Convex Optimization and Modeling	Jun.-Prof. Matthias Hein
Solution of Exercise Sheet 3 - 28.4.2010	

Exercise 5 - Convex, concave, quasiconvex and quasiconcave functions

a. **(3 Points)** Exercise 3.16a)-c) Determine for each function all the classes (convex, concave, quasiconvex, quasiconcave) to which it belongs.

b. **(2 Points)** Exercise 3.19a.

c. **(2 Bonus Points)** Exercise 3.13. (use the hint !). The Kullback-Leibler divergence $KL(p||q)$ is a classical measure of "distance" between two probability measures p and q. In the exercise it is used in a more general form for strictly positive measures on $\mathbb{R}^n_{++}$. Let $u, v \in \mathbb{R}^n_{++}$ (that means $u_i > 0$, $\forall\, i = 1, \ldots, n$)

$$KL(u||v) = \sum_{i=1}^{n} \Big[u_i \log\Big(\frac{u_i}{v_i}\Big) - u_i + v_i \Big].$$

The KL-divergence is a special case of the so called **Bregman-divergences** which have recently attracted some interest.

Solution:

a.

- $f(x) = e^x - 1$ on $\mathbb{R}$. We have $\frac{\partial^2 f}{\partial x^2} = e^x > 0$ for all $x \in \mathbb{R}$ and thus f is strictly convex and quasiconvex but not concave. It is however also quasiconcave.
- $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}_+ +^2$. The Hessian of f is

$$Hf(x_1, x_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

  This matrix has eigenvalues 1 and -1 and thus is indefinite. Therefore f is neither convex nor concave. The function is not quasiconvex as $(1, 0)$ and $(0, 1)$ are contained in each sublevel set but $(\frac{1}{2}, \frac{1}{2})$ only if the level is larger than $\frac{1}{4}$. The superlevelsets are convex (see Exercise 4) and thus f is quasiconcave.
- $f(x_1, x_2) = \frac{1}{x_1\, x_2}$ on $\mathbb{R}_+ +^2$ has Hessian,

$$Hf(x_1, x_2) = \frac{1}{x_1\, x_2} \begin{pmatrix} \frac{2}{x_1^2} & -\frac{1}{x_1\, x_2} \\ -\frac{1}{x_1\, x_2} & \frac{1}{x_2^2} \end{pmatrix}.$$

  The determinant $\det Hf = \frac{4}{x_1^2 x_2^2} + \frac{1}{x_1^2 x_2^2} = \frac{5}{x_1^2 x_2^2}$ is equal to the product of the eigenvalues and is positive. Thus both eigenvalues have the same sign and as the trace is also positive - they are both positive and Hf is positive-definite. Thus f is strictly convex and quasiconvex but not concave or quasiconcave.

b. The function $f(x) = \sum_{i=1}^{r} \alpha_i x_[i]$, where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_r \geq 0$ and $x_[s]$ is the s-th largest component of the vector x. We have,

$$\begin{aligned} f(x) &= \sum_{i=1}^{r} \alpha_i x_[i] = \alpha_r \sum_{i=1}^{r} x_[i] + \sum_{i=1}^{r} (\alpha_i - \alpha_r) x_[i] \\ &= \alpha_r \sum_{i=1}^{r} x_[i] + (\alpha_{r-1} - \alpha_r) \sum_{i=1}^{r-1} x_[i] + \ldots + (\alpha_1 - \alpha_2) \sum_{i=1}^{1} x_[i]. \end{aligned}$$

As $\sum_{i=1}^{k} x_{[i]}$ is convex in x and f is a non-negative linear combination of these functions, f is convex.

c. The hint said that

$$KL(u||v) = f(u) - f(v) - \langle \nabla f|_v, u - v \rangle \,,$$

where $f(v) = \sum_{i=1}^{n} v_i \log v_i$ is the negative entropy of v. We have

$$\frac{\partial}{\partial v_k} f = \log v_k + 1, \qquad \frac{\partial^2}{\partial v_k \partial v_l} f = \frac{1}{v_l} \delta_{lk}.$$

