# Assignment 2

**Problem 1**: Determine whether convex, concave, quasiconvex, or quasiconcave?

a) $f(x) = e^x - 1$ on $R$

Strictly convex as $f''(x) > 0$. Therefore, quasiconvex. Also quasiconcave, but not concave.

b) $f(x_1, x_2) = x_1 x_2$ on $R^2_{++}$

Hessian is $\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ which indefinite. Thus, $f$ is neither convex nor concave. It is quasiconcave, sine its ~~supporting~~ supperlevel sets $\{(x_1, x_2) \in R^2_{++} \mid x_1 x_2 \geq \alpha\}$ are convex. It is not quasiconvex.

c) $f(x_1, x_2) = \dfrac{1}{x_1 x_2}$ on $R^2_+$.

Hessian is $\nabla^2 f(x) = \dfrac{1}{x_1 x_2} \begin{bmatrix} \frac{2}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^2} \end{bmatrix} \succeq 0$

It is convex and quasiconvex, not quasiconcave, not concave.

d) $f(x_1, x_2) = \dfrac{x_1}{x_2}$ on $R^2_+$

Hessian is $\nabla^2 f(x) = \begin{bmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & 2\frac{x_1}{x_2^3} \end{bmatrix}$ is indefinite.

Thus, function is not convex, not concave. It's quasiconvex & quasiconcave (i.e., quasilinear), because the sublevel and superlevel sets are halfspaces.

e) $f(x_1,x_2)=\frac{x_1^2}{x_2}$ on $R\times R_+$

Hessian: $\nabla^2 f(x)=\begin{bmatrix}\frac{2}{x_2} & -\frac{2x_1}{x_2^2}\\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3}\end{bmatrix}=\frac{2}{x_2}\begin{bmatrix}1\\ -\frac{2x_1}{x_2}\end{bmatrix}\begin{bmatrix}1 & -2\frac{x_1}{x_2}\end{bmatrix}$

i.e. it is rank 1 matrix for which quadratic form is always nonnegative, and thus,

$\nabla^2 f(x)\succeq 0$

Thus, f is convex and quasiconvex. It's not concave and quasiconcave.

f) $f(x_1,x_2)=x_1^{\alpha}x_2^{1-\alpha}$, $0\le\alpha\le 1$ on $R_+^2$

Hessian: $\nabla^2 f(x)=\begin{bmatrix}\alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha}\\ \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} & (1-\alpha)(-\alpha)x_1^{\alpha}x_2^{-\alpha-1}\end{bmatrix}=$

$=\alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha}\begin{bmatrix}-\frac{1}{x_1^2} & \frac{1}{x_1x_2}\\ \frac{1}{x_1x_2} & -\frac{1}{x_2^2}\end{bmatrix}=$

$=-\alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha}\begin{bmatrix}1/x_1\\ -1/x_2\end{bmatrix}\begin{bmatrix}1/x_1 & -1/x_2\end{bmatrix}\preceq 0$

rank 1 matrix with '−' in front,

f is concave and quasiconcave, not convex, not quasiconvex.

<u>Problem 2:</u> x is real r.v. that takes vales on $\{a_1,\dots,a_n\}$, $a_1<a_2<a_3<\dots<a_n$ with $prob(x=a_i)=p_i$

On probability simplex $\{p\in R_+^n \mid 1^T p=1, p_i\ge 0\}$ determine if the following function: convex, concave, quasiconvex, quasiconcave?

<u>Solution:</u> 1) $E\,x=p_1a_1+\dots+p_na_n$ is liner. Thus, convex, concave, quasiconvex, quasiconcave,

2) $\text{prob}(x \geq \alpha)$

Let $j = \min\{i \mid a_i \geq \alpha\}$. Then $\text{prob}(x \geq \alpha) = \sum_{i=j}^{n} p_i$.

This is linear function of $p$. Thus, convex, concave, quasiconvex, quasiconcave.

3) $\text{prob}(\alpha \leq x \leq \beta)$.

Let $j = \min\{i \mid a_i \geq \alpha\}$ and $k = \max\{i \mid a_i \leq \beta\}$.

Then $\text{prob}(\alpha \leq x \leq \beta) = \sum_{i=j}^{k} p_i$. The same as above.

4) $\sum_{i=1}^{n} p_i \log p_i$

$p \log p$ is a convex function in $R_+$ (assuming $0 \log 0 = 0$), so negentropy is convex (thus, quasiconvex).

