098311 Optimization 1 Spring 2018
HW 6

Chen Tessler 305052680
Orr Krupnik 302629027

May 24, 2018

Problem 1. Prove that for any $x_1, ..., x_n \in \mathbb{R}_{++}$, the following inequality holds:

$$\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i} \leq \sqrt{\frac{\sum_{i=1}^n x_i^3}{\sum_{i=1}^n x_i}}$$

Solution Using Jensen's inequality with the convex function $f(x) = x^2$ (over all of $\mathbb{R}_++$), we have:

$$\left(\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i}\right)^2 = \left(\sum_{i=1}^n \underbrace{\frac{x_i}{\sum_{i=1}^n x_i}}_{\lambda_i} x_i\right)^2 \leq \sum_{i=1}^n \frac{x_i}{\sum_{i=1}^n x_i} x_i^2 = \frac{\sum_{i=1}^n x_i^3}{\sum_{i=1}^n x_i}$$

Problem 2. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

1. Show that if $f(0) = 0$ and f is an even function, that is $f(-x) = f(x)$, then $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

2. Suppose that f is homogeneous of degree p such that $p > 1$. That is, $f(tx) = t^p f(x)$ for all $x \in \mathbb{R}^n$ and all $t > 0$. Show that $f(0) = 0$ and $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

3. Show that if f is not a constant function, then there exists $0 \neq x_0 \in \mathbb{R}^n$ such that:

$$\lim_{t\to\infty} f(tx_0) = \infty$$

Solution

1. Suppose there exists some x_0 for which $f(x_0) < 0$. Then, since f is even, we have $f(-x_0) = f(x_0)$. Also, since f is convex, we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for any $x, y \in \mathbb{R}^n, \lambda \in [0, 1]$. However, with $\lambda = 0.5$, we have

$$0 = f(0) = f(0.5(-x_0) + 0.5x_0) > 0.5f(-x_0) + 0.5f(x_0) = f(x_0)$$

since $f(x_0) < 0$. This contradicts the convexity of f, and therefore $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

2. Since f is homogeneous, $f(0) = f(t \cdot 0) = t^p f(0), \forall t > 0$. Since $p > 1$ and this holds for any $t > 0$, the equality is only true for $f(0) = 0$.
   Using the homogeneity and convexity of f and the fact that $f(0) = 0$, we have for some $\lambda \in (0, 1)$ and any $x_0 \neq 0$:

$$\lambda^p f(x_0) = f(\lambda x_0) = f(\lambda x_0 + (1 - \lambda) \cdot 0) \leq \lambda f(x_0) + (1 - \lambda) f(0) = \lambda f(x_0)$$

   The inequality above only holds true for any $f(x_0) \geq 0$, but not if $f(x_0) < 0$. Since we selected x_0 arbitrarily, we require $f(x) \geq 0$ for all $x \in \mathbb{R}^n$.

3. First, we show that there exists some point $0 \neq x_0 \in \mathbb{R}^n$ such that $f(x_0) > f(0)$. Assume $\forall x \in \mathbb{R}^n$, $f(x) \leq f(0)$. Then, consider the vectors $x_1 \neq 0, x_2 = \frac{\lambda}{\lambda - 1} x_1 \in \mathbb{R}^n$, for some $\lambda \in (0, 1)$. Then:

$$f(\lambda x_1 + (1 - \lambda) x_2) = f(0) \geq \max\{f(x_1), f(x_2)\} \geq \lambda f(x_1) + (1 - \lambda) f(x_2)$$

   However, this contradicts the convexity of f, and therefore there exists some $0 \neq x_0 \in \mathbb{R}^n$ such that $f(x_0) > f(0)$.
   Now, using the convexity of f, and selecting $\lambda = \frac{1}{t}$ (this works for any $t > 1$) and x_0 such that $f(x_0) > f(0)$ (which we've shown above exists):

$$\frac{1}{t} f(tx_0) + \left(1 - \frac{1}{t}\right) f(0) \geq f\left(\frac{1}{t} \cdot tx_0 + \left(1 - \frac{1}{t}\right) \cdot 0\right) = f(x_0)$$
$$\Rightarrow f(tx_0) \geq t\left(f(x_0) - \left(1 - \frac{1}{t}\right) f(0)\right) > t\underbrace{(f(x_0) - f(0))}_{>0} \xrightarrow{t \to \infty} \infty$$

   And therefore $\lim_{t \to \infty} f(tx_0) = \infty$.