Since the domain of KL is on $\mathbb{R}^n_{++}$ (which is convex) we conclude that the Hessian of f is positive definite (diagonal matrix with positive entries). From the first order condition of a differentiable convex function we derive

$$f(u) \geq f(v) + \langle \nabla f|_v, u - v \rangle \,.$$

which implies $KL(u||v) \geq 0$. Moreover, since f is strictly convex we have in the above inequality only equality if $u = v$ and thus

$$KL(u||v) = 0 \quad \text{if and only if } u = v.$$

Exercise 7 - Subdifferential

- **(4 Points)** Derive the subdifferential of the general graph-based total variation energy functional,

$$F(f) = \sum_{i,j=1}^{n} w_{ij} |f_i - f_j|,$$

where $w \in \mathbb{R}^{n \times n}$ are the non-negative weights of a graph with n nodes and $f \in \mathbb{R}^n$ is a function on the nodes.

- **(2 Points)** Derive the subdifferential of $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = \|x\|_\infty$.

Hints:

- for a) it might be helpful to first derive the subdifferential of the function $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = |x - y|$ using the chain rule introduced in the lecture.
- for b) make a case distinction if $x = 0$ or $x \neq 0$. For $x = 0$ the dual norm introduced in Exercise 1 is helpful and for $x \neq 0$ the direct use of the definition is required.

Solution:

- In the lecture the following rules were introduced for convex functions f. For $f(x) = g(Ax+b)$ we have

$$\partial f(x) = A^T \partial g(Ax + b),$$

and for $f(x) = \sum_{i=1}^{r} \alpha_i f_i(x)$ where all f_i are convex we have,

$$\partial f(x) = \sum_{i=1}^{r} \alpha_i \, \partial f_i(x).$$

Using the hint we have $f(x, y) = |x - y| = \left| \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right|$ and thus,

$$\partial f(x, y) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \operatorname{sign}(x - y).$$

where we use the set-valued mapping,

$$\operatorname{sign}(x) = \begin{cases} -1, & \text{if } x < 0, \\ [-1, 1], & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases}$$

Thus we get in total for $F(f) = \sum_{i,j=1}^{n} w_{ij} |f_i - f_j|$,

$$\partial F(f) = \sum_{i,j=1}^{n} w_{ij} (e_i - e_j) \operatorname{sign}(f_i - f_j),$$

where e_i is the i-th unit vector in $\mathbb{R}^n$. In components, we get

$$\begin{aligned} \partial F(f)_r =& \sum_{j=1}^{n} w_{rj} \operatorname{sign}(f_r - f_j) - \sum_{i=1}^{n} w_{ir} \operatorname{sign}(f_i - f_r) \\ =& \sum_{j=1}^{n} w_{rj} \big(\operatorname{sign}(f_r - f_j) - \operatorname{sign}(f_j - f_r) \big) \\ =& 2 \sum_{j=1}^{n} w_{rj} u_{ij}, \end{aligned}$$

where $u_{ij} = -u_{ji}$ and

$$u_{ij} = \operatorname{sign}(f_i - f_j).$$

Note, that this is *not* the same as,

$$\partial F(f)_r = 2 \sum_{j=1}^{n} w_{rj} \operatorname{sign}(f_r - f_j),$$

as in the later expression the set-valued part need not be anti-symmetric.