It's not concave or quasiconcave.

For example, $n=2$, $p^1 = (1,0)$ and $p^2 = (0,1)$.

Both $p^1$ and $p^2$ have function value 0, but the convex combination $(\frac{1}{2}, \frac{1}{2})$ has function value $\log(\frac{1}{2}) < 0$. Thus, superlevel sets are not convex.

5) $\text{var}\, x = E(x - Ex)^2$

$$\text{var}\, x = Ex^2 - (Ex)^2 = \sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{i=1}^{n} p_i a_i\right)^2$$

Thus, $\text{var}\, x$ is concave quadratic function of $p$. But it is not convex (quasiconvex).

For example, $n=2$, $a_1=0$, $a_2=1$.
Both $(p_1, p_2) = (1/4, 3/4)$ and $(p_1, p_2) = (3/4, 1/4)$ lie in the probability simplex and have $\operatorname{var} x = \frac{3}{16}$, but the convex combination $(p_1, p_2) = (\frac{1}{2}, \frac{1}{2})$ has a variance $\operatorname{var} x = \frac{1}{4} > \frac{3}{16}$. Thus, the sublevel sets are not convex.

6) $\operatorname{quartile}(x) = \inf\{\beta \mid \operatorname{prob}(x \leq \beta) \geq 0{,}25\}$.
The sublevel and superlevel sets are convex. Indeed, $\operatorname{quartile}(x) \geq \alpha \Rightarrow$
~~$\Rightarrow \operatorname{prob}(x \leq a_k) = \sum_{i=1}^{k} p_i < 0{,}25$.~~
$\Rightarrow \operatorname{prob}(x \leq \beta) < 0{,}25 \quad \forall \beta < \alpha$
If $\alpha \leq a_1$, this is alway true.
Otherwise, define $k = \max\{i \mid a_i < \alpha\}$. This is fixed integer, independent on $p$, and the superlevel set constraint $f(p) \geq \alpha$ holds iff $\operatorname{prob}(x \leq a_k) = \sum_{i=1}^{k} p_i < 0{,}25$.
This is a strict linear inequality of $p$.
But quartile(x) is not continuous (it takes values in a discrete set $\{a_1, \dots a_n\}$, so it is not convex, not concave, but it is quasiconvex and quasiconcave.

7) The cardinality of the smallest set $A \subseteq \{a_1, \dots, a_n\}$ with probability $\geq 90\%$.

$f$ has integer values, and cannot be convex or concave. But it is quasiconcave because its superlevel sets are convex. $f(p) \geq \alpha$ iff

$$\sum_{i=1}^{k} p_{[i]} < 0.9.$$

where $k = \max\{i \mid i < \alpha\}$ is ~~the~~ largest than $\alpha$, and $p_{[i]}$ is $i$th largest component of $p$.
We know that $\sum_{i=1}^{k} p_{[i]}$ is a convex function on $p$, so the inequality $\sum_{i=1}^{k} p_{[i]} < 0.9$ defines a convex set.
$f$ is not quasiconvex. For example, for $n=2$, $a_1=0$, $a_2=1$, $p^1=(0.1, 0.9)$, $p^2=(0.9, 0.1) \Rightarrow$
$f(p^1)=f(p^2)=1$, but $f\left(\frac{p^1+p^2}{2}\right)=f(0.5, 0.5)=2$.
8) min width interval that contains 90% of the probabilities, i.e. $\inf\{\beta-\alpha \mid \text{prob}(\alpha \leq x \leq \beta) \geq 0.9\}$.
Find this min width interval $[a_i \; a_j]$ with $1 \leq i \leq j \leq n$. Indeed,

$$\text{prob}(\alpha \leq x \leq \beta) = \sum_{k=i}^{j} p_k = \text{prob}(a_i \leq x \leq a_k)$$

where $i = \min\{k \mid a_k \geq \alpha\}$, $j = \max\{k \mid a_k \leq \beta\}$
We have $f(p) \geq \gamma$ iff all intervals of width less than $\gamma$ have a probability $< 90\%$,
$\sum_{k=i}^{j} p_k < 0.9$, $\forall i, j$ that satisfy $a_j - a_i < \gamma$.
This defines a convex set. Thus, the function is quasiconcave.
It is not convex, concave nor quasiconvex.