Problem 3.

1. Let $f : C \to \mathbb{R}$ be a concave function over a convex set $C \subseteq \mathbb{R}^n$ and let $g : \mathbb{R} \to \mathbb{R}$ be a nondecreasing concave function. Prove that the composition of g with f defined by

$$h(x) = g(f(x))$$

   is a concave function over C.

2. Prove that the function $f(x) = \frac{1}{1+e^x}$ is strictly convex over $[0, \infty)$.

3. Prove that for any $a_1, a_2, ..., a_n \geq 1$ the inequality:

$$\sum_{i=1}^{n} \frac{1}{1 + a_i} \geq \frac{n}{1 + \sqrt[n]{a_1 \cdot a_2 \cdot ... \cdot a_n}}$$

Solution

1. Since g is nondecreasing, for any $z, w \geq z \in \mathbb{R}$ we have $g(w) \geq g(z)$ (a). Since f is concave, we have for any $x, y \in C$: $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$ (b). Considering also the concavity of g (c), for any $x, y \in C$:

$$h(\lambda x + (1-\lambda)y) = g(f(\lambda x + (1-\lambda)y)) \overset{(a)+(b)}{\geq} g(\lambda f(x) + (1-\lambda)f(y) \geq$$
$$\overset{(c)}{\geq} \lambda g(f(x)) + (1-\lambda)g(f(y)) = \lambda h(x) + (1-\lambda)h(y)$$

And therefore h is concave over C.

2. Based on the second order characterization of convexity, a twice differentiable function f is strictly convex over a set C iff $f''(x) > 0$ for any $x \in C$.

$$f(x) = \frac{1}{1+e^x}$$
$$f'(x) = -\frac{e^x}{(1+e^x)^2}$$
$$f''(x) = \frac{2e^{2x}}{(1+e^x)^3} - \frac{e^x}{(1+e^x)^2}$$

f is twice differentiable over the subset $C = [0, \infty) \in \mathbb{R}$. At $x = 0$ the value $f''(x = 0) = \frac{2}{2^2} - \frac{1}{2^2} = 0$.

In addition, for any $x > 0$:

$$f''(x) = \frac{2e^{2x}}{(1+e^x)^3} - \frac{e^x}{(1+e^x)^2} = \frac{2e^{2x} - e^x - e^{2x}}{(1+e^x)^3} = \frac{e^x(e^x-1)}{(1+e^x)^3} > 0 \ \ \forall x > 0$$

for $e^x > 1$ for any $x \in (0, \infty)$.

As $f''(x) > 0 \ \forall x \in (0, \infty)$ we conclude, using the second order characterization of convexity, that $f(x)$ is strictly convex on the set $(0, \infty)$.

Based on the first derivative and the linear approximation theorem, there exists a $c \in \mathbb{R}_{++}$ such that for any $y \in (0, c)$: $f(y) < f(0)$.

Need to show: $f(\eta x) < \eta f(x) + (1-\eta)f(0)$ In addition, the following holds for any $y \in (0, c)$ and $x > 0$:

$$\eta f(x) + (1-\eta)f(0) = \lim_{y \to 0^+} \eta f(x) + (1-\eta)f(y) > \lim_{y \to 0^+} f(\eta x + (1-\eta)y) = f(\eta x + (1-\eta)0)$$

where the equality holds due to the continuity of $f(x)$ and the inequality as shown above that $f(x)$ is strictly convex for any $x \in (0, \infty)$.
Hence combining both of the above, we have that $f(x)$ is strictly convex $\forall x \in [0, \infty)$.

3. We can write any $a_i \geq 1$ as e^{b_i} for some $b_i > 0$. Therefore, the inequality we are trying to prove is:

$$\frac{1}{n}\sum_{i=1}^{n} f(b_i) \geq \frac{1}{1 + e^{\frac{1}{n}\sum_{i=1}^{n} b_i}} = f\left(\frac{1}{n}\sum_{i=1}^{n} b_i\right)$$

using the notation of f from the previous section. Since we have shown f is strictly convex for the interval over which all b_i are defined, this is exactly Jensen's inequality.

