Name: ____________ Student ID: ____________ Scores: ____________

AIE6003 Homework 1
Instructor: Xiaopeng Li **Deadline**: Feb 6

Notice: If you are trying to invoke a result outside of this course, then you should put a reference to it. The questions with 1 star are easy, with 2 stars are medium, and with 3 stars are hard. Easy and medium level of questions are likely to occur in the exams, while hard questions are not likely to occur in the exams.

1*. (15 points) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex ℓ-smooth function. Assume an optimal solution $x^* \in \arg\min f$ exists and denote $f^* := f(x^*)$. Consider the following algorithm with initial points $x_{-1} = x_0$:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) + \beta_k (x_k - x_{k-1}), \quad k \in \mathbb{N},$$

Take $\beta_k = \frac{k}{k+2}$ and $\alpha_k = \frac{\alpha_0}{k+2}$ for $k \in \mathbb{N}$, where $\alpha_0 \in (0, 1/\ell]$.

(a) (3 points) Let $p_k = k(x_k - x_{k-1})$ for all $k \in \mathbb{N}$. Prove that

$$x_{k+1} + p_{k+1} = x_k + p_k - \alpha_0 \nabla f(x_k).$$

(b) (10 points) For all $k \in \mathbb{N}$, show that the sequence

$$\left(2\alpha_0 k (f(x_{k-1}) - f^*) + \|x_k + p_k - x^*\|^2\right)_{k \in \mathbb{N}}$$

is decreasing.

(c) (2 points) Conclude that for all $T \in \mathbb{N}$,

$$f(x_T) - f^* \le \frac{\|x_0 - x^*\|^2}{2\alpha_0 (T+1)}.$$

2**. (15 points) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and ℓ-smooth. Assume an optimal solution $x^\star \in \arg\min f$ exists and denote $f^\star := f(x^\star)$. Consider gradient descent with backtrack line search, with parameters $s > 0$, $\beta \in (0,1)$, $\sigma \in (0,1)$: given x^0, for $k = 0, 1, 2, \ldots$ set $g_k := \nabla f(x^k)$ and $d_k := -g_k$. Choose the smallest integer $m_k \ge 0$ such that, with $\alpha_k := s\beta^{m_k}$,

$$f(x^k + \alpha_k d_k) \le f(x^k) + \sigma \alpha_k g_k^\top d_k, \tag{1}$$

and then update $x^{k+1} := x^k + \alpha_k d_k$.

(a) (6 points) Prove that the backtracking procedure terminates for every k, and that the accepted step sizes satisfy the uniform lower bound

$$\alpha_k \ge \underline{\alpha} := \min\left\{ s, \frac{2\beta(1-\sigma)}{\ell} \right\}, \quad \forall k \ge 0.$$

(b) (5 points) Prove that for every $k \ge 0$,

$$f(x^k) - f^\star \le \frac{1}{2\alpha_k}\left(\|x^k - x^\star\|^2 - \|x^{k+1} - x^\star\|^2\right) + \frac{1}{2\sigma}(f(x^k) - f(x^{k+1})).$$

(c) (4 points) Using parts (a)-(b), prove that there exists an explicit constant $c > 0$ such that

$$f(x^k) - f^\star \le \frac{c}{k} \quad \forall k \ge 1,$$

with

$$c = \frac{\|x^0 - x^\star\|^2}{2\underline{\alpha}} + \frac{f(x^0) - f^\star}{2\sigma}.$$

3***. (20 points) Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite (SPD) and consider

$$\min_{x\in\mathbb{R}^n} \quad f(x) := \frac{1}{2}x^\top Ax - b^\top x,$$

whose unique minimizer is $x^\star = A^{-1}b$. Given $x_0 \in \mathbb{R}^n$, define the residual $r_k := b - Ax_k$, the error $e_k := x_k - x^\star$, and the Krylov subspace

$$\mathcal{K}_k(A, r_0) := \operatorname{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}. \tag{2}$$

The Conjugate Gradient (CG) iterations (assume $r_k \neq 0$) are:

$$\begin{aligned} r_0 &= b - Ax_0, \quad p_0 = r_0, \quad \alpha_k = \frac{r_k^\top r_k}{p_k^\top Ap_k}, \quad x_{k+1} = x_k + \alpha_k p_k, \\ r_{k+1} &= r_k - \alpha_k Ap_k, \quad \beta_{k+1} = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k}, \quad p_{k+1} = r_{k+1} + \beta_{k+1}p_k. \end{aligned} \tag{3}$$

(a) (5 points) Let $\phi_k(\alpha) := f(x_k + \alpha p_k)$. Show that the minimizer of ϕ_k satisfies

$$\alpha_k = \frac{p_k^\top r_k}{p_k^\top Ap_k}, \quad \text{and hence} \quad p_k^\top r_{k+1} = 0. \tag{4}$$

(b) (5 points) Prove that $x_k \in x_0 + \mathcal{K}_k(A, r_0)$ and $r_k \in \mathcal{K}_{k+1}(A, r_0)$ for all $k \geq 0$.

(c) (5 points) Show that for all $k \geq 1$, $p_{k-1}^\top r_k = 0$. Conclude that for all $k \geq 0$, $p_k^\top r_k = r_k^\top r_k$, and therefore the line-search formula in (4) is consistent with the CG step size in (3).

(d) (5 points) You may use the following standard CG property without proof: the search directions are A-conjugate, i.e., $p_i^\top Ap_j = 0$, for all $i \neq j$. Using this fact together with parts (a)-(b), prove that for all $k \geq 1$, $r_k \perp \mathcal{K}_k(A, r_0)$, and conclude that $r_i^\top r_j = 0$ for all $i \neq j$.

Page 2