Convex Optimisation: Assignment 4

Solutions should be typeset in LaTeX
and submitted via email to `matthew.tam@unimelb.edu.au`

Due: June 3rd, 2021 at 5pm

1. Let $y \in \mathbb{R}^d$ and $\Phi = [\phi_1 \ \dots \ \phi_n] \in \mathbb{R}^{d\times n}$ (*i.e.,* $\phi_i \in \mathbb{R}^n$ denotes the ith column of Φ). Consider the minimisation problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 \ \text{ s.t. } \ \|x\|_1 \leq 1. \tag{1}$$

This problem arises when a known "signal" y is to be represented as a *sparse* linear combination of predefined "atoms" (*i.e.,* the columns ϕ_i). In other words, we want to find scalar $x_1, \dots, x_n \in \mathbb{R}$ (most of which are zero) such that

$$y \approx \sum_{i=1}^{n} x_i\phi_i = \Phi x.$$

In (1), the objective function ensures $y \approx \Phi x$ and the constraint with the ℓ_1-norm promotes sparsity in the solution (the reasons for this are beyond the scope of this course).

(a) (1 point) Verify that (1) satisfies the assumptions of Theorem 3.1.2.

(b) (2 points) Given $C \subseteq \mathbb{R}^n$, a vector $x \in C$ is said to be an *extreme point* of C if there does not exists $y, z \in C$ with $y \neq z$ and $\lambda \in (0, 1)$ such that $x = \lambda y + (1 - \lambda)z$.

Characterise the extreme points of the set $C := \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$.

(c) (2 points) Let (s_k) and (x_k) be the sequences given by (3.2) in the notes. Show that (s_k) can be expressed as

$$\begin{cases} i_k &= \arg\max_{i\in\{1,\dots,n\}} |\langle \phi_i, y - \Phi x_k\rangle| \\ s_k &= \operatorname{sign}\big(\langle \phi_{i_k}, y - \Phi x_k\rangle\big)e_{i_k} \end{cases}$$

where sign denotes the sign function, and $e_1, \dots, e_n$ denote the n-dimensional standard basis vectors.

Hint: The following theorem may be useful.

Theorem. *Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Suppose*

$$P := \{x \in \mathbb{R}^n : Ax \le b\}$$

contains at least one extreme point. Then there exists an extreme point x^* of P such that $\langle c, x^*\rangle = \min_{x\in P}\langle c, x\rangle$.

(d) (1 point) Explain briefly why the formula in (c) provides a more efficient way to compute s_k then the original version of the formula in (3.2).

2. An alternative way to formulate the problem from Question 1 is to penalise constraint violations in the objective function. More precisely, let $\gamma > 0$ and consider the minimisation problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|^2 + \gamma\|x\|_1.$$

(a) (1 point) Propose an algorithm for solving this problem. You answer should briefly justify the suitability of your choice in terms of the problem's mathematical properties.

(b) (2 points) Explain how to (efficiently) compute the steps of the algorithm proposed in (a).

(*Hint:* Question 4 from Assignment 2)

3. Let $\gamma > 1$ and consider the unconstrained minimisation problem

$$\min_{x\in\mathbb{R}^2} f(x) \quad \text{where} \quad f(x) := \frac{1}{2}x_1^2 + \frac{\gamma}{2}x_2^2.$$

Given an initial point $x_0 \in \mathbb{R}$ and $\lambda > 0$, generate (x_k) according to

$$x_{k+1} := x_k - \lambda\nabla f(x_k) \quad \forall k \in \mathbb{N}. \tag{2}$$

(a) (2 points) Verify that f is μ-strongly convex and ∇f is L-Lipschitz continuous. In your answer, you should clearly state the values of μ and L.

(b) (1 point) By appealing to an appropriate theorem from the lectures, state a convergence result for the sequence (x_k). Your answer should include the interval for the stepsize λ.

(c) (1 point) Show that (x_k) satisfies

$$x_{k+1} = \begin{bmatrix} 1-\lambda & 0 \\ 0 & 1-\gamma\lambda \end{bmatrix} x_k.$$

(d) (1 point) Using a direct verification, show that the convergence result stated in (b) holds.

(e) (1 point) Suppose $x_0 = (a, b)$ where $a, b \ne 0$. Characterise the convergence of the sequence of objective function values (*i.e.*, the sequence $\big(f(x_k)\big)$) when the stepsize λ does not belong to the range specified by the result stated in (b).

4. Let $C, D \subseteq \mathbb{R}^n$ be closed, convex sets. The *convex feasibility problem* is to

$$\text{find } x \in C \cap D. \tag{FP}$$

In this exercise, we consider two different ways to formulate this problem as a minimisation problem. Specifically, we consider the problems (P1) and (P2) given by

$$\min_{x\in\mathbb{R}^n} \iota_C(x) + \iota_D(x), \tag{P1}$$

$$\min_{x\in\mathbb{R}^n} d_C^2(x) + \iota_D(x). \tag{P2}$$

Here $d_C\colon \mathbb{R}^n \to \mathbb{R}$ denotes distance function of the set C given by

$$d_C(x) := \inf_{c\in C} \|x - c\| \quad \forall x \in \mathbb{R}^n.$$

(a) (1 point) Explain the relationship between the solutions of (FP), (P1) and (P2). (*Hint:* Question 2 from Assignment 3 may be useful for d_C^2).

(b) (1 point) Compare the mathematical properties of (P1) and (P2).

(c) (2 points) By applying the Douglas–Rachford method to (P1), propose an algorithm for solving (FP) whose iteration uses the projections onto C and D. With reference to an appropriate result from the lectures, justify the convergence of the proposed algorithm.

(d) (2 points) By applying an appropriate method to (P2), propose an algorithm for solving (FP) whose iteration uses the projections onto C and D. With reference to an appropriate result from the lectures, justify the convergence of the proposed algorithm.

5. Suppose $f_i\colon \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable for all $i \in I := \{1,\dots,n\}$, and define $f(x) := \max_{i\in I} f_i(x)$ for all $x \in \mathbb{R}^n$. Consider the minimisation problem

$$\min_{x\in\mathbb{R}^n} f(x). \tag{3}$$

(a) (1 point) Described the mathematical properties of (3).

(b) (2 points) Let $x \in \mathbb{R}^n$ and denote $I(x) := \{i \in I : f(x) = f_i(x)\}$ for all $x \in \mathbb{R}^n$. Show that

$$\sum_{i\in I(x)} \lambda_i \nabla f_i(x) \subseteq \partial f(x)$$

whenever $\lambda \in \mathbb{R}_+^{|I(x)|}$ with $\sum_{i\in I(x)} \lambda_i = 1$. (Here $|S|$ denotes the cardinality of the set S.)

(c) (1 point) By applying an appropriate method from the lectures, propose an algorithm for solving (3). You should justify the convergence of the proposed algorithm.