Optimisation **Part IB - Easter 2018**

Example Sheet 1

Lecturer: Quentin Berthet

1. Properties of convex sets

a) Show that an intersection of convex sets is a convex set.

b) We recall that a norm N in $\mathbf{R}^n$ is a function from $\mathbf{R}^n$ to $\mathbf{R}_{\geq 0}$ such that

- For all $v \in \mathbf{R}^n$, and $a \in \mathbf{R}$, $N(ax) = |a|N(x)$.

- For all $v, w \in \mathbf{R}^n$, $N(v + w) \leq N(v) + N(w)$ (triangle inequality).

- $N(v) = 0$ implies $v = 0$.

Show that the unit ball of any norm is a convex set.

c) Show that a half space, defined for some $v \in \mathbf{R}^n$ and $t \in \mathbf{R}$, by $\{x \in \mathbf{R}^n \,:\, v^\top x \leq t\}$ is convex. Show that for any $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$, the set $\{x \in \mathbf{R}^n \,:\, Ax = b\}$ is convex. Deduce that the set $\{x \in \mathbf{R}^n \,:\, Ax = b\,,\; x \geq 0\}$ is convex.

2. Properties of convex functions

a) For a convex set $C \subseteq \mathbf{R}^n$ and a function f on C, the *epigraph* is the set of points in $C \times \mathbf{R}$ "above the graph", formally defined by $\{(x, y) \in C \times \mathbf{R} \,:\, y \geq f(x)\}$.

Show that a function is convex if and only if its epigraph is convex.

b) The level sets of a real-valued function on a set C are defined, for all $t \in \mathbf{R}$, as $\{x \in C \,:\, f(x) \leq t\}$. Show that the level sets of a convex function are convex. Is it true that a function with convex level sets is always convex?

c) Let $f_1, \ldots, f_k$ be convex functions on C and $\lambda_1, \ldots, \lambda_k$ be nonnegative reals. Show that the function f defined on C by $f(x) = \lambda_1 f_1(x) + \ldots + \lambda_k f_k(x)$ is convex.

d) Show that the supremum of convex functions is a convex function. Show that the absolute value function, mapping x to $|x|$, is convex.

e) Let $h : \mathbf{R} \to \mathbf{R}$ and $g : C \to \mathbf{R}$ be real valued functions, and $f = h \circ g$, i.e. $f(x) = h(g(x))$. Show that if g is convex and h is convex and nondecreasing, then f is convex.

f) Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function. Show that $h : \mathbf{R}^m \to \mathbf{R}$ defined for $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$ by $h(x) = f(Ax - b)$ is a convex function.

Remarks or errors can be addressed to q.berthet@statslab.cam.ac.uk

3. a) Show that if f is a differentiable and convex function, then

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)\,.$$

Deduce that

$$(\nabla f(x) - \nabla f(y)^\top (x - y) \geq 0\,.$$

For a convex function with global minimum x^*, taking $y = x^*$, what interpretation does this give about gradient descent?

4. Examples of convex sets and functions.

a) Show that the unit simplex in $\mathbf{R}^n$ defined by

$$\{x \in \mathbf{R}^n\; x_1 + \ldots + x_n = 1\,,\; x \geq 0\}$$

is a convex set. It is the set of all probability distributions on n elements. Show that for any given function $f : [n] \to \mathbf{R}$, the subset of distributions such that $\mathbf{E}f(X) \in [a, b]$, where $X \sim p$, is a convex set.

b) Show that the set of semidefinite positive matrices, and the set of positive definite matrices, are convex sets.

c) Show that the function defined by $f(x, y) = x^2/y$, for $y > 0$, is convex.

d) Show that the function from symmetric real matrices to $\mathbf{R}$ mapping M to its largest eigenvalue $\lambda_1(M)$ is convex.

5. Properties of smooth and/or strongly convex functions.

a) Show that if f is a β-smooth function, it holds that

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{\beta}{2}\|y - x\|^2\,.$$

b) Show that if a function f is twice differentiable, with semidefinite positive Hessian everywhere, then

- If the eigenvalues of the Hessian are lower bounded by α, then f is α-strongly convex.

- If the eigenvalues of the Hessian are upper bounded by β, then f is β-smooth.

Show that if the eigenvalues of the Hessian of f are in $[\alpha, \beta]$, the eigenvalues of the Hessian of $f(x) - \alpha\|x\|^2/2$ are in $[0, \beta - \alpha]$.

6. Application in statistics and machine learning

We observe, for n properties in a city, the price Y_i of the property, and a vector of variables $X_i \in \mathbf{R}^p$ of known information about the property (number of rooms, distance to public transport, rate of crime in the area, etc). We would like to find the vector

$\beta \in \mathbf{R}^p$ that creates the best linear fit $X_i^\top \beta$ for the price y_i. This can be done by taking the *least-squares*, which has a statistical motivation

$$\min_{\beta \in \mathbf{R}^p} f(\beta) = \min_{\beta \in \mathbf{R}^p} \sum_{i=1}^n (y_i - X_i^\top \beta)^2$$

a) Write $f(\beta)$ using the matrix $X \in \mathbf{R}^{n \times p}$ with i-th row $X_i^\top$ and of the vector y.

b) Show that f is a convex function.

c) If $n > p$, and X has rank p, show that f is α-smooth and β-strongly convex, for α and β depending on the eigenvalues of $X^\top X$. In this case, compute explicitly the minimizer β^* of f.

d) Compute an iterate of the gradient descent algorithm β_{t+1}, as a function of $\beta_t \in \mathbf{R}^p$. Using a result from the lecture, describe the behaviour of the iterates of this algorithm, when $\eta = 2/(\alpha + \beta)$.

e) What happens when applying Newton's algorithm to minimize this function? What is the advantage of using gradient descent rather than this algorithm, or computing explicitly the minimum?

7. a) Given constants $b_1,\ b_2$ such that $b_1 - e^{-b_2} \geq 0$ use the Lagrangian method to

maximise $2\tan^{-1} x_1 + x_2$ subject to $x_1 + x_2 \leq b_1,\ -\ln x_2 \leq b_2,\ x_1 \geq 0,\ x_2 \geq 0.$

[*Hint*: There will be two cases to check depending the constants b_1 and b_2.]

b) Given constant $b \geq 0$, solve

$$\min \frac{1}{x_1 + 1} + \frac{1}{x_2 + 2} \quad \text{such that} \quad x_1 + x_2 = b\, x_1, x_2 \geq 0\,.$$

8. Let $g : \mathbf{R}^n \to \mathbf{R}^m$ and $f : \mathbf{R}^n \to \mathbf{R}$ be convex functions and $\mathcal{X}$ be a convex set, and for every b let

$$\varphi(b) = \inf\{f(x) : \quad g(x) \leq b\,,\ x \in \mathcal{X}\}.$$

Assuming $\varphi(b)$ is finite for all b, show that the function φ is convex.

9. We denote by Δ_n the unit simplex defined in question 4. For any $c \in \mathbf{R}^n$ with distinct entries and $\alpha > 0$, we consider the optimization problem

$$\min_{x \in \Delta_n} c^\top x + \frac{1}{\alpha} \sum_{i=1}^n x_i \log(x_i)\,.$$

a) Solve this problem, using the Lagrangian sufficiency theorem or otherwise.

b) Let x_α^* be the solution to this problem, find its limit when α goes to 0, and $+\infty$.

c) What is the solution to $\min_{x \in \Delta_n} c^\top x$? Compare this result with the question above.