$

where  $[\cdot]^+$  denotes the Euclidean projection onto  $X$ , and  $D \succ 0$  is a symmetric matrix. Assume that  $\nabla^2 f \preceq H$  for all  $x$ , where  $H$  is a symmetric and positive definite matrix.

- 1) Analyze the convergence rate of the rescaled gradient method, i.e., give a bound for  $f(x^k) - f(x^*)$  using  $\alpha$ ,  $D$ , and  $H$ . (Prove rigorously, and try to get the sharpest possible bound).
- 2) What would be the optimal choice of step size  $\alpha$ ? What would be the bound then?
- 3) Assume in addition that  $f$  is strongly convex such that  $\nabla^2 f(x) \succeq \sigma I$  for some  $\sigma > 0$ . Show that there exists some  $\rho > 0$  such that

$$\|x^k - x^*\| \leq \rho^k \|x^0 - x^*\|$$

Find  $\rho$  as a function of  $H$ ,  $D$ ,  $\sigma$ ,  $\alpha$ .

- 4) Under the assumptions of part 3), what would be the optimal choice of step size  $\alpha$ ? What would be the error bound then?

**Answer:** ,

- 1) The first thing we shall do is to define the projection

$$\begin{aligned} x^{k+1} &= [x^k - \alpha D \nabla f(x^k)]_{D^{-1}}^+ \\ x^{k+1} &= \arg \min_{x \in X} \|x^k - \alpha D \nabla f(x^k) - x\|_{D^{-1}}^2 \end{aligned}$$

Where the norm is defined to be

$$\|x\|_{D^{-1}}^2 = \langle x, x \rangle_{D^{-1}} = x^T D^{-1} x$$

Using the characterization of projections, this new projection yields

$$\begin{aligned} \langle x - x^{k+1}, x^k - \alpha D \nabla f(x^k) - x^{k+1} \rangle_{D^{-1}} &\leq 0 \quad \forall x \in X \\ \iff (x - x^{k+1})^T D^{-1} (x^k - \alpha D \nabla f(x^k) - x^{k+1}) &\leq 0 \quad \forall x \in X \end{aligned}$$

Since  $D$  is positive definite and symmetric, we can define  $D = SS$  and  $D^{-1} = S^{-1}S^{-1}$ . We now also define the following using this decomposition.  $x^k = Sy^k$ ,  $x = Sy$ ,  $g(y) = f(Sy)$ ,  $\nabla g(y) = Sf(x)$ . Where  $g$  is convex and in  $C^2$  because  $f$  is just composed with a linear function. Thus, the above becomes

$$\begin{aligned} &\iff (Sy - Sy^{k+1})^T D^{-1}(Sy^k - \alpha SS \nabla f(x^k) - Sy^{k+1}) \leq 0 \quad \forall y \in S^{-1}(X) \\ &\iff (y - y^{k+1})^T SD^{-1}S(y^k - \alpha \nabla g(y^k) - y^{k+1}) \leq 0 \quad \forall y \in S^{-1}(X) \\ &\iff (y - y^{k+1})^T (y^k - \alpha \nabla g(y^k) - y^{k+1}) \leq 0 \quad \forall y \in S^{-1}(X) \\ &\iff y^{k+1} = [y^k - \alpha \nabla g(y^k)]^+ \end{aligned}$$

Where  $[\cdot]^+$  is the Euclidean projection. Notice that this is the original problem, but now  $D = I$ . Now, notice that  $\nabla^2 g = S \nabla^2 f S \succeq SHS$ . So,  $\|\nabla^2 g\| \leq \|S\|^2 \|H\|$ . Where  $\|S\|^2 = \lambda_{\max}(D)$  and  $\|H\|^2 = \lambda_{\max}(H)$  since both matrices are diagonalizable. We let  $L = \lambda_{\max}(D)\lambda_{\max}(H)$  and so  $g(x)$  is  $L$ -Lipschitz. Thus, we use the following inequality

$$g(y^{k+1}) \leq g(y^k) + \nabla g(y^k)^T (y^{k+1} - y^k) + \frac{L}{2} \|y^{k+1} - y^k\|^2 \quad (1)$$

Again, from the definition of a projection

$$(y - y^{k+1})^T (y^k - \alpha \nabla g(y^k) - y^{k+1}) \leq 0$$

Which gives us the bound

$$\nabla g(y^k)(y^{k+1} - y^k) \leq -\frac{1}{\alpha} \|y^{k+1} - y^k\|^2 \quad (2)$$

Bounding (1) with (2) yields

$$g(y^{k+1}) \leq g(y^k) + \left(\frac{L}{2} - \frac{1}{\alpha}\right) \|y^{k+1} - y^k\|^2$$

Which is decreasing for  $\alpha < \frac{2}{L}$ . We pick  $\alpha \leq \frac{1}{L}$  for convenience. Hence, we now have a contraction. By convexity of  $g$  we have

$$g(y^k) \leq g(y^*) + \nabla g(y^k)^T (y^k - y^*)$$

Which, when applied to the upper bound of (1) yields

$$\begin{aligned}
g(y^{k+1}) &\leq g(y^*) + \nabla g(y^k)^T(y^{k+1} - y^k) + \nabla g(y^k)^T(y^k - y^*) \\
&\quad + \frac{1}{2\alpha} \|y^{k+1} - y^k\|^2 \\
g(y^{k+1}) &\leq g(y^*) + \nabla g(y^k)^T(y^{k+1} - y^*) + \frac{1}{2\alpha} \|y^{k+1} - y^k\|^2 \quad (3)
\end{aligned}$$

