MATH 329 Nonlinear Optimization
HW1. GD and Convexity

Yujun Kim

1. $f(\theta) = -\log l(\theta))$

$$= -\log\left(\prod_{i=1}^{m} \sigma(\langle x_i, x\rangle + b)^{y_i} \sigma(-\langle x_i, x\rangle - b)^{1-y_i}\right)$$

$$= -\log\left(\prod_{i=1}^{m} \sigma(\langle \tilde{x}_i, \theta\rangle)^{y_i} \sigma(-\langle \tilde{x}_i, \theta\rangle)^{1-y_i}\right)$$

$$= \sum_{i=1}^{m} y_i \log \frac{1}{\sigma(\langle \tilde{x}_i, \theta\rangle)} + \Sigma (1-y_i) \log \frac{1}{\sigma(-\langle \tilde{x}_i, \theta\rangle)}$$

$$\frac{1}{\sigma(\langle \tilde{x}_i, \theta\rangle)} = 1+e^{-\langle \tilde{x}_i, \theta\rangle}, \quad \frac{1}{\sigma(-\langle \tilde{x}_i, \theta\rangle)} = 1+e^{-\langle \tilde{x}_i, \theta\rangle}.$$

Hence, the statement holds.

2. Let $\varphi(z) = \log(1+e^z)$. Then, $\varphi'(z) = \frac{e^z}{1+e^z} = 1 - \frac{1}{1+e^z}$.
As $z \mapsto 1+e^z$ is inc, $w \mapsto 1 - 1/w$ is increasing, it follows that φ' is inc.
Thus, φ is convex.

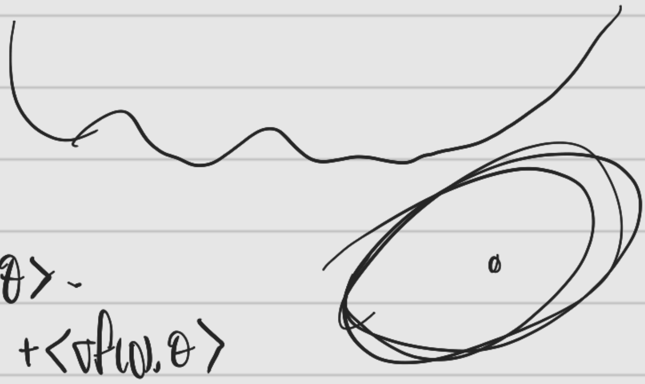
3. Convexity is preserved under addition, composition with linear map, and scala multiplication by nonnegative
For any $x \in \mathbb{R}^{n \times n}$, $\theta \mapsto \langle \tilde{x}, \theta\rangle$ is linear
$\theta \mapsto \langle -\tilde{x}, \theta\rangle$ is linear.

With 2, we have f: convex.

4. By conv, $f(\theta) \geq f(\phi) + \langle \nabla f(\phi), \theta - \phi\rangle$ $\quad f(\phi) \geq f(\theta) + \langle \nabla f(\theta), \phi - \theta\rangle$
$\Rightarrow \langle \nabla f(\phi) - \nabla f(\theta), \phi - \theta\rangle \geq 0.$
Thus, $\langle \nabla f_\lambda(\theta) - \nabla f_\lambda(\phi), \theta - \phi\rangle = \langle \nabla f(\theta) - \nabla f(\phi) + \lambda(\theta - \phi), \theta - \phi\rangle$
$\geq \lambda \|\theta - \phi\|^2.$

5. Suppose f_λ attains min at

$|\langle \tilde{x}_i, \theta \rangle| \le \|\tilde{x}_i\| \|\theta\|$.



$$\langle \nabla f_\lambda(\theta), \theta \rangle = \langle \nabla f(\theta) + \lambda\theta, \theta \rangle$$
$$= \langle \nabla f(\theta), \theta \rangle + \lambda \langle \theta, \theta \rangle$$
$$= \langle \nabla f(\theta) - \nabla f(0), \theta - 0 \rangle + \langle \nabla f(0), \theta \rangle + \lambda \|\theta\|^2$$
$$\ge \lambda \|\theta\|^2 - \|\nabla f(0)\| \|\theta\| = \lambda \left(\|\theta\| - \frac{\|\nabla f(0)\|}{2\lambda} \right)^2 -$$

If $\|\theta\| \ge \frac{\|\nabla f(0)\|}{2\lambda} = r_\lambda$, then $\exists \phi$ with $\|r_\lambda\|$ s.t. $f_\lambda(\phi) \le f_\lambda(\theta)$

Thus, $\min\limits_{\|\theta\| \le r_\lambda} f_\lambda(\theta)$ is global minima. ($\exists$ by compactness of $\{\theta \mid \|\theta\| \le r_\lambda\}$.

Uniqueness and $\nabla f_\lambda(\theta) = 0$ at minima is obvious.

6. $\nabla f_\lambda(\theta) = \nabla f(\theta) + \lambda\theta$.

$$\theta \overset{\varphi}{\longmapsto} \log(1 + e^{\langle \tilde{x}, \theta \rangle}) \Rightarrow \nabla \varphi(\theta) = \frac{e^{\langle \tilde{x}, \theta \rangle}}{1 + e^{\langle \tilde{x}, \theta \rangle}} \cdot \tilde{x}.$$

$$\Rightarrow \nabla f(\theta) = \sum_{i=1}^{m} y_i \frac{e^{-\langle \tilde{x}_i, \theta \rangle}}{1 + e^{-\langle \tilde{x}_i, \theta \rangle}} (-\tilde{x}_i) + \sum_{i=1}^{m} (1 - y_i) \frac{e^{\langle \tilde{x}_i, \theta \rangle}}{1 + e^{\langle \tilde{x}_i, \theta \rangle}} x_i$$