ORIE 6365 — Continuous Optimization: Algorithms and Complexity, **Spring 2026**

Homework 2

Released: Feb 27, 2026 *Due: Mar 14, 2026*

1 On Lower Bounds

In the lectures, we have proved the following result about the fundamental limitation of first-order methods on problem class of smooth convex minimization (see lectures 7 and 8):

Theorem 1. Let $L > 0$ and $K \geq 1$ be fixed. Then, for any first-order optimization algorithm, such that

$$x_{k+1} \quad \in \quad \mathrm{span}\Big\{\nabla f(x_0), \ldots, \nabla f(x_k)\Big\}, \tag{1}$$

running for K iterations, there exist a convex function $f : \mathbb{R}^n \to \mathbb{R}$ with $n \geq 2K+1$ such that

1. ∇f is Lipschitz with constant L;
2. For the output x_K of the algorithm, the following holds:

$$f(x_K) - f^\star \quad \geq \quad \tfrac{3L\|x_0 - x^\star\|^2}{2^6(K+1)^2}. \tag{2}$$

Problem 1. In Theorem 1, we allow the parameters $L > 0$ (the Lipschitz constant), and $K \geq 1$ (the number of iterations) to be arbitrarily fixed. However, another important parameter, that is the initial distance to the solution $\|x_0 - x^\star\|$, appears on the right hand side of (2) without any specification. In this problem, we aim to prove the following more general result:

Theorem 2. Let $L > 0$, $R > 0$, and $K \geq 1$ be fixed. Then, for any first-order optimization algorithm, such that

$$x_{k+1} \quad \in \quad \mathrm{span}\Big\{\nabla f(x_0), \ldots, \nabla f(x_k)\Big\}, \tag{3}$$

running for K iterations, there exists a convex function $f : \mathbb{R}^n \to \mathbb{R}$ with $n \geq 2K+1$ such that

1. ∇f is Lipschitz with constant L;
2. $\|x_0 - x^\star\| = R$, where $x^\star$ is the minimizer of f;
3. For the output x_K of the algorithm, the following holds:

$$f(x_K) - f^\star \quad \geq \quad c \cdot \tfrac{LR^2}{(K+1)^2}, \tag{4}$$

   where $c > 0$ is an absolute numerical constant that depends neither on L, K, nor R.

- Consider the following quadratic objective:

$$f_k(x) \quad := \quad \tfrac{\alpha}{2}\Big[\sum_{i=1}^{k-1}(x^{(i)} - x^{(i+1)})^2 + \sum_{i=k}^{n}(x^{(i)})^2\Big] - \beta x^{(1)}, \qquad x \in \mathbb{R}^n,$$

  where $\alpha > 0$ and $\beta > 0$ are parameters. Compute the Lipschitz constant of the gradient $L > 0$ and the squared norm of the solution: $\|x_k^\star\|_2^2$, where $x_k^\star = \arg\min_{x \in \mathbb{R}^n} f_k(x)$, for this objective.

- Prove Theorem 2.

2 On Composite Problems

We consider fully composite optimization problem,

$$\min_{x \in Q} \Big[\varphi(x) \quad := \quad F(x, f(x)) \Big], \tag{5}$$

where $Q \subseteq \mathbb{R}^n$ is a convex set, $F : Q \times \mathbb{R}^m \to \mathbb{R}$ is a convex function on its entire domain $Q \times \mathbb{R}^n$, but not-necessary differentiable, and $f : Q \to \mathbb{R}^m$ is a differentiable mapping, $f(x) = [f_1(x), \ldots, f_m(x)]^\top$. We use $\| \cdot \|$ for the standard Euclidean norms. We assume that

1. Each component $f_i : Q \to \mathbb{R}$, $1 \leq i \leq m$, is a differentiable convex function that has the Lipschitz continuous gradient with constant $L^{(i)} > 0$:

   $$\|\nabla f_i(x) - \nabla f_i(y)\| \quad \leq \quad L^{(i)} \|x - y\|, \qquad x, y \in Q \subseteq \mathbb{R}^n.$$

   We denote by $L = (L^{(1)}, \ldots, L^{(m)})^\top \in \mathbb{R}^m$ the vector of all Lipschitz constants.

