Optimization for AI **Name:** ______________________
Fall 2020
Midterm **ID:** ______________________

This exam contains 2 pages (including this cover page) and 7 questions.

---

1. (10 points) $f(x) = x + x^2 + e^x$ is α-strongly convex and β-smooth on the interval $[-2, 2]$. What is α? What is β?

2. (10 points) Let f be a convex function. Explain why the following inequality holds:
$$\frac{1}{T}\sum_{t=1}^{T} f(x_t) - f(x^\star) \geq f(\bar{x}) - f(x^\star) \quad \text{where} \quad \bar{x} = \frac{\sum_{t=1}^{T} x_t}{T}.$$

3. (15 points) Consider a constrained optimization problem:
$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0 \quad \text{for all} \quad 1 \leq i \leq m \\ & g_j(x) \leq 0 \quad \text{for all} \quad 1 \leq j \leq r. \end{aligned}$$
We would like to solve the constrained optimization problem using Lagrange dual problem. Explain how to define the Lagrange dual problem and how to find the solution.

4. (15 points) Let f be α-strongly convex and β smooth. Show that
$$\frac{1}{2\beta}\|\nabla f(x)\|^2 \leq f(x) - f(x^\star) \leq \frac{1}{2\alpha}\|\nabla f(x)\|$$
where $x^\star$ is the minimum point.

5. (15 points) This is Mirror descent.
Mirror descent:
    1. x_t is mapped to $\nabla\Phi(x_t)$
    2. Compute $\nabla\Phi(x_t) - \gamma\nabla f(x_t)$
    3. Find y_{t+1} such that $\nabla\Phi(y_{t+1}) = \nabla\Phi(x_t) - \gamma\nabla f(x_t)$
    4. Projection. $x_{t+1} = \Pi_{\mathcal{X}}^{\Phi}(y_{t+1}) = \arg\min_{x\in\mathcal{X}} D_\Phi(x, y_{t+1})$

    Find Φ such that mirror descent is exactly equivalent to projected (sub)gradient descent.

6. (15 points) Let f_i be a β-smooth convex function for all i and $f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$. In SVRG, for $s = 1, \ldots,$ we update
$$\begin{aligned} x_1^{(s)} &= y^{(s)} \\ x_{t+1}^{(s)} &= x_t^{(s)} - \gamma\left(\nabla f_{i_t^{(s)}}(x_t^{(s)}) - \nabla f_{i_t^{(s)}}(y^{(s)}) + \nabla f(y^{(s)})\right) \quad t = 1, \ldots, k, \end{aligned}$$
where $i_t^{(s)}$ is drawn uniformly at random. Then, update $y^{(s+1)} = \frac{1}{k}\sum_{t=1}^{k} x_t^{(s)}$. Explain why SVRG can reduce the variance.

7. (20 points) Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex and differentiable, $\mathcal{X} \subset \mathbb{R}^d$ closed and convex, $x^\star$ a minimizer of f over $\mathcal{X}$. Suppose that $\|x - x'\| \leq R$ for all $x, x' \in \mathcal{X}$, and stochastic gradient $\tilde{g}(x)$ such that $\mathbb{E}[\tilde{g}(x)] = \nabla f(x)$ satisfies $\|\tilde{g}(x)\| \leq B$ for all $x \in \mathcal{X}$. Show that with decreasing step size $\gamma_t = \frac{R}{B\sqrt{t}}$ (i.e. $y_t = x_{t-1} - \gamma_t \tilde{g}(x_{t-1})$), the projected gradient descent has

$$\frac{1}{T}\mathbb{E}\left(\sum_{t=0}^{T-1} f(x_t) - f(x^\star)\right) \leq \frac{3}{2}\frac{RB}{\sqrt{T}}.$$