Sample midterm questions

March 11, 2026

1. (100 points) Let $X \subset \mathbb{R}^n$ be nonempty. We say that a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is *quasi-Fejér monotone* with respect to X if there exists a summable positive sequence $(\epsilon_k)_{k\in\mathbb{N}}$ such that

$$\|x_{k+1} - x\|^2 \leqslant \|x_k - x\|^2 + \epsilon_k, \quad \forall x \in X, \quad \forall k \in \mathbb{N}.$$

   (a) (15 points) Prove that if a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is quasi-Fejér monotone with respect to X, then the sequence is bounded, i.e., there exists $M > 0$ such that $\|x_k\| \leqslant M$ for all $k \in \mathbb{N}$.

   (b) (20 points) Prove that if a sequence $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ is quasi-Fejér monotone with respect to X and admits a subsequence converging to $x^* \in X$, then $x_k \to x^*$.

   Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function whose infimum is obtained at $X \subset \mathbb{R}^n$. Assume that f is L-Lipschitz, that is, $|f(x) - f(y)| \leqslant L\|x - y\|$ for all $x, y \in \mathbb{R}^n$. Denote by ∂f the convex subdifferential of f.

   (c) (15 points) Prove that $\|v\| \leqslant L$ for all $v \in \partial f(x)$, $x \in \mathbb{R}^n$.

   We next study the iterates of the subgradient method for minimizing f. Let $(x_k)_{k\in\mathbb{N}}$ be a sequence generated by $x_0 \in \mathbb{R}^n$ and $x_{k+1} \in x_k - \alpha_k \partial f(x_k)$, where $(\alpha_k)_{k\in\mathbb{N}}$ is a positive sequence.

   (d) (25 points) Derive a condition on $(\alpha_k)_{k\in\mathbb{N}}$ under which $(x_k)_{k\in\mathbb{N}}$ is quasi-Fejér monotone with respect to X.

   (e) (25 points) Derive a condition on $(\alpha_k)_{k\in\mathbb{N}}$ under which $(x_k)_{k\in\mathbb{N}}$ converges to a minimizer of f.

2. (100 points) Let f be a convex and L-smooth function. Let $g : \mathbb{R}^n \to \mathbb{R}$ to be a continuously differentiable and σ-strongly convex function. Denote $D_g(x, y)$ as the Bregman divergence induced by g.

   (a) (25 points) Show that

$$D_g(x, y) + D_g(y, z) = D_g(x, z) + \langle \nabla g(z) - \nabla g(y), x - y \rangle.$$

   (b) (25 points) Consider the update rule $\nabla g(x_{k+1}) = \nabla g(x_k) - \alpha_k \nabla f(x_k)$. Show that for all $y \in \mathbb{R}^n$,

$$f(y) - f(x_k) \geq \frac{1}{\alpha_k}(D_g(y, x_{k+1}) - D_g(x_k, x_{k+1}) - D_g(y, x_k)).$$

   (c) (30 points) By selecting $\alpha_k = \sigma/L$, it holds that

$$D_g(x_k, x_{k+1}) \leq \alpha_k(f(x_k) - f(x_{k+1})).$$

   (d) (20 points) Show that for all $T \in \mathbb{N}^+$,

$$\min_{k=0,\ldots,T-1} f(x_k) \leq f(x^*) + \frac{LD_g(x^*, x_0)}{\sigma T}.$$

3. (100 points) Let $f := \sum_{i=1}^n f_i$, where $f_i : \mathbb{R}^n \to \mathbb{R}$ is convex and L-smooth for all $i = 1, \ldots, n$. Assume f is μ-strongly convex. Consider the following algorithm:

$$x_{k+1} = x_k - \alpha_k \nabla f_{i_k}(x_k), \quad i_k \sim \mathrm{Unif}[n], \quad \forall\, k \in \mathbb{N}.$$

   Define $\sigma_f^* := \inf_{x^* \in \arg\min f} \mathbb{E}[\|\nabla f_i(x^*) - \nabla f(x^*)\|^2]$.

