Yihang Chen.

1. (a) suppose $g(u) = \log(1+e^{-u})$, we have.

$$g'(u) = -\frac{e^{-u}}{1+e^{-u}} \qquad g''(u) = \frac{e^{-u}(1+e^{-u}) + e^{-u}(-e^{-u})}{(1+e^{-u})^2} = \frac{e^{-u}}{(1+e^{-u})^2} > 0$$

$\Rightarrow$ g is convex.

$$\Rightarrow \frac{\partial^2 f}{\partial \vec{x}^2} = \frac{\partial^2}{\partial \vec{x}^2} \sum_{i=1}^{n} g(b_i a_i^T \vec{x}) = \cancel{\frac{\partial^2}{\partial \vec{x}^2}} \sum_{i=1}^{n} g''(b_i a_i^T \vec{x}) b_i^2 a_i a_i^T$$

which is positive semidefinsee $\Rightarrow$ f is convex

~~Since f is defined on $x \in \mathbb{R}^p$, and bounded below~~

The minimum does not always exist. eg. $f(x) = \log(1+e^{-x})$. the minimum is 0 and reached only by $x = +\infty$.

(b) ~~infima is~~ $f: \Omega \mapsto \mathbb{R}$. (i) a_{min} is the minimum $\Leftrightarrow$ $a_{min} \in \Omega$, $\forall a' \in \Omega$, $f(a') \geq f(a_{min})$

(ii) a_{inf} is infinima $\Leftrightarrow$ $\exists\, a^i \in \Omega$, $i \geq 1$. $a^i \to a_{inf}$, $\forall a' \in \Omega$, $f(a') \geq f(a_{inf})$.

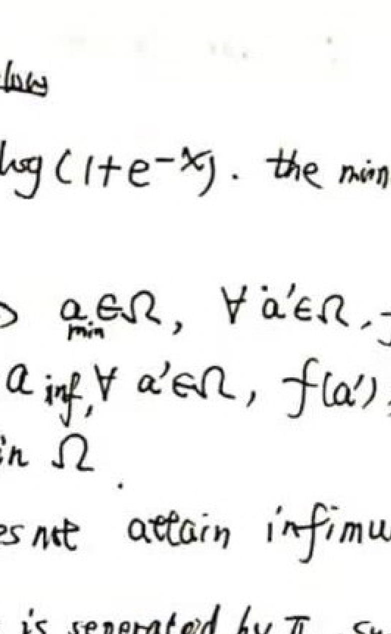
The difference is that $a_{min} \in \Omega$, but a_{inf} might not in Ω.

$f(x) = \log(1+e^{-x})$. $x \to +\infty$, $f(x) \to 0$, does not attain infimum.

(c). Say plane π is orthogonal to x^0, then $\{a_i\}$ is seperated by π, such that on one side it's 1, on the other side is -1.

~~say $f_x(x)$~~ say $g(\alpha) = f(\alpha x_0) = \sum_{i=1}^{n} \log(1+e^{-b_i \alpha a_i^T x_0})$

$\to 0$ as $\alpha \to +\infty$. so the minimum cannot be attained.



(d) $\nabla f_\mu(x) = \sum_{i=1}^{n} -b_i a_i \sigma(-b_i a_i^T x) + \mu x$ by the chain rule.

(e). $\nabla^2 f_\mu(x) = \sum_{i=1}^{n} b_i^2 \sigma''(-b_i a_i^T x) a_i a_i^T + \mu I$

$$= \sum_{i=1}^{n} \sigma(-b_i a_i^T x)(1 - \sigma(-b_i a_i^T x)) a_i a_i^T + \mu I.$$

since $\sigma''(t) = \dfrac{-e^{-t}(1+e^{-t})^2 + e^{-2t}(1+e^{-t})}{(1+e^{-t})^4} = \sigma(-t)(1-\sigma(t))$

and $b_i^2 = 1$

(f) since $f_\mu - \frac{\mu}{2}\|x\|^2$ is convex, we have.