For example, for $n=3$, $a_1=0$, $a_2=0{,}5$, $a_3=1$.
On the line $p_1+p_3=0{,}95$, we have

$$f(p)=\begin{cases}0, & p_1+p_3=0.95,\ p_1\in[0{,}05,0{,}1]\cup[0{,}9,0{,}95]\\ 0.5, & p_1+p_3=0{,}95,\ p_1\in(0{,}1,0{,}15]\cup[0{,}85,0{,}9)\\ 1, & p_1+p_3=0{,}95,\ p_1\in(0{,}15,\ 0{,}85)\end{cases}$$

which is not convex, concave nor quasiconvex on the line.

**Problem 4:** Prove that if $h(x)$ is convex and twice continuosly differentiable, then $\nabla^2 h(x)\geqslant 0$.

**Solution:** Prove the following lemma for 1-D case.

**Lemma:** A twice continuosly differentiable function $f(x): R\to R$ is convex iff $f''(x)\geqslant 0$.

**Proof:** $(\Rightarrow)$: Assume $f(x)$ to be convex.
For convex function 1st order Taylor approximation is $f(x)\geqslant f(y)+f'(y)(x-y)$
and $f(y)\geqslant f(x)+f'(x)(y-x)$ $\quad \forall x,y\in R$
Equivalently,

$$f'(y)(y-x)\geqslant f(y)-f(x)\geqslant f'(x)(y-x) \Rightarrow$$

$$\Rightarrow \frac{f'(y)-f'(x)}{y-x}\geqslant 0$$

Finally, from the definition of the derivative we have $f''(x)=\lim\limits_{y\to x}\dfrac{f'(y)-f'(x)}{y-x}\geqslant 0$.

$(\Leftarrow)$: Consider $f''(x)\geqslant 0$. In this case, $\forall x,y\in R$, we have the nonnegative integral:

$$0\geqslant\int_x^y f''(z)(y-z)\,dz=\left(f'(z)(y-z)\right)\Big|_x^y+$$

$$+\int_x^y f'(z)\,dz=-f'(x)(y-x)+f(y)-f(x)$$

which leads to $f(y)\geqslant f(x)+f'(x)(y-x)$.
Thus, $f(x)$ is convex ■

Generalize now for $n>1$. Use property that convex function is convex on lines, that is $g(t)=h(x_0+t\ell)$ is convex in $t$, for given $x_0$ and $\ell$. By ~~the~~ Lemma above, this is equivalent to the condition that $g''(t)\geq 0$. Therefore, the function $h(x)$ is convex iff $g''(t)=\frac{d^2(h(x_0+t\ell))}{dt^2}=$

$=v^T\nabla^2 h(x_0+t\ell)v\geq 0,\ \forall x_0, \ell\in R^n$.

But this is exactly the condition that the Hessian should satisfy $\nabla^2 h(x)\succeq 0,\ \forall x\in R^n$.

Problem 3: ~~1)~~ $f(X)=tr(X^{-1})$ is convex?

Consider any symmetric positive definite $X$ such that $X\succ 0$. Using the property that $f(X)$ is convex iff it is convex on all lines $X=Z+tV$ for given symmetric $Z$ and $V$, that is

$g(t)=f(Z+tV)$ is convex.

Since $X\succ 0$, the dom $g=\{t \mid Z+tV\succ 0\}$

Consider now the equivalence

$g(t)=tr\{(Z+tV)^{-1}\}=tr\{Z^{1/2}(I+tZ^{-1/2}VZ^{-1/2})Z^{1/2}\}^{-1}$

$=tr(Z^{-1}(I\,Z^{-1/2}VZ^{-1/2})^{-1}\}$

given that $Z^{1/2}Z^{1/2}=Z$ is always possible since $Z\succ 0$.

Now note that $Z^{-1/2}VZ^{-1/2}\succ 0$ such that it accepts

~~f is integer-values, and cannot be convex or concave, But it is quasiconcave because its sup~~ the eigendecomposition

$$Z^{-1/2} V Z^{-1/2} = U^T \Lambda U \quad \text{with}$$

$\Lambda = \mathrm{diag}\{\lambda_1, \dots, \lambda_n\}$ and $U^T U = I$. Thus,

$$g(t) = \mathrm{tr}\{Z^{-1}(I + tU^T\Lambda U)^{-1}\} = \mathrm{tr}\{Z^{-1} U (I + t\Lambda)^{-1} U^T\} =$$

$$= \mathrm{tr}\{\underbrace{U^T Z^{-1} U}_{W} (I + t\Lambda)^{-1}\} = \sum_{i=1}^{n} w_{ii} \frac{1}{1 + t\lambda_i}$$

