1. f is convex iff $\mathrm{epi}(f)$ is convex

- $\Rightarrow$ direction: For any two points $\binom{x_1}{t_1}, \binom{x_2}{t_2} \in \mathrm{epi}(f)$, their convex combination ($\lambda \in [0,1]$)
$\lambda \binom{x_1}{t_1} + (1-\lambda)\binom{x_2}{t_2} = \begin{pmatrix} \lambda x_1 + (1-\lambda)x_2 \\ \lambda t_1 + (1-\lambda)t_2 \end{pmatrix}$ is also in $\mathrm{epi}(f)$ because
$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) \le \lambda t_1 + (1-\lambda) t_2$

- $\Leftarrow$ direction: For any two points $x_1, x_2 \in \mathrm{dom}(f)$, let their function value be $f(x_1) = t_1$, $f(x_2) = t_2$. We know that then points $\binom{x_1}{t_1}, \binom{x_2}{t_2} \in \mathrm{epi}(f)$. Hence their convex combination $\begin{pmatrix} \lambda x_1 + (1-\lambda)x_2 \\ \lambda t_1 + (1-\lambda)t_2 \end{pmatrix}$, $\lambda \in [0,1]$, is also in $\mathrm{epi}(f)$. So, $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda t_1 + (1-\lambda)t_2 = \lambda f(x_1) + (1-\lambda) f(x_2)$. Therefore, f is a convex function.

2. Given f is a convex function, we know that for any two points $x, y \in \mathrm{dom}(f)$,
$f(y) \ge f(x) + \nabla f(x)^T(y-x)$ and $f(x) \ge f(y) + \nabla f(y)^T(x-y)$.

Hence, $(\nabla f(x) - \nabla f(y))^T (x-y) = -(\nabla f(x)^T(y-x) + \nabla f(y)^T(x-y)) \ge -[(f(y)-f(x)) + (f(x)-f(y))]$
$= 0$

Thus, $\nabla f(x)$ is a monotone mapping.

The converse is not true, consider the counter example
$\psi(x) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ It is monotone because $(\psi(x)-\psi(y))^T(x-y) = (x-y)^T\left(\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}x - \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}y\right)$
$= (x-y)^T \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}(x-y)$
$= (x-y)^T \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}(x-y)$
$= (x_1-y_1+x_2-y_2)^2 \ge 0$

But $\psi(x)$ is not the gradient of any function f because f need to satisfy $\frac{\partial f}{\partial x_1 \partial x_2} = 2$, $\frac{\partial f}{\partial x_2 \partial x_1} = 0$, which is impossible.

3. By definition, $D_{KL}(u,v) = \sum_{i=1}^{n}(u_i \log\frac{u_i}{v_i} - u_i + v_i) = \sum_{i=1}^{n}(u_i \log u_i - u_i \log v_i - u_i + v_i)$

$$= \sum_{i=1}^{n}(u_i \log u_i - (v_i - v_i + u_i)\log v_i - u_i + v_i)$$

$$= \sum_{i=1}^{n}(u_i \log u_i - v_i \log v_i - (\log v_i + 1)(u_i - v_i))$$

$$= \sum_{i=1}^{n} u_i \log u_i - \sum_{i=1}^{n} v_i \log v_i - \sum_{i=1}^{n}(\log v_i + 1)(u_i - v_i)$$

$$= f(u) - f(v) - \nabla f(v)^T(u-v) \leftarrow \text{the hint given by the problem}$$

Now, we know $f(u) = \sum_{i=1}^{n} u_i \log u_i$ (the negative entropy) is strictly convex and differentiable because it is sum of n strictly convex function. Thus, we know

$$f(u) > f(v) + \nabla f(v)^T(u-v), \text{ for } u \neq v$$

which is $f(u) - f(v) - \nabla f(v)^T(u-v) > 0$ for $u \neq v$. Therefore $D_{KL}(u,v) > 0$ for $u \neq v$.

$D_{KL}(u,v) = 0$ when $u = v$.

4. (a). For function $f(X) = tr(X^{-1})$, we can verify convexity by considering an arbitrary line, given by $X = Z + tV$, where $Z, V \in S^n$. We define $g(t) = f(Z+tV)$, and restrict g to the interval of values of t for which $Z + tV \in S^n_{++}$. Without loss of generality, we can assume $t=0$ is inside this interval, i.e., $Z \in S^n_{++}$. We have

$$g(t) = tr((Z+tV)^{-1}) = tr(Z^{-1}(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})^{-1}) = tr(Q^TZ^{-1}Q(I+t\Lambda)^{-1}) = \sum_{i=1}^{n}(Q^TZ^{-1}Q)_{ii}(1+t\lambda_i)^{-1}$$

where we used the eigenvalue decomposition $Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}} = Q\Lambda Q^T$. In the end, we write $g(t)$ as a positive weighted sum of convex functions $\frac{1}{1+t\lambda_i}$, hence it is convex.

