Problem Sheet on Convex Optimisation

---

1. Show that the following functions are convex:

   (a) $f(x) = |x|$, $x \in \mathbb{R}$.

   (b) $f(x) = x \log x$, $x \in (0, \infty)$. Logarithms are natural unless otherwise specified.

   (c) $f(x) = x^2 y^2$, $(x, y) \in \mathbb{R}^2$.

   (d) $f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2$, $\mathbf{x} \in \mathbb{R}^n$, where A is an $m \times n$ matrix, $\mathbf{b} \in \mathbb{R}^m$, and $\| \cdot \|$ denotes the Euclidean norm of a vector. In order to show that f is convex, you may need to show that a certain matrix is positive semi-definite.

   (e) $f(x) = \begin{cases} x, & x > 0, \\ y, & x = 0, \\ +\infty, & x < 0, \end{cases}$ where $y \geq 0$ is arbitrary.

2. (a) Let $\mathbb{S}$ denote the set of all real symmetric $n \times n$ matrices. Show that $\mathbb{S}$ is a convex subset of $\mathbb{R}^{n \times n}$.

   (b) Recall that all eigenvalues of a symmetric matrix are real. Hence, they can be ordered from largest to smallest. Let $\lambda_{\max}(A)$ denote the largest eigenvalue of the real symmetric matrix A. Let $f : \mathbb{S} \to \mathbb{R}$ be defined by $f(A) = \lambda_{\max}(A)$. Show that f is a convex function.

   Hint. Use the Rayleigh-Ritz formula, which states that

$$\lambda_{\max}(A) = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T A \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x}.$$

3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a convex function, and let C be a convex subset of $\mathbb{R}$. Define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \min_{y \in C} f(x, y).$$

   Show that g is a convex function.

   Hint. Start from the definition of convexity. Pick x_0 and x_1 in $\mathbb{R}$, $\alpha \in [0, 1]$ and define $x_\alpha = (1 - \alpha)x_0 + \alpha x_1$. Suppose that the minimum in the definition of $g(x_0)$ (respectively $g(x_1)$, $g(x_\alpha)$) is attained at $y_0 \in C$ (respectively at y_1, y^α. The use of the superscript for y^α is to make it clear that no assumption is made about the value of y^α, beyond that it is in C; specifically, y^α need not be equal to $y_\alpha = (1 - \alpha)y_0 + \alpha y_1$.).

4. (a) The linear regression problem in statistics has data in the form of an $n \times p$ matrix X of n observations of p independent variables, and a $n \times 1$ matrix (i.e. a column vector of length n) $\mathbf{y}$ of observations of a single dependent variable. The objective is to find a p-vector of coefficients β so as to solve the following least-squares problem:

$$\min_{\beta \in \mathbb{R}^p} f(\beta) := \|X\beta - \mathbf{y}\|^2,$$

where, as usual, $\|\mathbf{z}\|$ denotes the Euclidean norm of the vector $\mathbf{z}$; recall that $\|\mathbf{z}\|^2 = \mathbf{z}^T\mathbf{z}$.

   i. Show that $f : \mathbb{R} \to \mathbb{R}$ is a convex function. You may need to show that a certain matrix is positive semi-definite.

   ii. Use the first order sufficient condition for unconstrained optimisation to find the optimal β. You may assume that the matrix $X^T X$ is full rank (hence invertible).

   (b) Linear regression works well in practice when the number of observations n is much bigger than the number of variables p. In many "Big Data" problems, this is not the case. Often p is close to n in size, which can result in overfitting (finding spurious statistically significant dependencies just be chance), or p can even be bigger than n, in which case the least squares problem doesn't have a unique solution for β. One approach to dealing with such problems is to assume that β is sparse, i.e., that it has few non-zero coefficients. Often, there is good intuitive justification for this assumption. This motivates us to try and solve the least squares problem with a sparsity constraint (constraint on the number of non-zero elements in β). However, this problem is non-convex and, consequently, intractable. Instead, a common approach is to consider a "convex relaxation" of this problem.

   LASSO, or ℓ_1-penalised least squares, seeks to solve the folllowing modification of the least-squares optimisation problem:

$$\min_{\beta \in \mathbb{R}^p} g(\beta) := \|X\beta - \mathbf{y}\|^2 + \lambda\|\beta\|_1, \quad \text{where } \|\beta\|_1 := \sum_{i=1}^{p} |\beta_i| \text{ and } \lambda > 0.$$

   The quantity $\|\beta\|_1$ is called the ℓ_1 norm of β, and is simple the sum of the absolute values of its co-ordinates. The constant $\lambda > 0$ is the penalty on the ℓ_1 norm, and trades off between how accurately we want to fit the data (the first term in g), and how much we penalise large values of β, as measured by its ℓ_1 norm.

   Show that g is a convex function. You may use answers to previous questions to save yourself some work.

5. A quadratic programming (QP) problem involves the minimisation of a quadratic function subject to linear constraints. The general form is:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\mathbf{x}^T Q\mathbf{x} + \mathbf{c}^T\mathbf{x} \text{ subject to } A\mathbf{x} \leq \mathbf{b}.$$

Here, Q is a real symmetric $n \times n$ matrix, $\mathbf{c} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$; inequalities between vectors are to be interpreted to hold for each coordinate.

(a) Suppose Q is positive semi-definite. Then show that the QP is a convex optimisation problem, i.e., that the objective function is convex, and that the set of $\mathbf{x} \in \mathbb{R}$ satisfying the constraints is a convex set.

(b) Write down the Lagrangian for the problem, and formulate the dual problem.

(c) Write down the KKT conditions for optimality.

(d) Starting from an initial value $\mathbf{x}^0$ for the unconstrained version of the above problem, compute the value $\mathbf{x}^1$ you would obtain after one iteration of gradient descent with exact line search.

Do the same for one step of the Newton method, assuming Q is positive definite, and hence invertible.

What is the exact solution to the minimisation problem in this case (still unconstrained)?