

Yale University
CPSC 516, Spring 2023
Assignment 3

Chang Feng (Felix) Zhou cz397

P.1.

(a)

The gradient and Hessian are given by

$$\begin{aligned}\nabla f(x) &= \log x \\ \nabla^2 f(x) &= \text{Diag}(1/x_1, \dots, 1/x_n).\end{aligned}$$

Here the logarithm is applied component-wise and $\text{Diag} : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ maps a vector to the diagonal matrix whose diagonal entries are precisely given by the input vector.

(b)

By our work in class, f is strictly convex if and only if $\nabla^2 f \succ 0$ since f is twice differentiable. Indeed, the eigenvalues of $\nabla^2 f$ are given by

$$1/x_1, \dots, 1/x_n$$

and are all positive over $\mathbb{R}_{>0}^n$. Thus f is indeed strictly convex.

(c)

Suppose towards a contradiction that f is α -strongly convex for some $\alpha > 0$. Let e_1 be the vector which is all zero except for a one in the first entry. We have for all $\lambda > 0$,

$$\begin{aligned}f(\lambda e_1 + \lambda e_1) &\geq f(\lambda e_1) + \langle \nabla f(\lambda e_1), \lambda e_1 \rangle + \frac{\alpha}{2} \|\lambda e_1\|_2^2 \\ 2\lambda \log(2\lambda) - 2\lambda &\geq \lambda \log \lambda - \lambda + \lambda \log \lambda + \frac{\alpha}{2} \lambda^2 \\ 2\lambda \log\left(\frac{2\lambda}{\lambda}\right) &\geq \lambda + \frac{\alpha}{2} \lambda^2 \quad \checkmark \\ 2 \log 2 &\geq 1 + \frac{\alpha}{2} \lambda.\end{aligned}$$

This is a contradiction since we can make the RHS arbitrarily large while the LHS stays constant.

(d)

$$\begin{aligned} D_f(x, y) &:= f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\ &= \sum_{i=1}^n y_i \log y_i - y_i - x_i \log x_i + x_i - (y_i - x_i) \log x_i \\ &= \sum_i y_i \log y_i - y_i + x_i - y_i \log x_i \\ &= \boxed{\sum_{i=1}^n y_i \log \frac{y_i}{x_i} + x_i - y_i}. \end{aligned}$$

+2-1

By inspection, this is not symmetric for all $x, y > 0$.

Please give an example justifying the observation

(e)

We wish to show that

$$D_f(x, y) \geq \frac{1}{2} \|y - x\|_1^2$$

for all $x, y \in \Delta^n := \{x \in \mathbb{R}_{>0}^n : \sum_{i=1}^n x_i = 1\}$. First, we remark that in Δ^n ,

$$D_f(x, y) = \sum_{i=1}^n y_i \log \frac{y_i}{x_i} = \text{KL}(y \| x).$$

Now, Pinsker's inequality states that

$$D_f(x, y) = \text{KL}(y \| x) \geq \frac{1}{2} \|y - x\|_1^2.$$

and so f is 1-strongly convex with respect to the 1-norm, as desired.

+3

P.2.

(a)

First suppose that $d = n$. We have a system of linear inequalities

$$-1 \leq Ax \leq 1$$

where the i -th row of $A \in \mathbb{R}^{m \times n}$ is a_i . By assumption, A has full column rank so that A is injective. We can therefore define a linear left inverse $B: A(\mathbb{R}^n) \rightarrow \mathbb{R}^n$ such that $BAx = x$. But then

$$\|x\| = \|BAx\| \leq \|B\|_{\text{op}} \cdot \|Ax\|$$

for all $x \in \mathbb{R}^n$. In particular, $x \in P$ implies that $\|Ax\| \leq \sqrt{n}$ and consequently $\|x\| \leq \|B\|_{\text{op}} \sqrt{n}$. This shows that P is indeed bounded.

Now suppose that $d < n$. Then we can find some $0 \neq b \in \mathbb{R}^n$ orthogonal to all the a_i 's. Clearly $\lambda b \in P$ for every $\lambda \in \mathbb{R}$ since $\langle a_i, \lambda b \rangle = 0 \leq 1$ for all $i \in [m]$. However, P cannot be bounded since $\|\lambda b\| = |\lambda| \cdot \|b\|$ can be made arbitrarily large.

(b)

The gradient and Hessian are given by

$$g_j(x) = \sum_{i=1}^m 2 \frac{\langle a_i, x \rangle}{1 - \langle a_i, x \rangle^2} a_i$$

$$H(x) = \sum_{i=1}^m 2 \frac{1 + \langle a_i, x \rangle^2}{(1 - \langle a_i, x \rangle^2)^2} a_i a_i^T.$$

(c)

$F(x)$ is finite if and only if for all i ,

$$1 - \langle a_i, x \rangle^2 > 0$$

$$\Rightarrow |\langle a_i, x \rangle| < 1.$$

So $\text{dom } F = \text{int } P$ is the interior of P .

For every $x \in \text{dom } f$, we observe that $H(x)$ is a non-negative linear combination of symmetric rank 1 matrices which are positive semidefinite. It follows that H is positive semidefinite and so F is convex.

(d)

We know that the zeros of the gradient are precisely the global minimizers of F as it is convex. We have $g(0) = 0$ so that 0 is a minimizer.

Furthermore, for $x \in \text{dom } F \setminus 0$, there is some a_i such that $\langle a_i, x \rangle \neq 0$. Thus g is a sum of non-negative functions and one of them takes positive value at x . So $g(x) > 0$ and x is not a minimizer.

It follows that 0 is the unique minimizer.

(e)

We remark that $G(h) := h^T H(x) h$ is convex since its second derivative is $2H(x) \succeq 0$. Thus \mathcal{E}_x is a level-set of a convex function which is necessarily convex. Indeed, for all $h_1, h_2 \in \mathcal{E}_x, \lambda \in [0, 1]$,

$$\begin{aligned} G[\lambda h_1 + (1 - \lambda) h_2] &\leq \lambda G(h_1) + (1 - \lambda) G(h_2) \\ &\leq \lambda + (1 - \lambda) \\ &= 1 \end{aligned}$$

so that the line segment $[h_1, h_2] \subseteq \mathcal{E}_x$ as desired.

We claim that any $h \in \mathcal{E}_x$ satisfies $\langle a_i, h \rangle^2 \leq 1$ for all $i \in [m]$. Suppose otherwise, there is some \bar{i} such that the inequality is violated. We have

$$\begin{aligned} h^T H(x) h &= \sum_{i=1}^m 2 \frac{1 + \langle a_i, x \rangle^2}{(1 - \langle a_i, x \rangle^2)^2} \langle a_i, h \rangle^2 \\ &> 2 \frac{1 + \langle a_{\bar{i}}, x \rangle^2}{(1 - \langle a_{\bar{i}}, x \rangle^2)^2} \cdot 1 \\ &\geq 2. \end{aligned}$$

The last inequality follows from the observation that $\frac{1+z^2}{(1-z^2)^2}$ attains its minimum at $z = 0$. But then $h \notin \mathcal{E}_x$ which is a contradiction. By contradiction, $\mathcal{E}_x \subseteq P$ for all $x \in \text{dom } F$.

↓
This is only true in the domain $-\sqrt{3} \leq z \leq \sqrt{3}$