DATA130026.01 Optimization Assignment 9

Due Time: at the beginning of the class, May 29, 2024
姓名: 雍崔扬
学号: 21307140051

Problem 1

Let $F(x) = Ax + b$ be an affine function, with A an $n \times n$-matrix.
What properties of the matrix A correspond to the following conditions (a)-(e) on F?
Suppose that A is symmetric, so $F(x)$ is the gradient of a quadratic function.

(a) Monotonicity:
$(F(x) - F(y))^T(x-y) \geq 0 \ (\forall\ x, y)$

- **Solution:**
  对于任意 x, y 都有:
  $$\begin{aligned}(F(x) - F(y))^T(x-y) &= [(Ax+b) - (Ay+b)]^T(x-y) \\ &= (x-y)^T A(x-y) \\ &\geq 0\end{aligned}$$
  当且仅当 $A \succeq 0$

(b) Strict monotonicity:
$(F(x) - F(y))^T(x-y) > 0 \ \ (\forall\ x, y)$

- **Solution:**
  对于任意 x, y 都有:
  $$\begin{aligned}(F(x) - F(y))^T(x-y) &= [(Ax+b) - (Ay+b)]^T(x-y) \\ &= (x-y)^T A(x-y) \\ &> 0\end{aligned}$$
  当且仅当 $A \succ 0$

(c) Strong monotonicity (for the Euclidean norm):
$(F(x) - F(y))^T(x-y) \geq m\|x-y\|_2^2 \ \ (\forall\ x, y)$
where $m > 0$ is a positive constant.

- **Solution:**
  对于任意 x, y 都有:
  $$\begin{aligned}(F(x) - F(y))^T(x-y) - m\|x-y\|_2^2 &= [(Ax+b) - (Ay+b)]^T(x-y) - m\|x-y\|_2^2 \\ &= (x-y)^T(A - mI_n)(x-y) \\ &\geq 0\end{aligned}$$
  当且仅当 $A - mI_n \succeq 0$ (即 $\lambda_{\min}(A) \geq m$)

(d) Lipschitz continuity (for the Euclidean norm):
$\|F(x) - F(y)\|_2 \leq L\|x-y\|_2 \ \ (\forall\ x, y)$
where $L > 0$ is a positive constant.

- **Solution:**
  对于任意 x, y 都有:
  $$\begin{aligned}\|F(x) - F(y)\|_2^2 - L^2\|x-y\|_2^2 &= \|(Ax+b) - (Ay+b)\|_2^2 - L^2\|x-y\|_2^2 \\ &= (x-y)^T A^T A(x-y) - L^2(x-y)^T(x-y) \\ &= (x-y)^T(A^T A - L^2 I_n)(x-y) \\ &\leq 0\end{aligned}$$
  当且仅当 $A^T A - L^2 I_n \preceq 0$
  即 $\lambda_{\max}(A^T A) \leq L^2$
  即 $\|A\|_2 = \sigma_{\max}(A) \leq L$ (最大奇异值)

(e) Co-coercivity (for the Euclidean norm):
$(F(x) - F(y))^T(x-y) \geq \frac{1}{L}\|F(x) - F(y)\|_2^2 \ \ (\forall\ x, y)$
where $L > 0$ is a positive constant.

- **Solution:**
  对于任意 x, y 都有:
  $$(F(x)-F(y))^T(x-y)-\frac{1}{L}\|F(x)-F(y)\|_2^2$$
  $$=[(Ax+b)-(Ay+b)]^T(x-y)-\frac{1}{L}\|(Ax+b)-(Ay+b)\|_2^2$$
  $$=(x-y)^T A(x-y)-\frac{1}{L}(x-y)^T A^T A(x-y)$$
  $$=(x-y)^T(A-\frac{1}{L}A^T A)(x-y)$$
  $$\geq 0$$
  当且仅当 $A-\frac{1}{L}A^T A \succeq 0$
  即对于任意 $i=1,\dots,n$ 都有 $\lambda_i(A)-\frac{1}{L}(\lambda_i(A))^2 \geq 0$ 成立,
  即对于任意 $i=1,\dots,n$ 都有 $0 \leq \lambda_i(A) \leq L$ 成立.

Problem 2

Let f be a convex and continuously differentiable function over $\mathbb{R}^n$.
For a fixed $x \in \mathbb{R}^n$, define the function $g_x(y)=f(y)-\nabla f(x)^T y$

Suppose ∇f is L-Lipschitz continuous, i.e., $\|\nabla f(x)-\nabla f(y)\| \leq L\|x-y\|$ $(\forall\, x, y \in \mathbb{R}^n)$

(a) Prove that x is a minimizer of $g_x(y)$ over $\mathbb{R}^n$.

- **Solution:**
  容易验证 $g_x(y)=f(y)-\nabla f(x)^T y$ 是关于 $y \in \mathbb{R}^n$ 的凸函数,
  故其驻点记为全局最小值点.