Problem 4. Let C be a convex subset of $\mathbb{R}^n$. A function f is called **strongly convex** over C if there exists $\sigma > 0$ such that the function $f(x) - \sigma\frac{||x||^2}{2}$ is convex over C. The parameter σ is called the strong convexity parameter.

1. Prove that f is strongly convex over C with parameter σ if and only if
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)||x-y||^2$$
2. Prove that a strongly convex function over C is also strictly convex over C.
3. Suppose that f is continuously differentiable over C. Prove that f is strongly convex over C with parameter σ iff
$$f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\sigma}{2}||x-y||^2$$
for any $x, y \in C$.
4. Suppose that f is continuously differentiable over C. Prove that f is strongly convex over C with parameter σ iff
$$(\nabla f(x) - \nabla f(y))^T(x-y) \geq \sigma||x-y||^2$$
for any $x, y \in C$.

Solution

1. We show that the convexity of $f(x) - \sigma\frac{||x||^2}{2}$ is equivalent to the required inequality:
$$\begin{aligned}
&f(\lambda x + (1-\lambda)y) - \sigma\frac{||\lambda x + (1-\lambda)y||^2}{2} \leq \lambda\left(f(x) - \sigma\frac{||x||^2}{2}\right) + (1-\lambda)\left(f(y) - \sigma\frac{||y||^2}{2}\right)\\
\iff\ &f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\left(\lambda||x||^2 + (1-\lambda)||y||^2 - ||\lambda x + (1-\lambda)y||^2\right)\\
\iff\ &f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}(\lambda||x||^2 + (1-\lambda)||y||^2 -\\
&\qquad - \lambda^2||x||^2 - (1-\lambda)^2||y||^2 - 2\lambda(1-\lambda)x^Ty)\\
\iff\ &f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)\left(||x||^2 + ||y||^2 - 2x^Ty\right)\\
\iff\ &f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)||x-y||^2 \quad \square
\end{aligned}$$

2. Following the definitions of $\sigma > 0$ and $\lambda \in (0,1)$, we have that $\frac{\sigma}{2}\lambda(1-\lambda)||x-y||^2 > 0$ for any $x \neq y$. Therefore, for any $x \neq y \in C$, we have for a strongly convex function, using the inequality from the previous section:
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\sigma}{2}\lambda(1-\lambda)||x-y||^2 < \lambda f(x) + (1-\lambda)f(y)$$
and therefore it is also strictly convex by definition.

3. Let us denote $g(x) = f(x) - \sigma\frac{||x||^2}{2}$. Since f is strongly convex with parameter $\sigma > 0$ if g is convex, for any strongly convex f we have that g obeys the gradient inequality for any $x, y \in C$:

$$g(y) \geq g(x) + \nabla g(x)^T(y - x)$$

We have $\nabla g(x) = \nabla f(x) - \sigma x$. Therefore, we can plug in the gradient and the definition of g to obtain:

$$\begin{aligned}
f(y) - \sigma\frac{||y||^2}{2} &\geq f(x) - \sigma\frac{||x||^2}{2} + (\nabla f(x) - \sigma x)^T(y - x) \\
\iff f(y) &\geq f(x) + \frac{\sigma}{2}\left(||y||^2 - ||x||^2\right) + (\nabla f(x) - \sigma x)^T(y - x) \\
\iff f(y) &\geq f(x) + \frac{\sigma}{2}\left(||y||^2 - ||x||^2\right) + \nabla f(x)^T(y - x) - \sigma x^T y + \sigma||x||^2 \\
\iff f(y) &\geq f(x) + \frac{\sigma}{2}\left(||y||^2 - 2x^T y + ||x||^2\right) + \nabla f(x)^T(y - x) \\
\iff f(y) &\geq f(x) + \nabla f(x)^T(y - x) + \frac{\sigma}{2}||x - y||^2 \quad \square
\end{aligned}$$

4. As defined above, f is strongly convex with parameter $\sigma < 0$ iff $g(x) = f(x) - \sigma\frac{||x||^2}{2}$ is convex. Since f is continuously differentiable over C, so is g, and therefore g is convex iff its gradient is monotonous (lecture 7 slide 11), meaning for every $x, y \in C$:

$$\begin{aligned}
&(\nabla g(x) - \nabla g(y))^T(x - y) \geq 0 \\
\iff &(\nabla f(x) - \sigma x - \nabla f(y) + \sigma y)^T(x - y) \geq 0 \\
\iff &(\nabla f(x) - \nabla f(y))^T(x - y) - \sigma x^T x + 2\sigma x^T y - \sigma y^T y \geq 0 \\
\iff &(\nabla f(x) - \nabla f(y))^T(x - y) \geq \sigma||x - y||^2 \quad \square
\end{aligned}$$