(a) (25 points) Show that for every $x^* \in \arg\min f$, we have $\sigma_f^* = \mathbb{E}[\|\nabla f_i(x^*) - \nabla f(x^*)\|^2]$.

(b) (20 points) Show that

$$\mathbb{E}[\|\nabla f_i(x)\|^2] \leq 4L(f(x) - \inf f) + 2\sigma_f^*.$$

(c) (30 points) With step size $\alpha_k = \alpha \in (0, 1/(2L)]$, show that for all $k \geq 0$,

$$\mathbb{E}[\|x_k - x^*\|^2] \leq (1 - \alpha\mu)^k \|x_0 - x^*\|^2 + \frac{2\alpha}{\mu}\sigma_f^*.$$

(d) (25 points) Show that for every $\epsilon > 0$, we can guarantee that $\mathbb{E}[\|x_T - x^*\|^2] \leq \epsilon$ provided that

$$\alpha := \min\left\{\frac{\mu\epsilon}{4\sigma_f^*}, \frac{1}{2L}\right\}, \quad \text{and} \quad T \geq \max\left\{\frac{4\sigma_f^*}{\epsilon\mu^2}, \frac{2L}{\mu}\right\} \log\left(\frac{2\|x_0 - x^*\|^2}{\epsilon}\right).$$

Solution 1:

(a) Let $x \in X$. $\|x_k - x\|^2 \leqslant \|x_0 - x\|^2 + \sum_{k=0}^{\infty} \epsilon_k < \infty$.

(b) Let $E_k := \sum_{i=k}^{\infty} \epsilon_i$. We have that $E_k \to 0$. We know that $\|x_{\varphi_k} - x^*\| \to 0$ for some subsequence of $(x_k)_{k\in\mathbb{N}}$. By quasi-Fejér monotonicity, $\|x_{k'} - x^*\|^2 \leqslant \|x_{\varphi_k} - x^*\|^2 + E_{\varphi_k} \to 0$ for all $k' \geqslant \varphi_k$.

(c)

$$L\|v\| = L\|(x+v) - x\| \geqslant |f(x+v) - f(x)| \geqslant \langle v, v\rangle = \|v\|^2.$$

(d) Let $x^* \in X$, we have that

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - \alpha_k v_k - x^*\|^2 && \text{(1a)}\\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|v_k\|^2 - 2\langle x_k - x^*, \alpha_k v_k\rangle && \text{(1b)}\\
&\leqslant \|x_k - x^*\|^2 + \alpha_k^2 \|v_k\|^2 - 2\alpha_k (f(x_k) - f(x^*)) && \text{(1c)}\\
&\leqslant \|x_k - x^*\|^2 + \alpha_k^2 L^2. && \text{(1d)}
\end{aligned}$$

Thus a condition is $(\alpha_k)_{k\in\mathbb{N}}$ is square summable.

(e) The condition is

$$\sum_{k=0}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \tag{2}$$

Telescoping (1c) yields

$$2\sum_{k=0}^{\infty} \alpha_k (f(x_k) - f(x^*)) \leqslant \|x_0 - x^*\|^2 + \sum_{k=0}^{\infty} \alpha_k^2 \|v_k\|^2 \leqslant \|x_0 - x^*\|^2 + L^2 \sum_{k=0}^{\infty} \alpha_k^2$$

Thus $\liminf_{k\to\infty} f(x_k) = f(x^*)$. By question 1, $(x_k)_{k\in\mathbb{N}}$ is bounded, thus it exists a subsequence converging to some point in X (again denoted by x^*). By question 2, $x_k \to x^*$.