  $g_x(y)$ 的梯度为 $\nabla_y g_x(y)=\nabla f(y)-\nabla f(x)$
  取 $y=x$ 得到 $\nabla_y g_x(x)=\nabla f(x)-\nabla f(x)=0_n$
  因此 x 是 $g_x(y)$ 的驻点，从而是 $g_x(y)$ 的全局最小值点.

(b) Show that for any $x, y \in \mathbb{R}^n$, $g_x(x) \leq g_x(y)-\frac{1}{2L}\|\nabla g_x(y)\|^2$

- **Lemma：(Quadratic Upper Bound, Nonlinear Programming 附录 A 命题 A.24)**
  (另一来源：江如俊教授 Note 6 Lemma 1.2)
  设 $f: \mathbb{R}^n \mapsto \mathbb{R}$ 一阶连续可微，给定标量 $L>0$
  若 ∇f 是 L-Lipschitz 连续的,
  即 $\|\nabla f(x+d)-\nabla f(x)\| \leq L\|d\|$ $(\forall\, x, d \in \mathbb{R}^n)$
  则对于任意 $x, d \in \mathbb{R}^n$ 我们有:
  $f(x+d) \leq f(x)+\nabla f(x)^T d+\frac{L}{2}\|d\|^2$
  即等价于 $\frac{L}{2}x^T x-f(x)$ 是凸函数.
  - **Proof:**
    定义函数 $g(t)=f(x+td)$ $(t \in [0,1])$
    由链式求导法则有 $\frac{d}{dt}g(t)=d^T\nabla f(x+td)$
    进而有:
    $$\begin{aligned}
    f(x+d)-f(x) &= g(1)-g(0) \\
    &= \int_0^1 \frac{d}{dt}g(t)dt \\
    &= \int_0^1 d^T\nabla f(x+td)dt \\
    &\leq \int_0^1 d^T\nabla f(x)dt+|\int_0^1 d^T(\nabla f(x+td)-\nabla f(x))dt| \\
    &\leq d^T\nabla f(x)+\int_0^1 \|d\|\cdot\|\nabla f(x+td)-\nabla f(x))\|dt \\
    &\leq d^T\nabla f(x)+\|d\|\int_0^1 Lt\|d\|dt \\
    &= d^T\nabla f(x)+\frac{L}{2}\|d\|^2
    \end{aligned}$$

- **Solution:**
  根据引理可知，对于任意 $z, y \in \mathbb{R}^n$ 都有:
  $g_x(z) - g_x(y) \le \nabla g_x(y)^T(z-y) + \dfrac{L}{2}\|z-y\|^2 \quad (①)$
  注意到右式关于 $z \in \mathbb{R}^n$ 是凸函数，
  因此右式对 z 最小化，只需求解驻点方程 $\nabla g_x(y) + L(z-y) = 0_n$ 即可
  解得 $z_\star = y - \frac{1}{L}\nabla g_x(y)$
  将 $z = z_\star = y - \frac{1}{L}\nabla g_x(y)$ 代入到 ① 式中得到:
  $$\begin{aligned} g_x(x) - g_x(y) &\le g_x(z_\star) - g_x(y) \qquad (\text{这一步利用了(a)的结论}, x = \arg\min_z g_x(z)) \\ &\le \nabla g_x(y)^T(z_\star - y) + \frac{L}{2}\|z_\star - y\|^2 \\ &= \nabla g_x(y)^T(y - \frac{1}{L}\nabla g_x(y) - y) + \frac{L}{2}\|y - \frac{1}{L}\nabla g_x(y) - y\|^2 \\ &= -\frac{1}{L}\|\nabla g_x(y)\|^2 + \frac{1}{2L}\|\nabla g_x(y)\|^2 \\ &= -\frac{1}{2L}\|\nabla g_x(y)\|^2 \end{aligned}$$
  命题得证.

(c) Show that for any $x, y \in \mathbb{R}^n$, $f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y)$

- **Solution:**
  根据 (b) 的结论，对于任意 $x, y \in \mathbb{R}^n$ 都有:
  $g_x(x) \le g_x(y) - \frac{1}{2L}\|\nabla g_x(y)\|^2$
  代入 $\begin{cases} g_x(x) = f(x) - \nabla f(x)^T x \\ g_x(y) = f(y) - \nabla f(x)^T y \\ \nabla g_x(y) = \nabla f(y) - \nabla f(x) \end{cases}$
  即得 $f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y)$

(d) Show that $L\langle \nabla f(y) - \nabla f(x), y - x\rangle \ge \|\nabla f(y) - \nabla f(x)\|^2$ $(\forall\, x, y \in \mathbb{R}^n)$