- The subdifferential of $f(x) = \|x\|$ at $x = 0$ can be deduced from $\|x\| = \sup_{\|u\|_* \leq 1} \langle u, x \rangle$ and the definition $\|y\| \geq \|0\| + \langle \partial f|_0, (y - 0) \rangle = \langle \partial f|_0, y \rangle$. Thus at $x = 0$ we have

  $$\partial f|_0 = \{ u \in \mathbb{R}^n \mid \|u\|_* \leq 1 \}.$$

  For the infinity norm the dual norm is the 1-norm. Thus we have using the definition from Exercise 1,

  $$\partial \left\| \cdot \right\|_\infty (0) = \{ u \in \mathbb{R}^n \mid \|u\|_1 \leq 1 \}.$$

  Now for $x \neq 0$, the definition of $\partial f(x)$ requires for all $y \in \mathbb{R}^n$,

  $$\|y\| - \|x\| \;\geq\; \langle \partial f(x), y - x \rangle .$$

  In particular for $y = (1 + \lambda) x$, we have

  $$\lambda \|x\| \;\geq\; \lambda \langle \partial f(x), x \rangle .$$

  For $\lambda > 0$ we have $\|x\| \geq \langle \partial f(x), x \rangle$ and for $\lambda < 0$ we get $\|x\| \leq \langle \partial f(x), x \rangle$ and thus we conclude

  $$\|x\| \;=\; \langle \partial f(x), x \rangle .$$

Plugging this result into the definining property of $\partial f(x)$ we obtain

$$\|y\| \geq \langle \partial f(x), y \rangle \,, \quad \forall\, y \in \mathbb{R}^n,$$

from which we deduce using the result at $x = 0$, that $\partial f(x) \subset \{u \in \mathbb{R}^n \;\big|\; \|u\|_* \leq 1\}$. However, using

$$\langle u, v \rangle \leq \|u\| \, \|v\|_* \,,$$

we can achieve $\langle u, v \rangle = \|u\|$ if and only if $\|v\|_* = 1$ and thus we have finally,

$$\partial f(x) = \{u \in \mathbb{R}^n \;\big|\; \langle u, x \rangle = \|x\| \text{ and } \|u\|_* = 1\}.$$

From Exercise 1 one we know that $\langle u, x \rangle = \|x\|_\infty$ if and only if $u = \sum_{i \in R(x)} \lambda_i v^{(i)}$, where

$$v_s^{(i)} = \begin{cases} \operatorname{sign}(x_s), & \text{if } s = i, \\ 0, & \text{else.} \end{cases}$$

and $R(x) = \{i \;\big|\; |x_i| = \|x\|_\infty\}$ and $\lambda_i \geq 0$ and $\sum_{i \in R(x)} \lambda_i = 1$. Note that $\left\|v^{(i)}\right\|_1 = 1$ and thus for $x \neq 0$,

$$\partial \left\|\cdot\right\|_\infty (x) = \operatorname{conv}\{v^{(i)} \;\big|\; i \in R(x)\}.$$

Related to the infinity-norm one can show that given that $f(x) = \max_{i=1,\ldots,n} f_i(x)$, where all f_i are convex, one has

$$\partial f(x) = \operatorname{conv}\Big(\bigcup \{\partial f_k(x) \;\big|\; f_k(x) = f(x)\}\Big).$$

The inclusion in one direction can be shown as follows. Denote by $R(x)$ the "active" set at x, that is $R(x) = \{i \mid f_i(x) = f(x)\}$. Then for all $z \in \operatorname{dom} f$ and for any $\lambda_i \geq 0$, $i = 1, \ldots, n$ with $\sum_{i \in R(x)} \lambda = 1$,

$$f(z) \geq \sum_{i \in R(x)} \lambda_i f_i(z) \geq \sum_{i \in R(x)} \big(\lambda_i f_i(x) + \langle \partial f_i(x), z - x \rangle\big) = f(x) + \sum_{i \in R(x)} \lambda_i \langle \partial f_i(x), z - x \rangle \,,$$

Thus it follows that $\operatorname{conv} \bigcup \{\partial f_k(x) \;\big|\; f_k(x) = f(x)\} \subset \partial f(x)$.

In general, one can state that it is much easier to verify that a certain vector v is a subgradient of f than the identification of the whole subdifferential.