Problem 5. Show that the following functions are convex/concave over the specified domain C:

1. $f(x, y, z) = -\sqrt{xy} + 2x^2 + 2y^2 + 3z^2 - 2xy - 2yz$ over $C = \mathbb{R}^3_+$
2. $f(x) = \sum_{i=1}^n x_i \ln(x_i) - (\sum_{i=1}^n x_i)\ln(\sum_{i=1}^n x_i)$ over $C = \mathbb{R}^n_{++}$
3. $f(x) = \sqrt[n]{\prod_{i=1}^n x_i}$ over $C = \mathbb{R}^n_+$. (Hint: prove first for $\mathbb{R}^n_{++}$)

Solution

1. We can write f as:

$$f(x, y, z) = (x, y, z)\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} - \sqrt{xy}$$

The first part of the function is a quadratic function with $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}$, and $b = 0, c = 0$. A is PD (strictly diagonally dominant with positive values along the diagonal) and therefore the first part of the function is strictly convex (quadratic function with PD matrix A).
We shall now show the second part of the function is also convex. We start by defining $g(x,y,z) = -\sqrt{xy}$, and calculating $\nabla g(x,y,z)$:

$$\nabla(g(x,y,z)) = \nabla(-\sqrt{xy}) = \left(-\frac{y}{2\sqrt{xy}}, -\frac{x}{2\sqrt{xy}}, 0\right)^T = \frac{1}{2g(x,y,z)}(y,x,0)^T$$

Now, $g(x,y,z)$ is convex if and only if it obeys the gradient inequality for any two vectors $(x_1,y_1,z_1)^T, (x_2,y_2,z_2)^T \in \mathbb{R}^3_+$:

$$\begin{aligned}
g(x_2,y_2,z_2) &\geq g(x_1,y_1,z_1) + \frac{1}{2g(x_1,y_1,z_1)}(y_1,x_1,0)(x_2-x_1, y_2-y_1, z_2-z_1)^T = \\
&= -\sqrt{x_1,y_1} - \frac{y_1(x_2-x_1) + x_1(y_2-y_1)}{2\sqrt{x_1y_1}} = \\
&= -\sqrt{x_1y_1} + \frac{\sqrt{x_1y_1}}{2} + \frac{\sqrt{x_1y_1}}{2} - \frac{y_1x_2 + x_1y_2}{2\sqrt{x_1y_1}} = -\frac{y_1x_2 + x_1y_2}{2\sqrt{x_1y_1}} \\
&\iff \sqrt{x_1y_1x_2y_2} \leq \frac{x_1y_2 + x_2y_1}{2}
\end{aligned}$$

The latter is true according to the AM-GM inequality (recall $x, y \in \mathbb{R}_+$), and therefore $g(x,y,z)$ is convex.
Finally, since f is a summation of two convex functions over $\mathbb{R}^3_+$, it is also a convex function over $\mathbb{R}^3_+$.

2. Since $\mathbb{R}^n_{++}$ is open, f is convex over the set iff $\nabla^2(x) \succeq 0$ for any $x \in \mathbb{R}^n_{++}$. Calculating the Hessian:

$$\begin{aligned}
\nabla f(x)_i &= \ln(x_i) - \ln\left(\sum_{k=1}^{n} x_k\right) \\
\nabla^2 f(x)_{i,i} &= \frac{1}{x_i} - \frac{1}{\sum_{k=1}^{n} x_k} \\
\nabla^2 f(x)_{i,j} &= -\frac{1}{\sum_{k=1}^{n} x_k}
\end{aligned}$$

By definition, the Hessian matrix is PSD if $v^T \nabla^2 f(x) v \geq 0, \;\; \forall x, v \in \mathbb{R}^n$.