We now wish to upper bound these terms. Again, applying the characterization of projection of  $y^k$  we have

$$\begin{aligned}
(y^* - y^{k+1})^T(y^k - \alpha \nabla g(y^k) - y^{k+1}) &\leq 0 \\
(y^* - y^{k+1})^T(-\alpha \nabla g(y^k)) &\leq -(y^* - y^{k+1})^T(y^k - y^{k+1}) \\
(y^{k+1} - y^*)^T(\nabla g(y^k)) &\leq -\frac{1}{\alpha}(y^* - y^{k+1})^T(y^k - y^{k+1})
\end{aligned}$$

Applying this to (3) yields

$$\begin{aligned}
g(y^{k+1}) &\leq g(y^*) - \frac{1}{\alpha}(y^* - y^{k+1})^T(y^k - y^{k+1}) + \frac{1}{2\alpha} \|y^{k+1} - y^k\|^2 \\
g(y^{k+1}) &\leq g(y^*) + \frac{1}{2\alpha} [-2(y^* - y^{k+1})^T(y^k - y^{k+1}) + \|y^{k+1} - y^k\|^2] \\
g(y^{k+1}) &\leq g(y^*) + \frac{1}{2\alpha} [-2(y^* - y^{k+1})^T(y^k - y^*) \\
&\quad -2(y^* - y^{k+1})^T(y^* - y^{k+1}) + \|y^{k+1} - y^*\|^2 \\
&\quad + \|y^k - y^*\|^2 + 2(y^* - y^{k+1})^T(y^k - y^*)] \\
g(y^{k+1}) &\leq g(y^*) + \frac{1}{2\alpha} [\|y^k - y^*\|^2 - \|y^{k+1} - y^*\|^2]
\end{aligned}$$

Which, is a telescoping sum. Thus

$$\frac{1}{K} \sum_{k=0}^K g(y^k) - g(y^*) \leq \frac{1}{2\alpha K} \|y^0 - y^*\|^2$$

But, since  $g(y^k)$  is decreasing, we have  $g(y^K) \leq \frac{1}{K} \sum_{k=0}^K g(y^k)$ . This yields

$$0 \leq g(y^K) - g(y^*) \leq \frac{1}{2\alpha K} \|y^0 - y^*\|^2$$



Which is equivalent to

$$0 \leq f(x^K) - f(x^*) \leq \frac{1}{2\alpha K} \|x^0 - x^*\|_{D^{-1}}^2$$

- 2) We wish to minimize the upper bound by picking  $\alpha$  suitably. We notice that minimizing the bound is equivalent to maximizing  $\alpha$ . Since we bounded  $\alpha$  by  $\frac{1}{L}$  to get the answer in part 1), we pick  $\alpha = \frac{1}{L}$ . The bound then becomes

$$0 \leq f(x^K) - f(x^*) \leq \frac{L}{2K} \|x^0 - x^*\|_{D^{-1}}^2$$

- 3) We have that  $y^* = [y^* - \alpha \nabla g(y^*)]^+$  is a stationary point, similar to the proof in question 3). Then we have

$$\begin{aligned} \|x^{k+1} - x^*\|_{D^{-1}}^2 &= \|y^{k+1} - y^*\|^2 \\ &\leq \|y^k - \alpha \nabla g(y^k) - y^* + \alpha \nabla g(y^*)\|^2 \\ &\leq \|y^k - y^*\|^2 + 2\alpha(y^k - y^*)^T(\nabla g(y^*) - \nabla g(y^k)) + \alpha^2 \|\nabla g(y^*) - \nabla g(y^k)\|^2 \\ &\leq (1 - 2\alpha\sigma\lambda_{\min}(D) + L^2\alpha^2) \|y^* - y^k\|^2 \end{aligned}$$

Where the last line is a result of the following two inequalities. By strong convexity, we have

$$\begin{aligned} \nabla^2 g &= S \nabla^2 f S \succeq \sigma D \succeq \sigma \lambda_{\min}(D) I \\ \implies (y^k - y^*)^T (\nabla g(y^*) - \nabla g(y^k)) &\leq -\sigma \lambda_{\min}(D) \|y^k - y^*\|^2 \end{aligned}$$

By  $L$ -Lipschitz

$$\|\nabla g(y^*) - \nabla g(y^k)\|^2 \leq L \|y^k - y^*\|^2$$

Where  $L = \lambda_{\max}(D)\lambda_{\max}(H)$  as before. Applying the bound recursively yields,

$$\|x^k - x^*\|_{D^{-1}}^2 \leq (1 - 2\alpha\sigma\lambda_{\min}(D) + L^2\alpha^2)^k \|x^0 - x^*\|_{D^{-1}}^2$$

Thus, letting  $\rho = (1 - 2\alpha\sigma\lambda_{\min}(D) + L^2\alpha^2)^{1/2}$  gives us

$$\|x^k - x^*\|_{D^{-1}} \leq \rho^k \|x^0 - x^*\|_{D^{-1}}$$

4) Minimizing  $\rho$  by taking the derivative and setting it to 0 gives

$$\alpha = \frac{\sigma \lambda_{\min}(D)}{L^2}$$

Which means that  $\rho$  becomes

$$\begin{aligned}\rho &= \left(1 - \frac{\sigma^2 \lambda_{\min}^2(D)}{L^2}\right)^{1/2} \\ \rho &= \left(1 - \frac{\sigma^2 \lambda_{\min}^2(D)}{\lambda_{\max}^2(D) \lambda_{\max}^2(H)}\right)^{1/2} \geq 0\end{aligned}$$