- **Solution:**
  根据 2.(c) 的结论可知，对于任意 $x, y \in \mathbb{R}^n$ 都有:
  $f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y)$
  调换 x, y 的顺序，可知:
  $f(y) + \nabla f(y)^T(x-y) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \le f(x)$
  两式相加，整理即得 $L\langle \nabla f(y) - \nabla f(x), y - x\rangle \ge \|\nabla f(y) - \nabla f(x)\|^2$
  命题得证.
  - 一个经典的错误做法是利用一阶凸性条件:
    $$\begin{aligned} \|\nabla f(x) - \nabla f(y)\|^2 &\le 2L(f(x) - f(y) - \nabla f(y)^T(x-y)) \\ &= 2L(f(x) - (f(x) + \nabla f(x)^T(y-x)) - \nabla f(y)^T(x-y)) \\ &= 2L\langle \nabla f(y) - \nabla f(x), y - x\rangle \end{aligned}$$
    这样会得到 $2L\langle \nabla f(y) - \nabla f(x), y - x\rangle \ge \|\nabla f(y) - \nabla f(x)\|^2$ $(\forall\, x, y \in \mathbb{R}^n)$
    导致怀疑题目出错了 (然后贻笑大方)

Problem 3 (Faster convergence for gradient descent)

Let f be a convex and continuously differentiable function over $\mathbb{R}^n$.
Suppose ∇f is L-Lipschitz continuous.

Let $\{x^{(k)}\}$ be a sequence generated by the **gradient descent method** (GD)
for minimizing f with a constant step-size $t = \frac{\phi}{L}$, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.
Then, for any optimal solution $x_\star$ and nonnegative integer n, it holds that:
$f(x^{(n)}) - f(x_\star) \le \frac{1}{n\phi} \cdot \frac{L}{2}\|x^{(0)} - x_\star\|^2$

- **Hint:** Use the inequality in 2(c), the sufficient descent lemma, and the identity:

$$\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2=\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k+1)}-x_\star\rangle \quad (1)$$

- **Lemma：(最速下降法的充分下降引理，江如俊教授 Note 6 Lemma 1.3)**
  设 $f:\mathbb{R}^n\mapsto\mathbb{R}$ 一阶连续可微.
  若 ∇f 是 L-Lipschitz 连续的，
  即对于给定标量 $L>0$ 有 $\|\nabla f(x)-\nabla f(y)\|\leq L\|x-y\|\ \ (\forall\, x,y\in\mathbb{R}^n)$ 成立
  则对于任意 $0<t<\frac{2}{L}$ 我们有 $f(x)-f(x-t\nabla f(x))\geq t(1-\frac{Lt}{2})\|\nabla f(x)\|^2$ 成立.
  - 在实际应用中，我们通常选取 $t=\arg\max\limits_{0<t<\frac{2}{L}}\{t(1-\frac{Lt}{2})\}=\frac{1}{L}$，
    这样有 $f(x)-f(x-t\nabla f(x))\geq\frac{1}{2L}\|\nabla f(x)\|^2$ 成立.
    (江老师将这种情况标注为 Exact Line Search 的情况，我认为不太妥当)
  - **充分下降引理的证明：**
    根据**二次上界引理** (参见 Problem 2 (b) 的解答) 可知：
    对于任意下降方向 d，我们都有 $f(x+d)\leq f(x)+\nabla f(x)^Td+\frac{L}{2}\|d\|^2$ 成立.
    取 $d=-t\nabla f(x)\ \ (0<t<\frac{2}{L})$ **(最速下降法的下降方向)**，我们有：
    $$\begin{aligned}f(x-t\nabla f(x))&\leq f(x)+\nabla f(x)^T(-t\nabla f(x))+\frac{L}{2}\|-t\nabla f(x)\|^2\\&=f(x)-t\|\nabla f(x)\|^2+\frac{Lt^2}{2}\|\nabla f(x)\|^2\\&=f(x)-t(1-\frac{Lt}{2})\|\nabla f(x)\|^2\end{aligned}$$
    即得到 $f(x)-f(x-t\nabla f(x))\geq t(1-\frac{Lt}{2})\|\nabla f(x)\|^2$ 成立.

Solution:

- 根据题设，步长 $t=\frac{\phi}{L}=\frac{1+\sqrt{5}}{2L}<\frac{2}{L}$
  因此可以应用**充分下降引理**可知，对于任意 $x\in\mathbb{R}^n$，我们都有：
  $$f(x)-f(x-t\nabla f(x))\geq t(1-\frac{Lt}{2})\|\nabla f(x)\|^2$$
  取 $x=x^{(k)}$ 并代入 $t=\frac{\phi}{L}$ 即得：
  $$\begin{aligned}f(x^{(k)})-f(x^{(k+1)})&=f(x^{(k)})-f(x^{(k)}-t\nabla f(x^{(k)}))\\&\geq t(1-\frac{Lt}{2})\|\nabla f(x^{(k)})\|^2\\&=\frac{\phi}{L}(1-\frac{L}{2}\cdot\frac{\phi}{L})\|\nabla f(x^{(k)})\|^2\\&=\frac{1}{L}(\phi-\frac{\phi^2}{2})\|\nabla f(x^{(k)})\|^2\quad(\text{use golden ratio function }\phi^2=\phi+1)\\&=\frac{1}{L}(\phi-\frac{\phi+1}{2})\|\nabla f(x^{(k)})\|^2\\&=\frac{\phi-1}{2L}\|\nabla f(x^{(k)})\|^2\end{aligned}$$
  因此我们有：
  $$\begin{aligned}-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2&\geq-\frac{1}{\phi-1}(f(x^{(k)})-f(x^{(k+1)}))\quad(\text{use golden ratio function }\phi^2=\phi+1)\\&=-\frac{\phi^2-\phi}{\phi-1}(f(x^{(k)})-f(x^{(k+1)}))\\&=-\phi(f(x^{(k)})-f(x^{(k+1)}))\\&=\phi(f(x^{(k+1)})-f(x^{(k)}))\end{aligned}$$
- 根据 **2.(c)** 的结论，对于任意 $x,y\in\mathbb{R}^n$，我们都有：
  $$f(x)+\nabla f(x)^T(y-x)+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2\leq f(y)$$
  取 $\begin{cases}y=x_\star\\x=x^{(k)}\end{cases}$ 可得：
  $$\begin{aligned}f(x^{(k)})-f(x_\star)&\leq\nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})-\nabla f(x_\star)\|^2\\&=\nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2\end{aligned}$$