$$f_\mu(x) - \frac{\mu}{2}\|x\|^2 \geq f_\mu(y) - \frac{\mu}{2}\|y\|^2 + \nabla f_\mu(y)^T (y-x) + y^T(y-x)$$

$$= f_\mu(y) \cancel{-\frac{\mu}{2}\|y\|^2} + \nabla f_\mu(y)^T (y-x) + \frac{\mu}{2}\|y-x\|^2$$ $\Rightarrow$ f_μ is μ-strongly convex

(g) (1) since $\mathrm{rank}(a_i a_i^T)=1 \Rightarrow a_i a_i^T$ has only one non-zero eigenvalue.

$$\Rightarrow \lambda_{max}(a_i a_i^T) = tr(a_i a_i^T) = \|a_i\|_2^2.$$

(2)
$$\lambda_{max}(\nabla^2 f_\mu(x)) = \lambda_{max}\left(\sum_{i=1}^n a_i a_i^T \sigma(-b_i a_i^T x)(1-\sigma(-b_i a_i^T x))\right)$$
$$\le \lambda_{max}\left(\sum_{i=1}^n a_i a_i^T + \lambda I\right) \quad (\text{by } 0 \le \sigma(t)(1-\sigma(t)) \le 1)$$
$$\le \sum_{i=1}^n \lambda_{max}(a_i a_i^T) + \mu \le \sum_{i=1}^n \|a_i\|_2^2 + \mu.$$

(3). Clearly ∇f_μ is continuously differentiable.

∇f_μ is L-smooth $\Leftrightarrow \|\nabla f_\mu(x) - \nabla f_\mu(y)\|_2 \le L\|x-y\|_2$.

by Taylor expansion, $\nabla f_\mu(x) = \nabla f_\mu(y) + \nabla^2 f_\mu(\alpha x + (1-\alpha)y)(x-y)$

$\therefore \|\nabla f_\mu(x) - \nabla f_\mu(y)\|_2 \le \|\nabla^2 f_\mu(\alpha x+(1-\alpha)y)\|_2 \|x-y\|_2$

$\le \lambda_{max}(\nabla^2 f_\mu)\|x-y\|_2 = (\|A\|_F^2 + \mu)\|x-y\|_2$

2.2. (a) according to 1.(g)(3), f_i is $(\|a_i\|^2+\mu_i)$-Lip continuous, and hence $L_{max} = \max_i L(f_i)$-Lip continuous.

Then $\mathbb{E}_{i\sim \mathrm{Uniform}([1,n])} \nabla f_i(x) = \sum_{i=1}^n \frac{1}{n}\nabla f_i(x) = \nabla f(x)$.
$\Rightarrow \nabla f_i$ is an unbaised estimate of $\nabla f(x)$

2.3. (b) Take the subgradient: $\lambda \nabla g(y) + y - z = 0 \Rightarrow z = y + \lambda \nabla g(y)$

$\forall$ coordinate i, $z_i = y_i + \lambda \nabla(\|y\|_1)_i$

since $\nabla|x| = \begin{cases} 1 & x>0 \\ [-1,1] & x=0 \\ -1 & x<0 \end{cases}$ we have

if $z_i > \lambda$. y_i must >0. $\Rightarrow y_i = z_i - \lambda$.

if $z_i < -\lambda$ y_i must <0 $\Rightarrow y_i = z_i + \lambda$.

~~if $|z_i| \le \lambda$~~ if $0 \le z_i \le \lambda$. and if $y_i > 0$. $\Rightarrow y_i + \lambda > \lambda \ge z_i$. contradiction

Similarly $y_i < 0$ does not hold $\Rightarrow y_i = 0$.

Similarly if $-\lambda \le z_i \le 0 \Rightarrow y_i = 0$.

in sum, $\mathrm{prox}_{\lambda g}(z) = y = \mathrm{sign}(z) \circ \max(|z|-\lambda, 0)$