$$
\begin{aligned}
v^T \nabla^2 f(x) v = \sum_{i=1}^{n} \frac{v_i^2}{x_i} - \frac{(\sum_{i=1}^{n} v_i)^2}{\sum_{i=1}^{n} x_i} \geq 0 \\
\iff \frac{1}{n} \sum_{i=1}^{n} \frac{v_i^2}{x_i} - \frac{1}{n} \frac{(\sum_{i=1}^{n} v_i)^2}{\sum_{i=1}^{n} x_i} \geq 0 \\
\iff \sum_{i=1}^{n} \frac{1}{n} \frac{v_i^2}{x_i} \geq \frac{(\sum_{i=1}^{n} \frac{1}{n} v_i)^2}{\sum_{i=1}^{n} \frac{1}{n} x_i}
\end{aligned}
$$

Defining $y_i = (x_i, v_i)^T$ and $f(y) = \frac{y_1^2}{y_2}$, which is convex (quad-over-lin), the last inequality above is exacly Jensen's inequality and therefore $\nabla^2 f(x) \succeq 0$, meaning $f(x)$ is convex over C.

3. We first present a lemma, the "opposite" gradient inequality.

   Lemma. Let $f : C \to \mathbb{R}$ be a continuously differentiable function defined on a convex set $C \subseteq \mathbb{R}^n$. Then f is **concave** over C if and only if

   $$f(y) \leq f(x) + \nabla f(x)^T (y - x)$$

   for any $x, y \in C$.

   Proof. As we've seen in class, f is concave over a set C iff $-f$ is convex over the set. Additionally, a function is convex over a set C iff it obeys the gradient inequality for any $x, y \in C$. Combining, We have:

   $$
   \begin{aligned}
   &f \text{ is concave over } C \iff -f \text{ is convex over } C \iff \\
   &\qquad -f(y) \geq -f(x) - \nabla f(x)^T (y - x) \iff f(y) \leq f(x) + \nabla f(x)^T (y - x)
   \end{aligned}
   $$

   $\square$

   Now, we shall prove $f(x)$ is concave over $\mathbb{R}^n_{++}$. In this case, $x_i > 0 \ \forall i = \{1, ..., n\}$ and we have:

   $$\nabla f(x)_i = \frac{1}{n x_i} \sqrt[n]{\prod_{j=1}^{n} x_j} = \frac{1}{n x_i} f(x)$$

   Now, f is concave if it obeys the "opposite" gradient inequality (proven in the lemma above) for all $x, y \in \mathbb{R}^n_{++}$:

   $$
   \begin{aligned}
   f(y) \leq f(x) + \nabla f(x)^T (y - x) = f(x) + \sum_{i=1}^{n} \frac{f(x)(y_i - x_i)}{n x_i} = \frac{f(x)}{n} \sum_{i=1}^{n} \frac{y_i}{x_i} \\
   \iff \sqrt[n]{\prod_{i=1}^{n} \frac{y_i}{x_i}} = \frac{f(y)}{f(x)} \leq \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{x_i}
   \end{aligned}
   $$

Defining $z_i = \frac{y_i}{x_i}$ (well defined since both y_i and x_i are positive), the final inequality is exactly the generalized AM-GM inequality, and therefore f is concave over $\mathbb{R}^n_{++}$.
Note $C = \mathbb{R}^n_+$ can be defined as $\mathbb{R}^n_{++} \cup \underbrace{\{x : x_i \geq 0 \ \forall i, \exists i : x_i = 0\}}_{\triangleq D}$. We also note for $x \in D$ we have $f(x) = 0$, and therefore D itself is trivially concave.
It remains to check the case of two points, $x \in \mathbb{R}^n_{++}$ and $y \in D$ (WLOG). For any $\lambda \in (0, 1)$:

$$\begin{aligned}\lambda f(x) + (1-\lambda) f(y) = \lambda f(x) = \lambda \sqrt[n]{\prod_{i=1}^{n} x_i} &\overset{(a)}{\leq} \lambda \sqrt[n]{\prod_{i=1}^{n} \left(x_i + \frac{1-\lambda}{\lambda} y_i \right)} = \\ &= \sqrt[n]{\prod_{i=1}^{n} (\lambda x_i + (1-\lambda) y_i)} = f(\lambda x + (1-\lambda) y)\end{aligned}$$

where (a) is true since adding non-negative factors $(\frac{1-\lambda}{\lambda} y_i)$ to positive elements (x_i) of a product will always enlarge it. Therefore, in this case as well f is concave, and we can conclude that f is concave over $\mathbb{R}^n_+$.