Continuous Optimization: Problem Set 2

February 23, 2026

1. Recall the definitions of α-strong convexity and β-smoothness: for all x, y

$$\frac{\alpha}{2}\|y-x\|_2^2 \leq f(y) - f(x) - \langle \nabla f(x), y-x\rangle \leq \frac{\beta}{2}\|y-x\|_2^2.$$

   We will examine equivalent 1-st and 2-nd order definitions of these properties. For this question, assume f is twice continuously differentiable.

   1. Show if $f : \mathbb{R}^n \to \mathbb{R}$ is α-strongly convex or β-smooth, then the restriction $g(t) := f(x+tv)$ is $\alpha\|v\|_2^2$-strongly convex and $\beta\|v\|_2^2$-smooth for any $x, v \in \mathbb{R}^n$;
   2. Show α-strong convexity and β-smoothness are equivalent to the conditions $\alpha I \preceq \nabla^2 f(x)$ and $\nabla^2 f(x) \preceq \beta I$, respectively. (Here the notation $A \succeq B$ means $A - B \succeq 0$ or equivalently $\forall v : \langle v, Av\rangle \geq \langle v, Bv\rangle$)
   3. Show for an α-strongly convex function, we have distance bound on the optimizer

$$\|x - x^*\|_2 \lesssim \frac{\|\nabla f(x)\|_2}{\alpha}.$$

   4. Show β-smoothness is equivalent to Lipschitz-ness of the *gradient*:

$$\forall x, y : \quad \|\nabla f(y) - \nabla f(x)\|_2 \leq \beta\|y-x\|_2.$$

      (You can use the following fact without proof: $0 \preceq H \preceq \beta I \iff \forall x : \|Hx\|_2 \leq \beta\|x\|_2 \iff \forall x, y : |\langle x, Hy\rangle| \leq \beta\|x\|_2\|y\|_2$)

2. In the analysis of gradient descent for strongly convex and smooth functions we used the following crucial claim, known as the Polyak-Lojaseiwicz inequality (PL):

$$\|\nabla f(x)\|_2^2 \gtrsim \alpha(f(x) - f(x^*)),$$

   where x^* is the optimizer.

   1. Show that α-strong convexity implies this α-PL condition.
   2. Show that the analysis goes through if we assume only the PL condition as above, instead of strong convexity.

3. Find a function that satisfies $\alpha = 1$ PL condition but is not $\alpha = 1$-strongly convex. (Hint: consider convex quadratics.)
4. **Remark**: this condition even allows rigorous fast convergence analyses for certain non-convex functions!

3. (Q4 Ex4 from Daniel Dadush Course 2022) Assume $\mathsf{A}\mathsf{A}^\mathsf{T}$ is invertible. Give an explicit formula for $\nu^* := \max g(\nu) := \nu^\mathsf{T}\mathbf{b} - \|\mathsf{A}^\mathsf{T}\nu\|_2^2/2$ in terms of $\mathsf{A}\mathsf{A}^\mathsf{T}$ and $\mathbf{b}$. Prove that $\mathsf{A}^\mathsf{T}\nu^*$ is the optimal solution to the primal least-squares problem $\min \|\mathbf{x}\|_2^2/2, \mathsf{A}\mathbf{x} = \mathbf{b}$.

4. Recall our characterization of quadratic convex functions from last homework:

$$\langle x, Hx\rangle + \langle b, x\rangle + c$$

is convex iff $H \succeq 0$, i.e. H is symmetric and $\forall v \in \mathbb{R}^n : \langle v, Hv\rangle \geq 0$ (strictly convex iff $H \succ 0$, i.e. the inequality is strict for $v \neq 0$). In this question we will consider partitioned matrices

$$q(x,y) := \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle,$$

where A, C are symmetric.

1. Consider $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a convex function of variables $x \in \mathbb{R}^n, y \in \mathbb{R}^m$. Show $g(x) := \inf_y f(x,y)$ is also convex.
2. Let $q(x,y)$ be the quadratic defined above and assume $C \succ 0$. Show

$$\inf_y q(x,y) = \langle x, (A - BC^{-1}B^T)x\rangle.$$

(Hint: use first order optimality conditions for y).

