

Solutions for Homework 1

1.

(a) Define $g(t) := f(y + t(x - y))$. Then we have $\int_0^1 g'(t)dt = g(1) - g(0) = f(x) - f(y)$. Notice $g'(t) = \nabla f(y + t(x - y))^\top (x - y)$. Therefore, we have

$$\begin{aligned} f(x) - f(y) - \nabla f(y)^\top (x - y) &= \int_0^1 \nabla f(y + t(x - y))^\top (x - y)dt - \nabla f(y)^\top (x - y)dt \\ &= \int_0^1 (\nabla f(y + t(x - y)) - \nabla f(y))^\top (x - y)dt \\ &\leq \int_0^1 \|\nabla f(y + t(x - y)) - \nabla f(y)\| \cdot \|x - y\|dt \\ &\leq \int_0^1 L\|x - y\| \cdot \|x - y\|tdt = \frac{L}{2}\|x - y\|^2 \end{aligned}$$

In the last two steps, we used Cauchy-Schwartz inequality and the definition of L -smoothness. This completes the proof.

(b) A differentiable function g is convex if $g(x) \geq g(y) + \nabla g(y)^\top (x - y)$ for all x, y . In order to prove this part, we will use the following fact.

Fact 1. If g is convex and L -smooth, then the following inequality holds for all x, y :

$$(\nabla g(y) - \nabla g(x))^\top (y - x) \geq \frac{1}{L}\|\nabla g(y) - \nabla g(x)\|^2.$$

We first prove the above fact. At first, we define $h(x) = g(x) - x^\top \nabla g(y)$ and show $h(x)$ is L -smooth and $h(y) \leq h(x)$ for any x . Notice $\nabla h(x) = \nabla g(x) - \nabla g(y)$ and hence $\nabla h(y) = 0$. In addition, $h(x) - h(z) - \nabla h(z)^\top (x - z) = g(x) - g(z) - (x - z)^\top \nabla g(y) - (\nabla g(z) - \nabla g(y))^\top (x - z) = g(x) - g(z) - \nabla g(z)^\top (x - z)$ for all x, z . Based on this relation, g is convex and L -smooth if and only if h is also convex and L -smooth. The convexity of h implies $h(x) \geq h(z) + \nabla h(z)^\top (x - z)$. Substituting $z = y$ and $\nabla h(y) = 0$ leads to $h(x) \geq h(y)$ for all x . Therefore, we have

$$h(y) \leq h\left(x - \frac{1}{L}\nabla h(x)\right) \leq h(x) + \nabla h(x)^\top \left(-\frac{1}{L}\nabla h(x)\right) + \frac{L}{2}\|-\nabla h(x)/L\|^2 = h(x) - \frac{1}{2L}\|\nabla h(x)\|^2$$

In the above inequality, the first step follows from $h(y) \leq h(x)$ for all x . The second step follows from the L -smoothness of h . The third step is basic algebra. Since $h(x) = g(x) - x^\top \nabla g(y)$ and $\nabla h(x) = \nabla g(x) - \nabla g(y)$, the above inequality is equivalent to

$$g(y) - y^\top \nabla g(y) \leq g(x) - x^\top \nabla g(y) - \frac{1}{2L}\|\nabla g(x) - \nabla g(y)\|^2$$

Since the choice of y is arbitrary, the above inequality just holds for any x, y . We can exchange (x, y) in the above inequality and obtain

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

Now Fact 1 can be directly proved via summing the above two inequalities.

Now we have proved Fact 1. Define $g(x) := f(x) - \frac{m}{2} \|x\|^2$. We have $\nabla g(x) = \nabla f(x) - mx$. For any x, y , we have $g(x) - g(y) - \nabla g(y)^T(x - y) = f(x) - f(y) - \frac{m}{2} \|x\|^2 + \frac{m}{2} \|y\|^2 - \nabla f(y)^T(x - y) + mx^T y - m\|y\|^2 = f(x) - f(y) - \nabla f(y)^T(x - y) - \frac{m}{2} \|x - y\|^2$. Notice f is m -strongly convex and it is straightforward to verify $g(x) \geq g(y) + \nabla g(y)^T(x - y)$. Hence g is convex. Since f is L -smooth, the following inequality also holds

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

Therefore g is convex and $(L - m)$ -smooth. Now we can use Fact 1 to get

$$(\nabla g(y) - \nabla g(x))^T(y - x) \geq \frac{1}{L - m} \|\nabla g(y) - \nabla g(x)\|^2 \quad (2)$$

which is equivalent to

$$(\nabla f(y) - my - \nabla f(x) + mx)^T(y - x) \geq \frac{1}{L - m} \|\nabla f(y) - my - \nabla f(x) + mx\|^2$$

Manipulating the above inequality directly leads to the desired inequality we want to show. This completes the proof.

(c) There are many different ways to prove this fact. We present one here. Other proofs will also get the same credits.