4.(b). Similarly, define $g(t) = f(Z+tV)$ where $Z \in S^n_{++}$ and $V \in S^n$.

$$g(t) = (\det(Z+tV))^{1/n} = (\det Z^{1/2} \cdot \det(I + tZ^{-1/2} V Z^{-\frac{1}{2}}) \cdot \det Z^{1/2})^{1/n}$$
$$= (\det Z)^{\frac{1}{n}} \left(\prod_{i=1}^{n} (1+t\lambda_i) \right)^{\frac{1}{n}}$$

where λ_i, $i=1,\dots,n$, are the eigenvalues of $Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}}$. From the last equality we know that g is a concave function of t on $\{t \mid Z+tV \succ 0\}$, since $\det Z > 0$ and the geometric mean $(\prod_{i=1}^{n} x_i)^{1/n}$ is concave on R^n_{++}.

5.(a) For any two points x, y in the interval, $fg(\lambda x + (1-\lambda)y) = f(\lambda x + (1-\lambda)y) g(\lambda x + (1-\lambda)y)$

$$\le (\lambda f(x) + (1-\lambda) f(y))(\lambda g(x) + (1-\lambda) g(y)) = \lambda f(x) g(x) + (1-\lambda) f(y) g(y) + \lambda(1-\lambda)(f(y)-f(x))(g(x)-g(y))$$

The third term $\lambda(1-\lambda)(f(y)-f(x))(g(x)-g(y)) \le 0$ because f and g are both nondecreasing (or both nonincreasing), hence

$$f(\lambda x + (1-\lambda)y) g(\lambda x + (1-\lambda)y) \le \lambda f(x) g(x) + (1-\lambda) f(y) g(y)$$

Thus, fg is convex.

(b). Similarly,
$$f(\lambda x + (1-\lambda)y) g(\lambda x + (1-\lambda)y) \ge (\lambda f(x) + (1-\lambda) f(y))(\lambda g(x) + (1-\lambda) g(y))$$
$$= \lambda f(x) g(x) + (1-\lambda) f(y) g(y) + \lambda(1-\lambda)(f(y)-f(x))(g(x)-g(y))$$
$$\ge \lambda f(x) g(x) + (1-\lambda) f(y) g(y)$$

because the third term $\lambda(1-\lambda)(f(y)-f(x))(g(x)-g(y)) \ge 0$ since f, g are nonincreasing and the other nondecreasing.

Thus, fg is concave.

(c) $\frac{1}{g}$ is convex, nondecreasing, and positive. Hence, using result from (a), we know $f \cdot \frac{1}{g}$ is convex.

6. (a). $g^*(y) = \sup_{x\in dom(g)} (y^T x - g(x)) = \sup_{x\in dom(g) \atop f} (y^T x - f(x) - c^T x - d) = \sup_{x\in dom(f)} ((y-c)^T x - f(x)) - d$

$= f^*(y-c) - d$

(b). The perspective of f is $g(x,t) = t\cdot f(\frac{x}{t})$, where $t>0$, $\frac{x}{t}\in dom(f)$.

$g^*(y,s) = \sup_{[\frac{x}{t}]\in dom(g)} (y^T x + st - g(x,t)) = \sup_{[\frac{x}{t}]\in dom(g)} (y^T x + st - t\cdot f(\frac{x}{t}))$

$= \sup_{[\frac{x}{t}]\in dom(g)} (\cancel{t} y^T \frac{x}{t} + ts - tf(\frac{x}{t})) = \sup_{t>0} t\,(s + \sup_{\frac{x}{t}\in dom(f)} (y^T \frac{x}{t} - f(\frac{x}{t})))$

$= \sup_{t>0} t(s + f^*(y)) = \begin{cases} 0, & \text{if } s+f^*(y)\le 0 \\ \infty, & \text{otherwise} \end{cases}$

(c) $g^*(y) = \sup_x (y^T x - g(x)) = \sup_x (y^T x - \inf_z f(x,z)) = \sup_{x,z} (y^T x - f(x,z))$

$= \sup_{x,z} (\begin{bmatrix} y \\ 0 \end{bmatrix}^T \begin{bmatrix} x \\ z \end{bmatrix} - f(x,z)) = f^*(y,0)$

For the second part, we can apply the result we just got to the application.
Here, $f(x,z) = \begin{cases} h(z), & Az+b=x \\ \infty, & \text{otherwise} \end{cases}$ Then $f^*(y,s) = \sup_{x,z} (y^T x + s^T z - f(x,z))$

$= \sup_{Az+b=x} (y^T x + s^T z - h(z)) = \sup_z (y^T(Az+b) + s^T z - h(z)) = y^T b + \sup_z (y^T A z + s^T z - h(z))$

$= y^T b + \sup_z ((A^T y + s)^T z - h(z)) = y^T b + h^*(A^T y + s)$

Thus, $g^*(y) = f^*(y,0) = y^T b + h^*(A^T y)$

6.(d). By definition we have $f^*(y) = \sup_x (y^Tx - f(x))$.

This means for any ~~a fixed~~ y, $y^Tx - f(x) \le f^*(y) \Rightarrow f(x) \ge y^Tx - f^*(y)$ for all x.

This means an affine function $y^Tx + b$, where $b \le -f^*(y)$, is a global underestimator of $f(x)$.

Hence, by result of exercise 3.28,

$$f(x) = \sup_{y \in \mathrm{dom}(f^*)} (y^Tx - f^*(y)) = \cancel{f^{**}(y)} \sup_{y \in \mathrm{dom}(f^*)} (x^Ty - f^*(y)) = f^{**}(x)$$