- 根据**恒等式**我们有：
$$\begin{aligned}
&\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2\\
&=\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k+1)}-x_\star\rangle\\
&=\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k+1)}-x^{(k)}\rangle+L\langle x^{(k)}-x^{(k+1)},x^{(k)}-x_\star\rangle\\
&=-\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k)}-x_\star\rangle\\
&=-\frac{L}{2}\|\frac{\phi}{L}\nabla f(x^{(k)})\|^2+L\langle\frac{\phi}{L}\nabla f(x^{(k)}),x^{(k)}-x_\star\rangle\\
&=-\frac{\phi^2}{2L}\|\nabla f(x^{(k)})\|^2+\phi\nabla f(x^{(k)})^T(x^{(k)}-x_\star)\quad(\text{use golden ratio function }\phi^2=\phi+1)\\
&=-\frac{\phi+1}{2L}\|\nabla f(x^{(k)})\|^2+\phi\nabla f(x^{(k)})^T(x^{(k)}-x_\star)\\
&=\phi\cdot\{\nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2\}-\frac{1}{2L}\nabla\|f(x^{(k)})\|^2
\end{aligned}$$
- 向上式代入之前的结论 $\begin{cases}-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2\geq\phi(f(x^{(k+1)})-f(x^{(k)}))\\ \nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2\geq f(x^{(k)})-f(x_\star)\end{cases}$

  我们有：
$$\begin{aligned}
&\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2\\
&=\phi\cdot\{\nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2\}-\frac{1}{2L}\nabla\|f(x^{(k)})\|^2\\
&\geq\phi(f(x^{(k)})-f(x_\star))+\phi(f(x^{(k+1)})-f(x^{(k)}))\\
&=\phi(f(x^{(k+1)})-f(x_\star))
\end{aligned}$$
- 上式对 k 从 0 到 $n-1$ 求和，我们有：
$$\begin{aligned}
&\sum_{k=0}^{n-1}\{\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2\}\\
&=\frac{L}{2}(\|x^{(0)}-x_\star\|^2-\|x^{(n)}-x_\star\|^2)\\
&\geq\phi\sum_{k=0}^{n-1}\{f(x^{(k+1)})-f(x_\star)\}\\
&\geq\phi\cdot n(f(x^{(n)})-f(x_\star))\quad(\text{since the gradient method always descends for a convex function }f)
\end{aligned}$$
  于是我们有：
$$\begin{aligned}
f(x^{(n)})-f(x_\star)&\leq\frac{1}{n\phi}\cdot\frac{L}{2}(\|x^{(0)}-x_\star\|^2-\|x^{(n)}-x_\star\|^2)\\
&\leq\frac{1}{n\phi}\cdot\frac{L}{2}\|x^{(0)}-x_\star\|^2
\end{aligned}$$
  命题得证.

Problem 4 (Faster convergence for gradient descent)

Let f be a convex and continuously differentiable function over $\mathbb{R}^n$.
Suppose ∇f is L-Lipschitz continuous.
Consider problem $(\mathrm{P})\ \min_{x\in\mathbb{R}^n}\ f(x)$

(a) Double sufficient decrease.
Let $\{x^{(k)}\}$ be a sequence generated by the gradient descent method (GD)
with a constant step-size satisfying $Lt\in(0,1]$.
Then, for any nonnegative integer k, it holds that:
$f(x^{(k)})-f(x^{(k+1)})\geq\frac{t}{2}\|\nabla f(x^{(k)})\|^2+\frac{t}{2}\|\nabla f(x^{(k+1)})\|^2$

- **Hint:** Use 2(c)
- **Solution**:
  根据题设，我们知道 $0<Lt\leq 1$，因此 $\frac{1}{L}\geq t$

根据 **2.(c)** 的结论，对于任意 $x, y \in \mathbb{R}^n$，我们都有：
$f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y)$