2. For every $x \in Q$, the function $F(x, \cdot)$ is *monotone*. That is, for any two vectors $u_1, u_2 \in \mathbb{R}^m$ such that $u_1^{(i)} \leq u_2^{(i)}$ for all $1 \leq i \leq m$, it holds:

   $$F(x, u_1) \quad \leq \quad F(x, u_2).$$

3. For every $x \in Q$, the function $F(x, \cdot)$ is Lipschitz with constant $M > 0$. That is:

   $$|F(x, u_1) - F(x, u_2)| \quad \leq \quad M \|u_1 - u_2\|, \qquad u_1, u_2 \in \mathbb{R}^m.$$

Problem 2. Show that function $\varphi(x)$ is convex.

Problem 3. Consider iterations of the fully composite version of the *basic gradient method* for solving (5). Starting with some $x_0 \in Q$, we iterate, for $k \geq 0$:

$$x_{k+1} \quad = \quad \arg\min_{y \in Q} \Big\{ F(x_k, f(x_k) + \nabla f(x_k)(y - x_k)) + \tfrac{\alpha}{2} \|y - x_k\|^2 \Big\}, \tag{6}$$

where $\alpha > 0$ is a fixed regularization parameter, and $\nabla f(x_k) \in \mathbb{R}^{m \times n}$ denotes the Jacobian of the mapping f computed at point x_k. We aim to prove the convergence rate for this process. Note that due to the composite structure, our previous reasoning for the gradient method is not applicable here, and we need to come up with another proof technique.

- Using the properties of the fully composite problem and definition (6) of one step, show that

  $$\varphi(x_{k+1}) \quad \leq \quad \varphi(y) + \tfrac{\alpha}{2} \|y - x_k\|^2, \qquad \forall y \in Q, \tag{7}$$

  as soon as $\alpha \geq M \|L\|$.

- Substituting $y := \gamma x^\star + (1 - \gamma) x_k$ for $\gamma \in [0, 1]$ into (7), show that

  $$\varphi(x_{k+1}) - \varphi^\star \quad \leq \quad (1 - \gamma)\big(\varphi(x_k) - \varphi^\star\big) + \tfrac{\alpha D^2}{2} \gamma^2, \tag{8}$$

  for some $D > 0$, assuming that the initial sublevel set $\mathcal{F}_0 = \Big\{ x \in Q : \varphi(x) \leq \varphi(x_0) \Big\}$ is bounded.

- Minimizing the right hand side on (8) in $\gamma \in [0, 1]$, show the progress of each iterate in terms of the functional residual $\varphi_k := \varphi(x_k) - \varphi^\star$.

- Show that $\varphi_k = O(1/k)$.

Hint. Consider the case of unconstrained smooth minimization $\varphi(x) \equiv f_1(x)$ first, as the new analysis should recover the rate of the classical gradient descent.

Problem 4. Let $F(x,u) := u^{(1)} + \psi(x)$, where ψ is the indicator of a given convex set $Q \subseteq \mathbb{R}^n$:

$$\varphi(x) \quad = \quad \begin{cases} 0, & x \in Q \\ +\infty, & \text{otherwise.} \end{cases}$$

Show that each iteration (6) can be represented as follows:

$$x_{k+1} \quad = \quad \pi_Q(x_k - \tfrac{1}{\alpha}\nabla f(x_k)),$$

where $\pi_Q(x) := \arg\min\limits_{y \in Q} \|y - x\|$ is the projection onto set Q in the Euclidean norm.

Problem 5. Let $F(x,u) := u^{(1)} + \lambda\|x\|_1$, for $\lambda > 0$, where $\|x\|_1 := \sum_{i=1}^n |x^{(i)}|$. Thus we are interested to minimize a smooth convex function $f : \mathbb{R}^n \to \mathbb{R}$ with ℓ_1 regularizer:

$$\min_{x \in \mathbb{R}^n} \Bigl\{ \varphi(x) \quad \equiv \quad f(x) + \lambda\|x\|_1 \Bigr\}.$$

Provide an explicit formula for one step $x_k \mapsto x_{k+1}$ of method (6).

