In this homework, we work on exercises from textbook including level sets of convex, concave, quasi-convex, quasi-concave functions (3.1, 3.2), second-order conditions for convexity on affine sets (3.9), Kullback-Leibler divergence (3.13), saddle points of convex-concave functions (3.14) determination of convex, concave, quasi-convex, quasi-concave functions (3.16), conjugate functions (3.36), and gradient and Hessian of conjugate functions (3.40). Extra assignments are given on Kullback-Leibler divergence, and softmax functions.

Exercises are graded by completion, assignments are graded by content. We may just grade a subset of the problems.

I. Exercises from textbook chapter 3 (8 pts)


II. Assignments (42 pts)

II. 1. Kullback Leibler Divergence.
Let us define Kullback Leibler (KL) divergence as a function \( D_{KL}(u, v) = \sum_i u_i \log(u_i/v_i) - u_i + v_i \), with the constraints \( 1^T u = 1, 1^T v = 1 \), where \( u, v \in \mathbb{R}_{++}^n \). Derive the following that honors the constraints. (hint: Convert the feasible set from qualification to enumeration expression) (17 pts)

II.1.1. Design a numerical example to show the value of the function. For this example, let us set \( n = 5 \). (4 pts)

II.1.2. Derive the first order derivative of the KL divergence, i.e. \( \nabla_{u,v} D_{KL}(u, v) \). (4 pts)

II.1.3. Derive the Hessian matrix of the KL divergence, i.e. \( \nabla_{u,v}^2 D_{KL}(u, v) \). (4 pts)

II.1.4. Is KL divergence a convex function? Show your proof if it is convex. (5 pts)

II. 2. Conjugate Functions and Softmax Functions. (25 pts)

Find the conjugate function of functions \( f_1(x), f_2(x), f_3(x) \).

II.2.1. \( f_1(x) = ||x||_p, p = 100 \), where \( x \in \mathbb{R}_{++}^n \). (6 pts)

II.2.2. \( f_2(x) = \log \sum_i \exp(x_i/\gamma) \), where \( x \in \mathbb{R}^n \). (6 pts)

II.2.3. \( f_3(x) = \sum_i x_i \exp(x_i/\gamma) / \sum_i \exp(x_i/\gamma) \), where \( x \in \mathbb{R}^n \). (6 pts)
II.2.4. Suppose that we use the above functions as a softmax function. Derive the worst error bounds, i.e. $|\max_i x_i - f(x)|$, of the above functions. (7 pts)