here the second and third equalities follow from the properties of $\mathrm{tr}\{\cdot\}$ (can be rotated under trace) and the last equality follows from the definition of the trace given the fact that $(I + t\Lambda)^{-1}$ is a diagonal matrix. Finally,

$$g''(t) = \sum_{i=1}^{n} w_{ii} \frac{2\lambda_i^2}{(1 + t\lambda_i)^3} \geq 0$$

where the last inequality follows from $W \succ 0$ (its diagonal sum is always positive), and the fact that eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ are all positive. Thus, since $g''(t) \geq 0$, the function $g(t)$ is convex and so $f(X)$.

2) $f(X) = (\det X)^{1/n}$ is concave?

Similar to the previous case, we show that $f(X)$ is concave in all lines $X = Z + tV$, with $Z, V$ symmetric. Thus, consider the following one-dimentional function

$g(t) = f(Z + tV)$ such that,

$g(t) = (\det(Z + tV))^{1/n} = \det(Z^{1/2}(I + tZ^{-1/2} V Z^{-1/2})Z^{1/2})^{1/n}$

$= (\det(Z^{1/2}) \det(I + tZ^{-1/2} V Z^{-1/2}) \det(Z^{1/2}))^{1/n} =$

$= (\det Z)^{1/n} (\det(I + tZ^{-1/2} V Z^{-1/2}))^{1/n}$.

Again consider the eigendecomposition

$Z^{-1/2} V Z^{-1/2} = U^T \Lambda U > 0$

$\Lambda = \operatorname{diag}\{\lambda_1, \dots, \lambda_n\}$ the eigenvalues of $Z^{-1/2} V Z^{-1/2}$, which are positive.

Now notice that the eigenvales of $(I + tU^T \Lambda U)$ are $\tilde{\lambda}_i = (1 + t\lambda_i)$, $\forall\, i = 1, \dots, n$

Using the property that

$$\det(A) = \prod_{i=1}^{n} \lambda_i(A) \text{ we have}$$

$$g(t) = (\det Z)^{1/2} \left(\prod_{i=1}^{n} (1 + t\lambda_i)\right)^{1/n}$$

Since $Z > 0$, we know that $(\det Z)^{1/n} \geq 0$. Moreover, we know that the geometric sum $\left(\prod_{i=1}^{n} (1 + t\lambda_i)\right)^{1/n}$ is concave on $\mathbb{R}^n_+$.

Thus, since $f(X)$ consists of the positive scaling of a convex function, it's also concave.

<u>Problem 5:</u> 1) sum of $r$ largest componets of a vector $x \in \mathbb{R}^n$ is a convex function.

<u>Solution:</u> $f(x) = \sum_{i=1}^{r} x_{[i]}$ given sorting

$x_{[1]} \geq x_{[2]} \geq x_{[3]} \geq \dots \geq x_{[n]}$, where $x_{[i]}$ is the ith largest element of $x$.

Adding the largest $r$ element of a vector corresponds to the largest possible sum of $r$ elements of such vector. That is, given any $x_j \le x_{[k]}$, inequality

$$\sum_{i=1}^{k} x_{[i]} + x_j > \sum_{i=1}^{k} x_{[i]} + x_{[k+1]}$$ would imply

that $x_{[k+1]} < x_j < x_{[k]}$, which contradicts the definition that $x_{[k+1]}$ is the $(k+1)$st larges element for any $1<k<r$.

Therefore, we can represent the function as

$$f(x) = \max\left\{ \sum_{i=1}^{r} x_{k_i} \,\middle|\, 1 \le k_1 < k_2 < \ldots\ldots < k_r \le n \right\}$$

which consists of nonnegative linear operations that preserve convexity. Thus, $f(x)$ is convex.

2) $X$ in $n\times n$ symmetric positive semi-definite

$$\lambda_1(X) \geq \lambda_2(X) \geq \ldots\ldots \geq \lambda_n(X)$$

Prove that $f(X) = \sum_{i=1}^{k} \lambda_k(X)$, $k \le n$ is convex?

Hint: $f(X) = \sup\{ tr(U^T X U) \mid U \in R^{n\times k}, U^T U = I_k \}$.