Solution 2:

(a) The Bregman divergence induced by g is defined as:

$$D_g(x, y) = g(x) - g(y) - \langle \nabla g(y), x - y\rangle.$$

Expanding the terms, we obtain:

$$D_g(x, y) + D_g(y, z) = g(x) - g(z) - \langle \nabla g(y), x - y\rangle - \langle \nabla g(z), y - z\rangle.$$

Rewriting the Bregman divergence $D_g(x, z)$:

$$D_g(x, z) = g(x) - g(z) - \langle \nabla g(z), x - z\rangle.$$

Using $x - z = (x - y) + (y - z)$, we get:

$$\langle \nabla g(z), x - z\rangle = \langle \nabla g(z), x - y\rangle + \langle \nabla g(z), y - z\rangle.$$

Substituting this identity:

$$D_g(x, y) + D_g(y, z) = D_g(x, z) + \langle \nabla g(z) - \nabla g(y), x - y\rangle.$$

(b) From the convexity of f:

$$f(y) \geq f(x_k) + \langle \nabla f(x_k), y - x_k\rangle.$$

Using the update equation:

$$\nabla g(x_{k+1}) - \nabla g(x_k) = -\alpha_k \nabla f(x_k),$$

we substitute:

$$\langle \nabla f(x_k), y - x_k \rangle = -\frac{1}{\alpha_k}\langle \nabla g(x_{k+1}) - \nabla g(x_k), y - x_k \rangle.$$

By applying the identity from part (a):

$$\langle \nabla g(x_{k+1}) - \nabla g(x_k), y - x_k \rangle = D_g(y, x_k) + D_g(x_k, x_{k+1}) - D_g(y, x_{k+1}),$$

we obtain:

$$f(y) - f(x_k) \geq \frac{1}{\alpha_k}(D_g(y, x_{k+1}) - D_g(x_k, x_{k+1}) - D_g(y, x_k)).$$

(c) Since f is L-smooth and g is σ-strongly convex, we have

$$f(x_{k+1}) - f(x_k) \leq \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\left\|x_{k+1} - x_k\right\|^2.$$

and $D_g(x_{k+1}, x_k) \geq \frac{\sigma}{2}\|x_{k+1} - x_k\|^2$. Thus it follows that

$$\begin{aligned} D_g(x_k, x_{k+1}) &= \alpha_k \langle \nabla f(x_k), x_k - x_{k+1} \rangle - D_g(x_{k+1}, x_k) \\ &\leq \alpha_k(f(x_k) - f(x_{k+1}) + \frac{L}{2}\|x_{k+1} - x_k\|^2) - \frac{\sigma}{2}\|x_{k+1} - x_k\|^2 \\ &= \alpha_k(f(x_k) - f(x_{k+1})) \end{aligned}$$

by selecting $\alpha_k = \sigma/L$.

(d) By part (b) and (c), it gives that

$$f(x^*) - f(x_{k+1}) \geq \frac{1}{\alpha_k}(D_g(x^*, x_{k+1}) - D_g(x^*, x_k)).$$

Summing over both sides from 0 to $T-1$, we obtain that

$$\frac{1}{T}\sum_{k=1}^{T} f(x_k) \leq f(x^*) + \frac{D_g(x^*, x_0)}{\alpha_k T} = f(x^*) + \frac{LD_g(x^*, x_0)}{\sigma T}.$$

Solution 3:

(a) Let $x^*, x' \in \arg\min f$. Using refined gradient inequality in Week 1 and taking expectation, the inner product term vanishes and we obtain

$$\frac{1}{2L}\mathbb{E}[\|\nabla f_i(x') - \nabla f_i(x^*)\|^2] \leq f(x') - \inf f = 0$$

This means that for every $i = 1, \dots, n$, we have $\|\nabla f_i(x') - \nabla f_i(x^*)\| = 0$. In other words, $\nabla f_i(x') = \nabla f_i(x^*)$, and thus $\mathbb{E}[\|\nabla f_i(x^*) - \nabla f(x^*)\|^2] = \mathbb{E}[\|\nabla f_i(x') - \nabla f(x')\|^2]$. Since x^*, x' are chosen arbitrarily, the desired result follows.

