UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Tuesday, 31 May, 2016 9:00 am to 11:00 am

PAPER 325

SET-VALUED ANALYSIS AND OPTIMISATION

*Attempt no more than **TWO** questions.*

*There are **THREE** questions in total.*

The questions carry equal weight.

In your answers, you may use any of the theorems and lemmas from the lecture notes, explicitly stating their use. The exception is when you are asked to prove that particular theorem or lemma; then you must provide the proof. You may however refer to any other theorems or lemmas used in the proof.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script 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

This question concerns convex analysis, and in particular the notions of infimal convolution and conjugacy. We start by recalling the basic definitions by answering

(i) What are the definitions of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ being convex and proper? What is the domain of f? How is the subdifferential of a convex function defined?

The infimal convolution $f \mathbin{\varnothing} g$ of two convex functions $f, g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as

$$(f \mathbin{\varnothing} g)(x) := \inf_{y \in \mathbb{R}^n} \big(f(x-y) + g(y)\big), \quad (x \in \mathbb{R}^n).$$

The convex conjugate of f is defined as

$$f^*(y) := \sup_{x \in \mathbb{R}^n} (\langle x, y\rangle - f(x)).$$

Now, answer the following questions, assuming throughout that f and g are convex, proper, and lower semicontinuous, *and that* $\operatorname{dom} f$ *is bounded.* You may take for granted that both f^* and $f \mathbin{\varnothing} g$ are proper and lower semicontinuous, and that f^* is convex.

(ii) Show that $f \mathbin{\varnothing} g$ is convex, and that

$$(f \mathbin{\varnothing} g)^* = f^* + g^*. \tag{1}$$

(iii) We recall that

$$y \in \partial f(x) \iff x \in \partial f^*(y). \tag{2}$$

Use (1) and (2) to compute in terms of the proximal map of f an expression for the subdifferential of the Moreau–Yosida regularisation f_τ of f, defined as

$$f_\tau := f \mathbin{\varnothing} g^\tau \quad \text{for} \quad g^\tau(x) := \frac{1}{2\tau}\|x\|_2^2 \quad \text{and some} \quad \tau > 0.$$

You may use the fact that the conjugate of g^τ is $g^{\tau,*}(y) = \frac{\tau}{2}\|y\|_2^2$.

(iv) What is ∂f_τ for $f(x) = \|x\|_2$? How could replacing f by f_τ be useful for optimisation?

2

This question is about fixed point theorems and their application to proving the convergence of optimisation methods. To get started, let us recall basic concepts by briefly answering the following questions:

(i) What does it mean for a map $T : \mathbb{R}^n \to \mathbb{R}^n$ to be α-averaging?

(ii) How about firmly non-expansive? What is the relationship to α-averaging operators?

(iii) What is the Browder fixed point theorem for averaging operators?

Now, let us put the fixed point theorem to use. We consider a convex function $f : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$, $(x, y) \mapsto f(x, y)$. We assume that f is L-smooth, and introduce the following alternating optimisation scheme:

$$y^{i+1} := \operatorname{prox}_{\tau \partial_y f(x^i, \cdot)}(y^i) := \operatorname*{arg\,min}_{y \in \mathbb{R}^m} \left(f(x^i, y) + \frac{1}{2\tau} \|y - y^i\|_2^2 \right), \tag{1a}$$

$$x^{i+1} := x^i - \tau \nabla_x f(x^i, y^{i+1}). \tag{1b}$$

Here $\partial_y f(x^i, \cdot)$ means the subdifferential of the function $y \mapsto \partial_y f(x^i, y)$.

It is your job to show the scheme (1) works by doing the following.

(iv) Show that smoothness implies co-coercivity.

(v) Show that the scheme (1) converges to a minimiser of f when $0 < \tau L \leqslant 1$.

(vi) Let $\phi : \mathbb{R}^m \to \mathbb{R}$ be convex, $c \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$. Write down the steps of the scheme (1) for

$$f(x, y) := \|Ax + y - c\|_2^2 + \phi(y),$$

in terms of the proximal map of ϕ.

3

This question is set-valued differentatiation and sensitivity analysis. To get started, let us recall basic concepts and results by briefly answering the following questions:

(i) What is the Aubin property? How can it be used for sensitivity analysis?

(ii) What is the Mordukhovich criterion? Provide an exact statement, including a definition of the coderivative therein. How is the Mordukhovich criterion useful for sensitivity analysis?

Let us now consider the following simple one-dimensional linear support vector machine

$$\min_{x \in \mathbb{R}} \frac{\alpha}{2} x^2 + \frac{1}{n} \sum_{i=1}^{n} \max\{0, 1 - a_i x\}, \tag{1}$$

where $a_i > 0$ means that sample i is class "A" and $a_i < 0$ that sample i is in class "B". The parameter $\alpha > 0$ is our regularisation parameter.

Now, answer the following questions:

(iii) What are the necessary and sufficient optimality conditions for (1)?

(iv) Let $\hat{x}$ be an optimal solution to (1) for the parameter $a = \hat{a}$. In the case that $1 - \hat{a}_i \hat{x} \neq 0$ for all $i \in \{1, \ldots, n\}$, does the solution map of (1) have the Aubin property? (Note that you will not need to compute graphical derivatives or coderivatives, although that route is also possible.)

END OF PAPER