取 $\begin{cases} y = x^{(k)} \\ x = x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)}) \end{cases}$ 即得：

$$\begin{aligned} f(x^{(k)}) - f(x^{(k+1)}) &= \nabla f(x^{(k+1)})^T(x^{(k+1)} - x^{(k)}) + \frac{1}{2L}\|\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\|^2 \\ &\ge \nabla f(x^{(k+1)})^T(-t\nabla f(x^{(k)})) + \frac{t}{2}\|\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\|^2 \\ &= -t\nabla f(x^{(k+1)})^T\nabla f(x^{(k)}) + \frac{t}{2}(\|\nabla f(x^{(k+1)})\|^2 - 2\nabla f(x^{(k+1)})^T\nabla f(x^{(k)}) + \|\nabla f(x^{(k)})\|^2) \\ &= \frac{t}{2}\|\nabla f(x^{(k)})\|^2 + \frac{t}{2}\|\nabla f(x^{(k+1)})\|^2 \end{aligned}$$

命题得证.

(b) Monotonicity of the gradient.
Let $\{x^{(n)}\}$ be a sequence generated by the gradient descent method (GD)
for solving problem (P) with a constant step-size satisfying $Lt \in (0, 2]$.
Then, for any nonnegative integer k, it holds that:
$\|\nabla f(x^{(k+1)})\| \le \|\nabla f(x^{(k)})\|$

- **Hint:** Use 2(d)
- **Solution:**
  根据 2 (d) 的结论可知，对于任意 $x, y \in \mathbb{R}^n$ 我们都有：
  $L\langle \nabla f(y) - \nabla f(x), y - x\rangle \ge \|\nabla f(y) - \nabla f(x)\|^2$
  取 $\begin{cases} x = x^{(k)} \\ y = x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)}) \end{cases}$ 上式就变为：
  $L\langle \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), x^{(k+1)} - x^{(k)}\rangle = \|\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\|^2$
  代入 $x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$ 即得：
  $-Lt\langle \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}), \nabla f(x^{(k)})\rangle = \|\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})\|^2$
  将左右式分别拆开，合并同类项，可得：
  $$\begin{aligned} \|\nabla f(x^{(k+1)})\|^2 &= (Lt-1)\|\nabla f(x^{(k)})\|^2 + (2-Lt)\nabla f(x^{(k+1)})^T\nabla f(x^{(k)}) \\ &\le (Lt-1)\|\nabla f(x^{(k)})\|^2 + (2-Lt)\|\nabla f(x^{(k+1)})\| \cdot \|\nabla f(x^{(k)})\| \\ &= \|\nabla f(x^{(k+1)})\| \cdot \|\nabla f(x^{(k)})\| \end{aligned}$$
  于是有 $\|\nabla f(x^{(k+1)})\| \le \|\nabla f(x^{(k)})\|$ 成立.
  命题得证.

(c) Ordering of optimality measures.
Consider problem (P).
Then, for any $x_\star \in X_\star$, it holds that:
$\frac{1}{2L}\|\nabla f(x)\|^2 \le f(x) - f(x_\star) \le \frac{L}{2}\|x - x_\star\|^2 \quad (\forall\, x \in \mathbb{R}^n)$

Solution:

- 一方面，根据 **2.(c)** 的结论，对于任意 $x, y \in \mathbb{R}^n$，我们都有：
  $f(x) + \nabla f(x)^T(y-x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y)$
  不失一般性，我们调换 x, y 位置，即有：
  $f(y) + \nabla f(y)^T(x-y) + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2 \le f(x)$
  取 $y = x_\star$ 可得：
  $$\begin{aligned} f(x) - f(x_\star) &\ge \nabla f(x_\star)^T(x - x_\star) + \frac{1}{2L}\|\nabla f(x_\star) - \nabla f(x)\|^2 \\ &= \frac{1}{2L}\|\nabla f(x)\|^2 \quad (\text{for all } x_\star \in X_\star \text{ it holds that } \nabla f(x_\star) = 0_n) \end{aligned}$$

- 另一方面，根据 **2.(b)** 中给出的**二次上界引理**，
  由 ∇f 是 L-Lipschitz 连续的可推出：
  对于任意 $x, d \in \mathbb{R}^n$ 我们都有 $f(x+d) \leq f(x) + \nabla f(x)^T d + \frac{L}{2}\|d\|^2$ 成立，
  即 $\frac{L}{2}x^T x - f(x)$ 是凸函数.
  做记号代换 $\begin{cases} x \to x_\star \\ d \to x - x_\star \end{cases}$ 即得：
  $$\begin{aligned} f(x) &\leq f(x_\star) + \nabla f(x_\star)^T (x - x_\star) + \frac{L}{2}\|x - x_\star\|^2 \\ &= f(x_\star) + \frac{L}{2}\|x - x_\star\|^2 \quad (\text{for all } x_\star \in X_\star \text{ it holds that } \nabla f(x_\star) = 0_n) \end{aligned}$$
  即 $f(x) - f(x_\star) \leq \frac{L}{2}\|x - x_\star\|^2$

综上所述，$\frac{1}{2L}\|\nabla f(x)\|^2 \leq f(x) - f(x_\star) \leq \frac{L}{2}\|x - x_\star\|^2 \ \ (\forall\, x \in \mathbb{R}^n)$
命题得证.

