CS395T: Continuous Algorithms
Homework II

Kevin Tian

Due date: February 15, 2024, start of class (3:30 PM).

Please list all collaborators on the first page of your solutions. Unless we have discussed and I have specified otherwise, homework is not accepted if it is not turned in by hand at the start of class, or turned in electronically on Canvas by then. Send me an email to discuss any exceptions.

1 Problem 1

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable, convex function, and for all $\lambda > 0$, let

$$x^\star_\lambda := \mathrm{argmin}_{x \in \mathbb{R}^d} f(x) + \frac{\lambda}{2} \|x\|_2^2 .$$

(i) Prove that for any $\lambda' > \lambda > 0$, we have $\|x^\star_\lambda\|_2 \ge \|x^\star_{\lambda'}\|_2$.

(ii) Suppose $\|\nabla f(\mathbb{0}_d)\|_2 \le L$ and let $R > 0$. Give a λ_R such that for any $\lambda \ge \lambda_R$, $\|x^\star_\lambda\|_2 \le R$.

2 Problem 2

Let $\mathbf{M} \in \mathbb{S}^{d \times d}_{\succ \mathbf{0}}$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^d$.

(i) Prove that if $\mathbf{M}[v, v] \le \|v\|^2$ for all $v \in \mathbb{R}^d$, then $\|\mathbf{M}v\|_* \le \|v\|$ for all $v \in \mathbb{R}^d$.[1]

(ii) Prove that $\mathbf{M}[v, v] \le \|v\|^2$ for all $v \in \mathbb{R}^d$ iff $\mathbf{M}^{-1}[v, v] \ge \|v\|_*^2$ for all $v \in \mathbb{R}^d$.[2]

3 Problem 3

Let $\mathcal{X} \subseteq \mathbb{R}^d$, let $f : \mathcal{X} \to \mathbb{R}$ be convex, differentiable, and L-Lipschitz in $\|\cdot\|$, and let $\psi : \mathcal{X} \to \mathbb{R}$ be convex and admit subgradients everywhere in $\mathcal{X}$. Let $F(x) = f(x) + \psi(x)$.[3] Finally, let $\varphi : \mathcal{X} \to \mathbb{R}$ be 1-strongly convex in $\|\cdot\|$ and of Legendre type. Consider iterating the update

$$x_{t+1} \leftarrow \mathrm{argmin}_{x \in \mathcal{X}} \left\{ \langle \eta \nabla f(x_t), x \rangle + \eta \psi(x) + D_\varphi(x \| x_t) \right\}, \text{ for } 0 \le t < T, \ \eta > 0.$$

Give an initialization strategy for choosing x_0, and prove that following your initialization strategy,

$$F(\bar{x}) - F(x^\star) \le \frac{\Theta}{\eta T} + \frac{\eta L^2}{2},$$

where $\bar{x} = \frac{1}{T} \sum_{0 \le t < T} x_t$, $x^\star \in \mathrm{argmin}_{x \in \mathcal{X}} F(x)$, and $\Theta \ge \sup_{x_0 \in \mathcal{X}} D_\varphi(x^\star \| x_0)$.

[1] This justifies that for general norms, the second claim in Lemma 14, Part II is true. Note that the last part of the proof of Lemma 6, Part II implicitly used this claim in the ℓ_2 case.

[2] Combined with $\nabla^2 f^*(\nabla f(x)) = (\nabla^2 f(x))^{-1}$, shown in class when f, f^* are twice-differentiable, this gives a simple proof of smoothness-strong convexity duality for twice-differentiable convex functions in general norms.

[3] The purpose of this problem is to explore how to generalize the mirror descent framework in Theorem 2, Part III, to handle composite objectives just as gradient descent can, as shown by Section 5.2, Part II.

4 Problem 4

(i) Prove that the following $2d \times 2d$ matrix is positive semidefinite, for any $a \in \mathbb{R}$ and $\mathbf{M} \in \mathbb{S}_{\succeq \mathbf{0}}^{d \times d}$:

$$\begin{pmatrix} a^2\mathbf{M} & a\mathbf{M} \\ a\mathbf{M} & \mathbf{M} \end{pmatrix}.$$

(ii) Prove that for any $\{a_i\}_{i\in[n]} \subset \mathbb{R}$, and $\{\mathbf{M}_i\}_{i\in[n]} \subset \mathbb{S}_{\succeq \mathbf{0}}^{d \times d}$ satisfying $\sum_{i\in[n]} \mathbf{M}_i \preceq \mathbf{I}_d$, we have[4]

$$\left(\sum_{i\in[n]} a_i \mathbf{M}_i\right)^2 \preceq \sum_{i\in[n]} a_i^2 \mathbf{M}_i.$$

5 Problem 5

Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable, and let $x^\star := \mathrm{argmin}_{x\in\mathbb{B}(R)} f(x)$ exist and be unique. Suppose that $\mathcal{A}$ is a randomized algorithm which produces a point $x \in \mathbb{B}(R)$ satisfying

$$\mathbb{E}\left[f(x) - f(x^\star)\right] \le \epsilon.$$

Let $g : \mathbb{B}(R) \to \mathbb{R}^d$ be a randomized estimator satisfying

$$\mathbb{E}g(x) = \nabla f(x),\ \mathbb{E}\left\|g(x)\right\|_2^2 \le L^2,\ \text{for all } x \in \mathbb{B}(R).$$

(i) Let $\alpha > 0$, $\delta \in (0, 1)$. Given $x, x' \in \mathbb{B}(R)$, design an algorithm which calls g $O(\frac{L^2R^2}{\alpha^2}\log\frac{1}{\delta})$ times, and estimates $f(x) - f(x')$ to additive error α with probability $\ge 1 - \delta$.[5]

(ii) Give an algorithm which calls $\mathcal{A}$ $O(\log\frac{1}{\delta})$ times and g $O(\frac{L^2R^2}{\epsilon^2}\mathrm{polylog}(\frac{1}{\delta}))$ times,[6] and produces x such that with probability $\ge 1 - \delta$,

$$f(x) - f(x^\star) \le 3\epsilon.$$

[4]This can be viewed as a matrix extension of the Cauchy-Schwarz inequality.
[5]It may be helpful to first write $f(x) - f(x')$ as an integral.
[6]I am curious to see how low people can get the polylog :)