UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Wednesday, 8 June, 2022 9:00 am to 11:00 am

PAPER 339

TOPICS IN CONVEX OPTIMISATION

Before you begin please read these instructions carefully

Candidates have TWO HOURS to complete the written examination.

Attempt no more than **TWO** questions.
There are **THREE** questions in total.
The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury tag
Script paper
Rough paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and assume that $\min_{x\in\mathbb{R}^n} f(x)$ exists and is finite. Recall that the proximal operator associated to f is defined by

$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{u\in\mathbb{R}^n} \left\{ f(u) + \frac{1}{2}\|u - x\|_2^2 \right\}.$$

(a) Give the definition of a *subgradient* of f at a point $x \in \mathbb{R}^n$. State (without proof) a necessary and sufficient condition to have $u = \mathrm{prox}_f(x)$ in terms of the subdifferential of f. [5]

(b) The *proximal point* algorithm for minimizing f on $\mathbb{R}^n$ proceeds as follows: starting from any $x_0 \in \mathbb{R}^n$ take

$$x_{k+1} = \mathrm{prox}_{tf}(x_k), \qquad k = 0, 1, \ldots \tag{1}$$

where $t > 0$ is some fixed step size.

(i) Show that for any integer $k \geqslant 0$ and any $u \in \mathbb{R}^n$ we have

$$f(u) \geqslant f(x_{k+1}) + \frac{1}{t}(x_k - x_{k+1})^T(u - x_{k+1}).$$ [5]

(ii) Deduce that for any $u \in \mathbb{R}^n$ we have

$$t(f(x_{k+1}) - f(u)) \leqslant \frac{1}{2}(\|u - x_k\|_2^2 - \|u - x_{k+1}\|_2^2).$$ [10]

(iii) Deduce that algorithm (1) satisfies, for any integer $k \geqslant 1$

$$f(x_k) - f^* \leqslant \frac{\|x_0 - x^*\|_2^2}{2kt}.$$ [10]

(c) For $t > 0$ define the function

$$M_f(x) = \min_{u\in\mathbb{R}^n} \left\{ f(u) + \frac{1}{2t}\|u - x\|_2^2 \right\}.$$

(i) State the definition of the *Fenchel conjugate* of a function h. [4]

(ii) Show that M_f is always smooth, even if f is not. [*Hint: write* $M_f(x) = \frac{1}{2t}\|x\|_2^2 - \frac{1}{t}g^*(x)$ *where* g *is a strongly convex function, and* g^* *denotes the Fenchel conjugate of* g]. [8]

(iii) What is $\nabla M_f(x)$? Deduce that (1) is a gradient descent algorithm for a well-chosen function that you should specify. You may use (without proof) the expression for the gradient of the Fenchel conjugate of a strongly convex function that we saw in lectures. [8]

2

(a) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function, and consider the optimization problem

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{subject to} \quad Ax = b$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. State the definition of the *Lagrangian* and the *dual problem*. Explain what is *weak duality*, *strong duality* and give sufficient conditions to have strong duality. [10]

Let Q be a real symmetric positive definite $n \times n$ matrix and consider the optimization problem

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2} x^T Q x \quad \text{subject to} \quad Ax \geqslant b, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $Ax \geqslant b$ is interpreted component wise, i.e., $(Ax)_i \geqslant b_i$ for all $i = 1, \ldots, m$.

(b) By rewriting problem (1) as

$$\min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} \frac{1}{2} x^T Q x + I(s) \quad \text{subject to} \quad Ax - b - s = 0,$$

where $I(s)$ is the indicator function of $\mathbb{R}^n_+$ ($I(s) = 0$ if $s \geqslant 0$ and $+\infty$ otherwise), write the Lagrangian dual of the problem, and show that it has the form

$$\max_{z \in \mathbb{R}^m} \quad h(z) \quad \text{subject to} \quad z \geqslant 0 \tag{2}$$

where $h(z)$ is a concave quadratic function that you should specify explicitly in terms of A, b and Q. Give explicit sufficient conditions (in terms of A and b only) for (2) to have the same optimal value as (1). [10]

(c) (i) Write down the projected gradient ascent method to solve (2). Your iterations should be completely explicit. [10]

(ii) Is the function $-h(z)$ strongly convex? If yes, give (without proof) a lower bound on the strong convexity parameter. Also give (without proof) an upper bound on the Lipschitz constant of the gradient of h. [5]

(iii) Give, without proof, an upper bound on the number of iterations k of projected gradient ascent that are needed to reach a point z_k such that $h^* - h(z_k) \leqslant \epsilon$, where h^* is the optimal value of (2). Your bound should have the form $k = O(c\phi(\epsilon))$ where c is a constant and ϕ is a function that you should both specify. How many iterations would be needed if we used the *fast* projected gradient method of Nesterov? [5]

(iv) Write the projected gradient descent for the original problem (1). Explain why using the projected gradient ascent for (2) might be preferable over the projected gradient descent algorithm for (1). [10]

3

(a) State the definition of a *firmly nonexpansive map* $T : \mathbb{R}^n \to \mathbb{R}^n$. Show that the proximal operator prox_f of a convex lower semi-continuous function f is firmly nonexpansive. [8]

(b) The Douglas-Rachford algorithm to minimize the sum of two convex functions $f(x) + h(x)$ is given by:

$$\begin{cases} x_{k+1} &= \mathrm{prox}_f(y_k - z_k) \\ y_{k+1} &= \mathrm{prox}_h(x_{k+1} + z_k) \\ z_{k+1} &= z_k + (x_{k+1} - y_{k+1}). \end{cases} \tag{1}$$

Show that by letting $w_{k+1} = x_{k+1} + z_k$, the iterates can be written as $w_{k+1} = T(w_k)$ for some map $T : \mathbb{R}^n \to \mathbb{R}^n$. Show that T is firmly nonexpansive. [8]

(c) Let $C, D \subset \mathbb{R}^n$ be two closed convex sets. We are interested in finding a point in the intersection $C \cap D$, which we assume is nonempty, using only the Euclidean projection operators P_C and P_D on C and D respectively.

(i) Show how the Douglas-Rachford algorithm can be used to achieve this. [4]

(ii) We now consider an alternative algorithm. Let $d_D(x) = \min_{y \in D} \left\{ \frac{1}{2} \|y - x\|_2^2 \right\}$. Show that d_D is convex, and that a point in $C \cap D$ can be obtained by finding a solution to the convex optimization problem

$$\min\{d_D(x) : x \in C\}. \tag{2}$$ [8]

(iii) Show that d_D is 1-smooth with respect to the Euclidean norm, and give an expression for $\nabla d_D(x)$. [*Hint: express $d_D(x) = \|x\|_2^2/2 - g^*(x)$ where g^* is the Fenchel conjugate of a 1-strongly convex function.*] [8]

(iv) Write down the projected gradient descent method for (2); your answer should involve only P_C and P_D, the Euclidean projection operators on C and D. [4]

(d) Assume now we have ℓ closed convex sets $C_1, \ldots, C_\ell \subset \mathbb{R}^n$, and we are interested in finding a point in their common intersection $C_1 \cap \cdots \cap C_\ell$, which we assume is nonempty. Show that this problem can be reduced to a problem of finding a point in the common intersection of a convex set $C \subset \mathbb{R}^{n\ell}$ and a subspace $D \subset \mathbb{R}^{n\ell}$. Show that the projections on C and D can be computed explicitly, in terms of the projection operators $P_{C_1}, \ldots, P_{C_\ell}$ on $C_1, \ldots, C_\ell$ respectively. [10]

END OF PAPER