Solution: ← not using the hint. The epigraph of $f(X)$ ~~admits~~

$\{(X,t) \mid f(X) \le t\}$ admits the SDR

(a) $t - ks - Tr\{Z\} \geq 0$ where $Z \in S^{m\times m}$

(b) $Z \succeq 0$ and $s \in R$ are additional

(c) $Z - X + sI_m \succeq 0$ variables

We should prove that:

(i) If a given pair $(X,t)$ can be extended by properly chosen $(s,Z)$ to a solution of the LMI (a)-(c), then $f(X)\le t$.

(ii) Vise versa, if $f(x)\le t$, that the pair $(X,t)$ can be extended by properly chosen $(s,Z)$ to a solution of (a)-(c).

Proof of (i): Basic fact: The vector $\lambda(X)$ is monotone function of $X\in S^{n\times n}$, the space of symmetric martices being equiped with the order ≥

$$X\succeq X' \Leftarrow \lambda(x)\ge\lambda(X')$$

Assuming $(x,t,s,Z)$ is a solution to (a)-(c), we get $X\preceq Z+sI_m$ so that

$$\lambda(x)\le\lambda(Z+sI_m)=\lambda(Z)+s\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

Thus, $f(x)\le f(Z)+sk$

Since $Z\succeq 0\Rightarrow f(x)\le Tr\{Z\}$, or together

$$f(X)\le Tr\{Z\}+sk \underset{\text{becase of (a)}}{\Rightarrow} f(x)\le t.$$

Proof of (ii): assume that we are given $(X,t)$ with $f(x)\le t$, and let us use $s=\lambda_k(X)$.

Then the $k$ largest eigenvalues of matrix $X-sI_m$ are nonnegative, and the remaining ope nonpositive. Let $Z$ being symmetric with the same eigenvalues as $X$ and such that the $k$ largest eigenvalues of $Z$ are the same as those of $X - sI_m$ and the remaining eigenvalues are zeros, $Z$ and $Z - X + sI_m$ are clearly PSD (first by construction, second since in the eigenbasis of $X$ this matrix is diagonal with the first $k$ diagonal entries being $0$ and remaining being the same as those of the matrix $sI_m - X$, i.e., being nonnegative). Thus, $Z$ and real $s$ we have built satisfy (b) and (c). To see that (a) is also satisfied, note that by construction $Tr\{Z\} = f(X) - sk$,

$$t - sk - Tr\{Z\} = t - f(X) \geq 0.$$

Problem 6: A quadratic-over-linear composition.

$f: R^n \to R$ is nonnegative and convex,
$g: R^n \to R$ is positive and concave.
Show that $f^2/g$, with domain $\mathrm{dom} f \cap \mathrm{dom}\, g$, is convex.

Solution: Show that the function can be represented as $h(f^2(x), g^{-1}(x)) = f^2(x) g^{-1}(x)$, which is convex by the property of $k$-dimentional composition: if $h: R_{\geq 0} \times R_{>0} \to R$ is convex nondecreasing in each argument, and both ~~nondecreasing in each argument~~ $f^2: R^n \to R_{\geq 0}$ and $g^{-1}: R \to R_{>0}$ are convex.

a) First, we show that $f^2$ is convex.
Consider the function $b_1(x)=x^2$. It is nondecreasing and concave in $R_{\geq 0}$ since $\nabla^2 b_1(x)=2\geq 0$ and the domain $R_{\geq 0}$ is convex. Given that $f$ is non-negative and convex, the composition $b_1(f(x))=f^2(x)$ is thus convex.

b) Now, we show that $g^{-1}$ is convex.
Consider the function $b_2(x)=\frac{1}{x}$. It is nonicreasing and convex in $R_{>0}$ sinse $\nabla^2 b_2(x)=2x^{-3}\geq 0$ and the domain $R_{>0}$ is convex. Given that $g$ is positive and concave, the composition $b_2(g(x))=1/g(x)$ is thus convex.

c) Finally, we show that $h(\theta_1(x), b_2(x))=a_1(x)a_2(x)$ is convex, nondecreasing, for $a_1(x)$ and $a_2(x)$ non-negative and convex. As $a_1(x)\in R_{\geq 0}$ and $a_2(x)\in R_{\geq 0}$, clearly $h$ is nondecreasing. Consider the following inequality

$$a_1(\theta x+(1-\theta)y)\, a_2(\theta x+(1-\theta)y)\leq$$

$$\leq(\theta a_1(x)+(1-\theta)a_1(y))(\theta a_2(x)+(1-\theta)a_2(y))=$$

$$=\theta a_1(x)a_2(x)+(1-\theta)a_1(y)a_2(y)+\theta(1-\theta)(a_1(x)-a_1(y))\times$$

$$\times(a_2(x)-a_2(y))\leq\theta a_1(x)a_2(x)+(1-\theta)a_1(y)a_2(y)$$

where the last inequality follows from
~~$\theta(1-\theta)(a_1(x)-a_1(y))\leq\theta a_1(x)a_2(x)+(1-\theta)a_1(y)a_2(y)$~~
$\theta(1-\theta)(a_1(x)-a_1(y))(a_2(x)-a_2(y))$ being always nonnegative, this means that $h(a_1(x),a_2(x))=a_1(x)a_2(x)$ satisfy the Jensen's inequality, and thus is

a convex function.