(b) Start by writing

$$\|\nabla f_i(x)\|^2 \leq 2\|\nabla f_i(x) - \nabla f_i(x^*)\|^2 + 2\|\nabla f_i(x^*)\|^2$$

Applying refined gradient inequality in Week 1, taking the expectation over the above inequality (similar to part (a)), and using part (a) gives the result.

(c) Let $x^* \in \arg\min f$. Using the update of algorithm and expanding the squares we have

$$\begin{aligned} \|x^{t+1} - x^*\|^2 &= \|x^k - x^* - \alpha \nabla f_i(x^k)\|^2 \\ &= \|x^k - x^*\|^2 - 2\alpha\langle x^k - x^*, \nabla f_i(x^k)\rangle + \alpha^2\|\nabla f_i(x^k)\|^2. \end{aligned}$$

Taking expectation conditioned on x^k we obtain

$$\begin{aligned}\mathbb{E}_k[\|x^{t+1} - x^*\|^2] &= \|x^k - x^*\|^2 - 2\alpha\langle x^k - x^*, \nabla f(x^k)\rangle + \alpha^2 \mathbb{E}_k[\|\nabla f_i(x^k)\|^2] \\ &\leq (1 - \alpha\mu)\|x^k - x^*\|^2 - 2\alpha[f(x^k) - f(x^*)] + \alpha^2 \mathbb{E}_k[\|\nabla f_i(x^k)\|^2].\end{aligned}$$

Taking expectations again and using part (b) gives

$$\begin{aligned}\mathbb{E}[\|x^{t+1} - x^*\|^2] &\leq (1 - \alpha\mu)\mathbb{E}\|x^k - x^*\|^2 + 2\alpha^2\sigma_f^* + 2\alpha(2\alpha L - 1)\mathbb{E}[f(x^k) - f(x^*)] \\ &\leq (1 - \alpha\mu)\mathbb{E}[\|x^k - x^*\|^2] + 2\alpha^2\sigma_f^*\end{aligned}$$

where we used in the last inequality that $2\alpha L \leq 1$ since $\alpha \leq \frac{1}{2L}$. Recursively applying the above and summing up the resulting geometric series gives

$$\begin{aligned}\mathbb{E}\|x^k - x^*\|^2 &\leq (1 - \alpha\mu)^k \|x_0 - x^*\|^2 + 2\sum_{j=0}^{k-1}(1 - \alpha\mu)^j \alpha^2 \sigma_f^* \\ &\leq (1 - \alpha\mu)^k \|x_0 - x^*\|^2 + \frac{2\alpha\sigma_f^*}{\mu}.\end{aligned}$$

(d) It suffices to require

$$(1 - \alpha\mu)^k \|x_0 - x^*\|^2 \leq \frac{\epsilon}{2}, \quad \frac{2\alpha\sigma_f^*}{\mu} \leq \frac{\epsilon}{2}.$$

This is equivalent to

$$k \log(1 - \alpha\mu) \leq \log \frac{\epsilon}{2\|x_0 - x^*\|^2}, \quad \alpha \leq \frac{\mu\epsilon}{4\sigma_f^*}.$$

The first term can be further simplified to

$$k \geq \frac{\log \frac{\epsilon}{2\|x_0 - x^*\|^2}}{\log(1 - \alpha\mu)} \implies k \geq \left(\log \frac{1}{1 - \alpha\mu}\right)^{-1} \log\left(\frac{2\|x_0 - x^*\|^2}{\epsilon}\right).$$

Since $\log \frac{1}{1-\alpha\mu} \geq \alpha\mu$, it suffices to have

$$k \geq \frac{1}{\alpha\mu} \log\left(\frac{2\|x_0 - x^*\|^2}{\epsilon}\right)$$

which is exactly the desired complexity.