3. Show q is strictly convex iff

$$H := \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succ 0$$

iff $C \succ 0$ and $A - BC^{-1}B^T$.

4. (Not a question, just a note): This last expression $A - BC^{-1}B^T$ is known as the *Schur complement* of H onto the x variables. It can be seen as the result of Gaussian elimination of the y variables and is extremely useful throughout math and computer science.

5. We use the following lemma: for convex ℓ and convex differentiable r, let $f := \ell + r$ with $x_* = \arg\min f$. Then

$$f(x) = \ell(x) + r(x) \geq f(x_*) + D_r(x, x_*) = \ell(x_*) + r(x_*) + D_r(x, x_*).$$

1. Prove the Bregman Pythagorean result: for strictly convex $r : \mathbb{R}^n \to \mathbb{R}$, let $\Pi_K(x) := \arg\min_{z \in K} D_r(z, x)$, then

$$\forall z \in K : \quad D_r(z, x) \geq D_r(\Pi(x), x) + D_r(z, \Pi(x)).$$

2. Let $f(z) := D_r(z, x)$, then $D_f = D_r$, i.e.

$$D_f(z, y) = D_r(z, x) - D_r(y, x) - \langle \nabla r(y) - \nabla r(x), z - y \rangle = D_r(z, y).$$

6. (Lap Chi Lau Course CS798, HW2 Q1) We analyze the multiplicative weights method for solving the following linear program:

$$\min \sum_{i \in [n]} x_i \qquad \text{s.t.} \qquad Ax \geq \vec{1}, \quad x \geq 0,$$

where $A \in \{0, 1\}^{m \times n}$, i.e. every entry of A is either 0 or 1. Design an algorithm to solving this linear program using the multiplicative weights update method (including the design of the oracle), and analyze its total running time.

7. In our proof of the lower bound for first-order methods for convex Lipschitz functions we made the following technical assumption: the initial point $x_0 = 0$ and the queries $x_t \in \text{span}\{x_0, g_0, ..., g_{t-1}\}$. In this question we show how to design an *adaptive* adversarial gradient oracle to remove this assumption.

    1. Show that we can remove the assumption $x_0 = 0$ by shifting the function based on the initial query.
    2. Show that for each query sequence $x_0, ...x_T$ with $T < n$ we can design a function of the form
    $$f_T(x) = \max_{i \in [n]} \langle g_i, x - x_0 \rangle$$
    such that $f(x_t) \leq 0$ for all $t \in [T]$.
    3. Show that we can choose functions $\{f_0, ..., f_T\}$ as above in a *consistent* way:
    $$\forall s \leq t < n : f_t(x_s) = f_s(x_s).$$
    Conclude that we can remove the technical assumption about queries.

8. (Lap Chi Lau Course CS798, HW2 Q1) Suppose you are given an 'accelerated' algorithm $\mathcal{A}$ for minimizing strongly convex and smooth functions: given f that is α-strongly convex and β-smooth, the algorithm outputs x_{alg} satisfying $f(x_{\text{alg}}) - f(x^*) \leq \varepsilon$ in

$$T \lesssim \sqrt{\frac{\beta}{\alpha}} \cdot \log\left(\frac{f(x_0) - f(x^*)}{\varepsilon}\right)$$

iteration, where $x^* = \arg\min_x f(x)$ (unknown), and x_0 is the initial point.

- You would like to use algorithm $\mathcal{A}$, but apply it to f that is *not strongly convex.* Prove that you can still use $\mathcal{A}$ as a black box to give an 'accelerated' algorithm with the following guarantee: for convex f that is β-smooth and minimizer $x^* = \arg\min_x f(x)$, find output x such that $f(x) - f(x^*) \leq \varepsilon$ in

$$T \lesssim \sqrt{\frac{\beta \|x_0 - x^*\|_2^2}{\varepsilon}} \cdot \log\left(\frac{f(x_0) - f(x^*)}{\varepsilon}\right)$$

  iterations, where again x_0 is the initial point. You are allowed to assume that you know the values $\|x_0 - x^*\|_2$ and $f(x_0) - f(x^*)$.

- Use this to derive the lower-bound for first-order methods for smooth convex functions via the lower-bound for smooth and strongly convex functions.