If $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is positive semidefinite, by definition we have $ax^2 + 2bxy + cy^2 \geq 0$ for any $x, y \in \mathbb{R}$. To show $\begin{bmatrix} aI & bI \\ bI & cI \end{bmatrix}$ is positive semidefinite, we just pick any vector $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ with compatible dimension and a straightforward calculation yields

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}^T \begin{bmatrix} aI & bI \\ bI & cI \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = a\|v_1\|^2 + 2bv_1^T v_2 + c\|v_2\|^2 = \sum_j \left(a(v_1^{(j)})^2 + 2bv_1^{(j)} v_2^{(j)} + c(v_2^{(j)})^2 \right) \geq 0$$

where $v_1^{(j)}$ is the j -th entry of v_1 and $v_2^{(j)}$ is the j -th entry of v_2 .

Now consider the other direction. Suppose $\begin{bmatrix} aI & bI \\ bI & cI \end{bmatrix}$ is positive semidefinite. For any $x, y \in \mathbb{R}$, we choose

$$v_1 = \begin{bmatrix} x \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} y \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then we have

$$\begin{bmatrix} x \\ y \end{bmatrix}^\top \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}^\top \begin{bmatrix} aI & bI \\ bI & cI \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \geq 0$$

Therefore, $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is positive semidefinite.

2.

(a) Assume x_1^* and x_2^* are both global min for f . We have $\nabla f(x_1^*) = \nabla f(x_2^*) = 0$. Since f is m -strongly convex, we have

$$\begin{aligned} f(x_1^*) &\geq f(x_2^*) + \frac{m}{2} \|x_1^* - x_2^*\|^2 \\ f(x_2^*) &\geq f(x_1^*) + \frac{m}{2} \|x_1^* - x_2^*\|^2 \end{aligned}$$

Adding the above two inequalities leads to the conclusion $\|x_1^* - x_2^*\| \leq 0$. Hence $x_1^* = x_2^*$. Therefore the global min for f is unique.

(b) The matrix $\begin{bmatrix} 1 - \rho^2 & -\alpha \\ -\alpha & \alpha^2 \end{bmatrix} + \lambda \begin{bmatrix} -2mL & m + L \\ m + L & -2 \end{bmatrix}$ is negative semidefinite if and only if

$$\begin{aligned} \rho^2 &\geq 1 - 2mL\lambda - \frac{(\lambda(m + L) - \alpha)^2}{\alpha^2 - 2\lambda} \\ \lambda &\geq \frac{\alpha^2}{2} \end{aligned}$$

Therefore we have $\|x_k - x^*\| \leq \rho^k \|x_0 - x^*\|$ if we can find $0 < \rho < 1$ and $\lambda \geq 0$ satisfying the above inequalities.

Now set $\lambda = \frac{1+t}{2}\alpha^2$ with some $t > 0$. Clearly $\lambda \geq \frac{\alpha^2}{2}$. Substituting $\lambda = \frac{1+t}{2}\alpha^2$ to the first inequality $\rho^2 \geq 1 - 2mL\lambda - \frac{(\lambda(m+L)-\alpha)^2}{\alpha^2-2\lambda}$ leads to the following inequality

$$\begin{aligned} \rho^2 &\geq 1 - mL(1+t)\alpha^2 + \frac{((1+t)\alpha(m+L) - 2)^2}{4t} \\ &= 1 - mL\alpha^2 - mL\alpha^2 t + \frac{(t\alpha(m+L) + \alpha(m+L) - 2)^2}{4t} \\ &= 1 - mL\alpha^2 - mL\alpha^2 t + \frac{\alpha^2(m+L)^2 t^2 + 2(\alpha(m+L) - 2)(m+L)\alpha t + (\alpha(m+L) - 2)^2}{4t} \\ &= 1 + \frac{\alpha^2(m^2 + L^2)}{2} - (m+L)\alpha + \frac{(L-m)^2 \alpha^2 t}{4} + \frac{(\alpha(m+L) - 2)^2}{4t} \end{aligned}$$

We want to choose the smallest ρ and associated λ that satisfy the above inequality. Hence we can choose t satisfying $\frac{(L-m)^2\alpha^2 t}{4} = \frac{(\alpha(m+L)-2)^2}{4t}$ and ρ satisfying

$$\rho^2 = 1 + \frac{\alpha^2(m^2 + L^2)}{2} - (m + L)\alpha + \frac{1}{2}((L - m)\alpha)\sqrt{(\alpha(m + L) - 2)^2}$$

When $\alpha \leq \frac{2}{m+L}$, we have

$$\begin{aligned}\rho^2 &= 1 + \frac{\alpha^2(m^2 + L^2)}{2} - (m + L)\alpha + \frac{1}{2}(L - m)\alpha(2 - \alpha(m + L)) \\ &= 1 - 2m\alpha + m^2\alpha^2 \\ &= (1 - m\alpha)^2\end{aligned}$$

Similarly, we have $\rho^2 = (1 - L\alpha)^2$ when $\alpha > \frac{2}{m+L}$. This is equivalent to $\rho = \max\{|1 - m\alpha|, |1 - L\alpha|\}$. Also notice that ρ^2 is required to be greater than 0 and smaller than 1, hence the formulas for ρ^2 only work for $\alpha < \frac{2}{L}$. Otherwise $(1 - L\alpha)^2 \geq 1$ when $L\alpha \geq 2$. Hence the statement is true and we prove the desired conclusion.