(d) Let $\{x^{(k)}\}$ be a sequence generated by the gradient descent method (GD)
for solving problem (P) with a constant step-size satisfying $Lt \in (0,1]$.
Then, for any $x_\star \in X_\star$ and nonnegative integer m, it holds that:

- **Objective value gap:** $f(x^{(m)}) - f(x_\star) \leq \frac{1}{2Ltm+1}\frac{L}{2}\|x^{(0)} - x_\star\|^2$
- **Gradient norm gap:** $\|\nabla f(x^{(m)})\| \leq \frac{L}{Ltm+1}\|x^{(0)} - x_\star\|$

Hint:
You may first show that:

$$f(x^{(n)}) - f(x^{(m)}) \geq 2(m-n)\frac{t}{2}\|\nabla f(x^{(m)})\|^2 \ \ (\forall\, m \geq n) \tag{2}$$

You may use (1), 2(c) and 4(b) to obtain:
$\frac{L}{2}\|x^{(k)} - x_\star\|^2 - \frac{L}{2}\|x^{(k+1)} - x_\star\|^2 \geq Lt[f(x^{(k)}) - f(x_\star)] + Lt(1 - Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2 \ \ (\forall\, m > k)$
Telescope the above inequality and use (2).
Recall the identity (1) in problem 3:

$$\frac{L}{2}\|x^{(k)} - x_\star\|^2 - \frac{L}{2}\|x^{(k+1)} - x_\star\|^2 = \frac{L}{2}\|x^{(k)} - x^{(k+1)}\|^2 + L\langle x^{(k)} - x^{(k+1)}, x^{(k+1)} - x_\star\rangle \tag{1}$$

Solution:

- 首先根据 **4.(a)** 的结论我们知道：
  在步长 t 满足 $Lt \in (0,1]$ 的情况下，对于任意非负整数 k 都有：
  $f(x^{(k)}) - f(x^{(k+1)}) \geq \frac{t}{2}\|\nabla f(x^{(k)})\|^2 + \frac{t}{2}\|\nabla f(x^{(k+1)})\|^2$
  结合 **4.(b)** 的结论我们知道：
  在步长 t 满足 $Lt \in (0,2]$ 的情况下，对于任意非负整数 k 都有：
  $\|\nabla f(x^{(k+1)})\| \leq \|\nabla f(x^{(k)})\|$
  综合上述两个结论可知：
  $$\begin{aligned} f(x^{(k)}) - f(x^{(k+1)}) &\geq 2 \cdot \frac{t}{2}\|\nabla f(x^{(k+1)})\|^2 \\ &\geq 2 \cdot \frac{t}{2}\|\nabla f(x^{(m)})\|^2 \quad (\forall\, m > k) \end{aligned}$$
  此式对 k 从 n 到 $m-1$ 求和 (假设 $n < m$)，便得到：
  $$\begin{aligned} f(x^{(n)}) - f(x^{(m)}) &= \sum_{k=n}^{m-1}\{f(x^{(k)}) - f(x^{(k+1)})\} \\ &\geq \sum_{k=n}^{m-1} t\|\nabla f(x^{(m)})\|^2 \\ &= 2(m-n)\frac{t}{2}\|\nabla f(x^{(m)})\|^2 \end{aligned}$$
  此式对 $n = m$ 的情况同样成立，于是我们有：

$$f(x^{(n)}) - f(x^{(m)}) \geq 2(m-n)\frac{t}{2}\|\nabla f(x^{(m)})\|^2 \ \ (\forall\, m \geq n) \tag{2}$$

- 根据恒等式 (1) 可知:
  $\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2$
  $=\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k+1)}-x_\star\rangle$
  $=\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k+1)}-x^{(k)}\rangle+L\langle x^{(k)}-x^{(k+1)},x^{(k)}-x_\star\rangle$
  $=-\frac{L}{2}\|x^{(k)}-x^{(k+1)}\|^2+L\langle x^{(k)}-x^{(k+1)},x^{(k)}-x_\star\rangle$
  $=-\frac{L}{2}\|t\nabla f(x^{(k)})\|^2+L\langle t\nabla f(x^{(k)}),x^{(k)}-x_\star\rangle$
  $=-\frac{Lt^2}{2}\|\nabla f(x^{(k)})\|^2+Lt\nabla f(x^{(k)})^T(x^{(k)}-x_\star)$
  - 根据 **2.(c)** 的结论，对于任意 $x,y\in\mathbb{R}^n$，我们都有:
    $f(x)+\nabla f(x)^T(y-x)+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2\le f(y)$
    取 $\begin{cases}y=x_\star\\x=x^{(k)}\end{cases}$ 可得:
    $$\begin{aligned}f(x^{(k)})-f(x_\star)&\le\nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})-\nabla f(x_\star)\|^2\\&=\nabla f(x^{(k)})^T(x^{(k)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(k)})\|^2\end{aligned}$$
    因此 $Lt\nabla f(x^{(k)})^T(x^{(k)}-x_\star)\ge Lt[f(x^{(k)})-f(x_\star)]+Lt\cdot\frac{1}{2L}\|\nabla f(x^{(k)})\|^2$