Problem 6. Let $F(x,u) := u^{(1)} + \frac{\lambda}{3}\|x\|_2^3$, for $\lambda > 0$. That is, the problem of minimizing a smooth convex function $f : \mathbb{R}^n \to \mathbb{R}$ with cubic regularization:

$$\min_{x \in \mathbb{R}^n} \Bigl\{ \varphi(x) \quad \equiv \quad f(x) + \tfrac{\lambda}{3}\|x\|_2^3 \Bigr\}.$$

Provide an explicit formula for one step $x_k \mapsto x_{k+1}$ of method (6).

3 On Accelerated Method

Problem 7. Consider the problem of unconstrained minimization:

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a strongly convex differentiable function with the Lipschitz continuous gradient. That is, for any $x, y \in \mathbb{R}^n$ it holds

$$\tfrac{\mu}{2}\|y - x\|^2 \quad \leq \quad f(y) - f(x) - \langle \nabla f(x), y - x \rangle \quad \leq \quad \tfrac{L}{2}\|y - x\|^2,$$

where $0 < \mu \leq L$ are two parameters, that we assume to be known. We aim to develop a modification of the fast gradient method on this problem class that achieves the optimal complexity of

$$O\Bigl(\sqrt{\tfrac{L}{\mu}} \log \tfrac{f(x_0) - f^\star}{\varepsilon} \Bigr) \tag{9}$$

first-order oracle calls to reach $f(x_k) - f^\star \leq \varepsilon$, without restarts.

We prove by induction the following inequality, for any $k \geq 0$:

$$\tfrac{\beta_0}{2}\|x - x_0\|^2 + A_k f(x) \quad \geq \quad \tfrac{\beta_k}{2}\|x - v_k\|^2 + A_k f(x_k), \qquad x \in \mathbb{R}^n, \tag{10}$$

for two sequences of points $\{x_k\}_{k \geq 0}$ and $\{v_k\}_{k \geq 0}$, starting from $x_0 = v_0$, and for two sequences $\{\beta_k\}_{k \geq 0}$, $\{A_k\}_{k \geq 0}$ of increasing non-negative coefficients.

We denote $a_{k+1} := A_{k+1} - A_k > 0$ and $\gamma_k := \frac{a_{k+1}}{A_{k+1}} \in (0,1]$. The indermediate points are denoted by

$$y_k \quad := \quad \gamma_k v_k + (1 - \gamma_k) x_k.$$

- Assuming that (10) holds for some $k \geq 0$, show that

$$\tfrac{\beta_0}{2}\|x - x_0\|^2 + A_{k+1} f(x) \quad \geq \quad m_k(x),$$

where

$$m_k(x) \quad := \quad \tfrac{\beta_k}{2}\|x - v_k\|^2 + a_{k+1}\Big[f(y_k) + \langle \nabla f(y_k), x - y_k \rangle + \tfrac{\mu}{2}\|x - y_k\|^2 \Big] + A_k f(x_k).$$

- Find the formula for $v_{k+1} = \arg\min\limits_x m_k(x)$ and $\beta_{k+1} > \beta_k$ such that

$$m_k(x) \quad \geq \quad \tfrac{\beta_{k+1}}{2}\|x - v_{k+1}\|^2 + m_k(v_{k+1}), \qquad x \in \mathbb{R}^n.$$

- Show that choosing $x_{k+1} = \gamma_k v_{k+1} + (1 - \gamma_k) x_k$, it is possible to ensure

$$m_k(v_{k+1}) \quad \geq \quad A_{k+1}\Big[f(y_k) + \langle \nabla f(y_k), x_{k+1} - y_k \rangle + \tfrac{\beta_k}{2\gamma_k^2 A_{k+1}}\|x_{k+1} - y_k\|^2 \Big].$$

- Show how to choose $A_{k+1} > A_k$ such that $m_k(v_{k+1}) \geq A_{k+1} f(x_{k+1})$ and thus prove (10).
- Obtain the optimal complexity (9) for the iterations of this algorithm.
- Write down the iterations of this algorithm using only one sequence of points. Compare it with the iterations of the Heavy Ball method:

$$z_{k+1} \quad = \quad z_k - \alpha \nabla f(z_k) - \beta(z_k - z_{k-1}).$$