Thus, as $f^2$ is nonnegative, $g^{-1}$ is positive, and both function are convex, and since $h$ is nondecreasing in each argument and convex when its arguments are positive, the composition $h(f^2(x), g^{-1}(x)) = f^2(x)/g(x)$ is convex.

<u>Problem 6:</u> Prove the information theoretic inequality: $D_{KL}(u, v) \geq 0 \quad \forall\, u, v$.
Also show that $D_{KL}(u, v) = 0$ iff $u = v$.

<u>Hint:</u> $D_{KL} = f(u) - f(v) - \nabla f^T(v)(u - v)$ where $f(v) = \sum_{i=1}^{n} v_i \log v_i$ is the negentropy of $v$.

<u>Solution:</u> Consider the original definition of the Kullback-Leibler divergence:

$$D_{KL}(u,v) = \sum_{i=1}^{n} \left(u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i\right) =$$

$$= \underbrace{\sum_{i=1}^{n} u_i \log u_i}_{f(u)} - \underbrace{\sum_{i=1}^{n} v_i \log v_i}_{f(v)} - \underbrace{\sum_{i=1}^{n} (1 + \log v_i)(u - v)}_{\nabla f^T(v)(u - v)}$$

where $f(v) = \sum v_i \log v_i$ is the negative entropy of $v \in R^n_{++}$.

Now, note that the negative entropy is convex: it is a nonnegative weighted sum of convex function $g(v_i) = v_i \log v_i$ (the second derivative satisfies $g''(v_i) = \frac{1}{v_i} > 0$ for $v_i \in R_{++}$). Being convex, the function $f$ satisfies the first-order Taylor approximation condition,

$$f(u) \geq f(v) + \nabla f^T(v)(u-v)$$

$$\Rightarrow f(u) - f(v) - \nabla^T(v)(u-v) \geq 0, \forall u, v \in \mathbb{R}^n_{++}$$

which directly proves that

$$D_{KL}(u,v) \geq 0 \quad \forall u, v \in \mathbb{R}^n_{++},$$

We now proceed to prove that

$$D_{KL}(u,v) = 0 \text{ iff } u = v,$$

$(\Rightarrow)$ is obvious. If $u=v$, then clearly

$$D_{KL}(u,v) = f(u) - f(v) - \nabla f^T(v)(u-v) = 0$$

$(\Leftarrow)$: Assume $D_{KL}(u,v) = 0$. without loss of generality, set $u = \frac{1}{1^T x}$ and $v = \frac{y}{1^T y}$, given some $x, y \in \mathbb{R}^n_{++}$ such that $\sum_{i=1}^n u_i = 1$ and $\sum_{i=1}^n v_i = 1$. That is $u$ and $v$ are probabilities, as it is in original KL divergence!

Considering that the function $-\log(x)$ is a convex function, we have the Jensen's inequality

$$0 = -D_{KL}(u,v) = \sum_{i=1}^n u_i \log\left(\frac{v_i}{u_i}\right) \leq$$

$$\leq \log\left(\sum_{i=1}^n u_i \left(\frac{v_i}{u_i}\right)\right) = \log 1 = 0.$$

Therefore, since the equality

$$\sum_{i=1}^n u_i \log\left(\frac{v_i}{u_i}\right) = \log\left(\sum_{i=1}^n u_i\left(\frac{v_i}{u_i}\right)\right) \text{ holds,}$$

by property of the Jensen's inequality we have that the arguments satisfy

$$\frac{v_1}{u_1} = \frac{v_2}{u_2} = \dots\dots = \frac{v_n}{u_n},$$

Since $D_{KL}(\mu, \nu) = 0$, this is only possible if $\frac{\nu_i}{\mu_i} = 1$ for all $0 \leq i \leq n$.
Thus, $\mu = \nu$.