  将刚刚得到的不等式代入恒等式，便得到:
  $\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2$
  $=-\frac{Lt^2}{2}\|\nabla f(x^{(k)})\|^2+Lt\nabla f(x^{(k)})^T(x^{(k)}-x_\star)$
  $\ge-\frac{Lt^2}{2}\|\nabla f(x^{(k)})\|^2+Lt[f(x^{(k)})-f(x_\star)]+Lt\cdot\frac{1}{2L}\|\nabla f(x^{(k)})\|^2$
  $=Lt[f(x^{(k)})-f(x_\star)]+Lt(1-Lt)\frac{1}{2L}\|\nabla f(x^{(k)})\|^2$
  $\ge Lt[f(x^{(k)})-f(x_\star)]+Lt(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2\quad(\text{use conclusion of }4.(b),\ \forall\, m\ge k)$

  这样我们就得到了下面的不等式 $(\forall\, m\ge k)$:
  $$\frac{L}{2}\|x^{(k)}-x_\star\|^2-\frac{L}{2}\|x^{(k+1)}-x_\star\|^2\ge Lt[f(x^{(k)})-f(x_\star)]+Lt(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2\tag{3}$$

- (3) 式对 k 从 0 到 $m-1$ 求和，即得:
  $\frac{L}{2}\{\|x^{(0)}-x_\star\|^2-\|x^{(m)}-x_\star\|^2\}$
  $=\frac{L}{2}\sum_{k=0}^{m-1}\{\|x^{(k)}-x_\star\|^2-\|x^{(k+1)}-x_\star\|^2\}\quad(\text{use inequality }(3))$
  $\ge Lt\sum_{k=0}^{m-1}\{f^{(k)}-f(x_\star)\}+Lt\sum_{k=0}^{m-1}(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2$
  $=Lt\sum_{k=0}^{m-1}\{f^{(k)}-f(x_\star)\}+m\cdot Lt(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2$

  因此我们有:
  $$\frac{L}{2}\|x^{(0)}-x_\star\|^2\ge\frac{L}{2}\|x^{(m)}-x_\star\|^2+Lt\sum_{k=0}^{m-1}\{f^{(k)}-f(x_\star)\}+Ltm(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2\tag{4}$$

- 根据 (2) 式我们知道:
  $f(x^{(n)})-f(x^{(m)})\ge2(m-n)\frac{t}{2}\|\nabla f(x^{(m)})\|^2\ \ (\forall\, m\ge n)$
  因此对于任意 $k\le m$ 我们都有:
  $f(x^{(k)})-f(x_\star)\ge f(x^{(m)})-f(x_\star)+2(m-k)\frac{t}{2}\|\nabla f(x^{(m)})\|^2$

因此我们有:

$$\sum_{k=0}^{m-1}\{f^{(k)}-f(x_\star)\} \geq \sum_{k=0}^{m-1}\{f(x^{(m)})-f(x_\star)+2(m-k)\frac{t}{2}\|\nabla f(x^{(m)})\|^2\}$$
$$= m\cdot\{f(x^{(m)})-f(x_\star)\}+2\sum_{k=0}^{m-1}(m-k)\frac{t}{2}\|\nabla f(x^{(m)})\|^2$$
$$= m\cdot\{f(x^{(m)})-f(x_\star)\}+2\frac{m(m+1)}{2}\frac{t}{2}\|\nabla f(x^{(m)})\|^2$$

将上式代入 (4) 式 (对应于下式的第二个不等号) 即得:

$$\frac{L}{2}\|x^{(0)}-x_\star\|^2$$
$$\geq \frac{L}{2}\|x^{(m)}-x_\star\|^2+Lt\sum_{k=0}^{m-1}\{f^{(k)}-f(x_\star)\}+Ltm(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2$$
$$\geq \frac{L}{2}\|x^{(m)}-x_\star\|^2+Ltm\{f(x^{(m)})-f(x_\star)\}+Ltm(m+1)\frac{t}{2}\|\nabla f(x^{(m)})\|^2+Ltm(1-Lt)\frac{1}{2L}\|\nabla f(x^{(m)})\|^2$$
$$= \frac{L}{2}\|x^{(m)}-x_\star\|^2+Ltm\{f(x^{(m)})-f(x_\star)\}+Ltm(\frac{mt}{2}+\frac{1}{2L})\|\nabla f(x^{(m)})\|^2$$

因此我们有:

$$\frac{L}{2}\|x^{(0)}-x_\star\|^2 \geq \frac{L}{2}\|x^{(m)}-x_\star\|^2+Ltm\{f(x^{(m)})-f(x_\star)\}+Ltm(\frac{mt}{2}+\frac{1}{2L})\|\nabla f(x^{(m)})\|^2 \quad (5)$$

- 根据 **2.(c)** 的结论，对于任意 $x,y\in\mathbb{R}^n$，我们都有:
  $f(x)+\nabla f(x)^T(y-x)+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2\leq f(y)$
  取 $\begin{cases} y=x_\star \\ x=x^{(m)} \end{cases}$ 可得:
  $$f(x^{(m)})-f(x_\star)\leq \nabla f(x^{(m)})^T(x^{(m)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(m)})-\nabla f(x_\star)\|^2$$
  $$= \nabla f(x^{(m)})^T(x^{(m)}-x_\star)-\frac{1}{2L}\|\nabla f(x^{(m)})\|^2$$
  因此我们有:
  $$\nabla f(x^{(m)})^T(x^{(m)}-x_\star)\geq f(x^{(m)})-f(x_\star)+\frac{1}{2L}\|\nabla f(x^{(m)})\|^2 \quad (6)$$
  利用 (5) 式我们有:
  $$\frac{L}{2}\|x^{(0)}-x_\star\|^2$$
  $$\geq \frac{L}{2}\|x^{(m)}-x_\star\|^2+Ltm\{f(x^{(m)})-f(x_\star)\}+Ltm(\frac{mt}{2}+\frac{1}{2L})\|\nabla f(x^{(m)})\|^2 \quad \text{(this is (5))}$$
  $$= Ltm\{f(x^{(m)})-f(x_\star)\}+\frac{L}{2}\|x^{(m)}-x_\star\|^2+\frac{(Ltm+1)^2}{2L}\|\nabla f(x^{(m)})\|^2-\frac{Ltm+1}{2L}\|\nabla f(x^{(m)})\|^2$$
  $$\text{(use inequality of Arithmetic and Geometric means)}$$
  $$\geq Ltm\{f(x^{(m)})-f(x_\star)\}+2\cdot\sqrt{\frac{L}{2}\frac{(Ltm+1)^2}{2L}}\nabla f(x^{(m)})^T(x^{(m)}-x_\star)-\frac{Ltm+1}{2L}\|\nabla f(x^{(m)})\|^2$$
  $$\text{(use inequality (6))}$$
  $$\geq Ltm\{f(x^{(m)})-f(x_\star)\}+(Ltm+1)\{f(x^{(m)})-f(x_\star)+\frac{1}{2L}\|\nabla f(x^{(m)})\|^2\}-\frac{Ltm+1}{2L}\|\nabla f(x^{(m)})\|^2$$
  $$= (2Ltm+1)\{f(x^{(m)})-f(x_\star)\}$$
  因此我们有 **Objective value gap** $f(x^{(m)})-f(x_\star)\leq\frac{1}{2Ltm+1}\frac{L}{2}\|x^{(0)}-x_\star\|^2$
  4.(d) 的第一个命题得证.
- 根据 **4.(c)** 我们有 $\frac{1}{2L}\|\nabla f(x)\|^2\leq f(x)-f(x_\star)\leq\frac{L}{2}\|x-x_\star\|^2\ \ (\forall\, x\in\mathbb{R}^n)$
  因此 $\frac{1}{2L}\|\nabla f(x^{(m)})\|^2\leq f(x^{(m)})-f(x_\star)\leq\frac{L}{2}\|x^{(m)}-x_\star\|^2$

将 $\begin{cases} \frac{L}{2}\|x^{(m)}-x_\star\|^2 \geq \frac{1}{2L}\|\nabla f(x^{(m)})\|^2 \\ f(x^{(m)})-f(x_\star) \geq \frac{1}{2L}\|\nabla f(x^{(m)})\|^2 \end{cases}$ 代入 (5) 式可知:

$$\begin{aligned}
&\frac{L}{2}\|x^{(0)}-x_\star\|^2 \\
&\geq \frac{L}{2}\|x^{(m)}-x_\star\|^2 + Ltm\{f(x^{(m)})-f(x_\star)\} + Ltm(\frac{mt}{2}+\frac{1}{2L})\|\nabla f(x^{(m)})\|^2 \quad \text{(this is (5))} \\
&\geq \frac{1}{2L}\|\nabla f(x^{(m)})\|^2 + Ltm\cdot\frac{1}{2L}\|\nabla f(x^{(m)})\|^2 + Ltm(\frac{mt}{2}+\frac{1}{2L})\|\nabla f(x^{(m)})\|^2 \\
&= \frac{1}{2L}(1+Ltm+Ltm\frac{mt}{2}\cdot 2L+Ltm)\|\nabla f(x^{(m)})\|^2 \\
&= \frac{1}{2L}(L^2t^2m^2+2Ltm+1)\|\nabla f(x^{(m)})\|^2 \\
&= \frac{1}{2L}(Ltm+1)^2\|\nabla f(x^{(m)})\|^2
\end{aligned}$$

因此我们有 $\frac{L}{2}\|x^{(0)}-x_\star\|^2 \geq \frac{1}{2L}(Ltm+1)^2\|\nabla f(x^{(m)})\|^2$

即有 **Gradient norm gap:** $\|\nabla f(x^{(m)})\| \leq \frac{L}{Ltm+1}\|x^{(0)}-x_\star\|$

4.(d) 的第二个命题得证.

综上所述，命题得证.