Methods for Nonsmooth Convex Minimization

INMA 2460:
Nonlinear Programming,
Exercise # 2

April 2020

1 Motivation

The goal of this exercise consists in practical implementation and comparison of two non-smooth convex optimization methods. The students are asked to create the corresponding computer programs, run several series of tests and write down a report with the analysis of the results. The formal requirements are as follows.

- **Schedule:** The final report has to be presented no later than last Friday in May 2020.
- **Marks:** This exercise is not obligatory. However, we strongly recommend to do it in the best possible way since it adds up to 5 points at the final examination.
- **Software:** The students can use any programming language (*C, Fortran*, etc.) or environment (*MatLab*).
- **Final Report.** This report is usually composed by the following parts:
  1. Description of the problems and the numerical methods.
  2. Description of the set of test problems and the testing strategy.
  3. Analysis of the results.
  4. Appendix (listing of the code and full computation results)

  Items 2 and 3 are of the most importance for the final evaluation.

2 Problem Formulation

The problem formulation is as follows:

$$\min_{x \in R^n} f(x), \tag{2.1}$$

where f is a nondifferentiable convex function.

Assumption 2.1 *For simplicity, we assume that the minimum x^* of problem (2.1) exists, and we know an estimate ρ:*

$$\| x_0 - x^* \| \leq \rho,$$

where x_0 is the starting point of the method.

3 Methods (Lecture 8)

In this exercise, it is necessary to test two numerical methods.

3.1 Subgradient Method

$$\begin{aligned} x_0 &\in R^n; \\ x_{k+1} &= x_k - \tfrac{\rho}{\sqrt{k+1}} \tfrac{g_k}{\|g_k\|}, \quad k \geq 0, \end{aligned}$$

where $g_k \in \partial f(x_k)$.

3.2 Ellipsoid Method

$$\begin{aligned} x_0 &\in R^n, & x_{k+1} &= x_k - \tfrac{1}{n+1} \cdot \tfrac{H_k g_k}{\langle H_k g_k, g_k \rangle^{1/2}}, \\ H_0 &= \rho^2 I_n, & H_{k+1} &= \tfrac{n^2}{n^2-1} \left(H_k - \tfrac{2}{n+1} \cdot \tfrac{H_k g_k g_k^T H_k}{\langle H_k g_k, g_k \rangle} \right), \quad k \geq 0, \end{aligned}$$

where I_n is the unit $n \times n$ matrix and $g_k \in \partial f(x_k)$.

4 Computation of the subgradients (Lecture 7)

Recall that vector $g \in R^n$ is called *subgradient* of function $f(\cdot)$ at point x_0, if for any $x \in R^n$ we have

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle.$$

For nondifferentiable convex functions, the subgradient is not always unique. The set of all subgradients at x_0 is denoted by $\partial f(x_0)$.

The following rules can be applied for computing the subgradients.

1. If $f(\cdot)$ is differentiable at x_0, then the subdifferential consists of a single vector, the gradient:

$$\partial f(x_0) \equiv \{f'(x_0)\}.$$

2. If $f(x) = \alpha f_1(x) + \beta f_2(x)$ with α, $\beta \geq 0$, then

$$\partial f(x_0) = \alpha \partial f_1(x_0) + \beta \partial f_2(x_0).$$

This means that for any $g_1 \in \partial f_1(x_0)$ and $g_2 \in \partial f_2(x_0)$, we have

$$\alpha g_1 + \beta g_2 \in \partial f(x_0).$$

3. If $f(x) = \max\{f_1(x), f_2(x)\}$, then

$$\partial f(x_0) = \begin{cases} \partial f_1(x_0), & \text{if } f_1(x_0) > f_2(x_0), \\ \partial f_2(x_0), & \text{if } f_1(x_0) < f_2(x_0), \\ \mathrm{Conv}\{\partial f_1(x_0), \partial f_2(x_0)\}, & \text{if } f_1(x_0) = f_2(x_0). \end{cases}$$

In the latter case, for any $g_1 \in \partial f_1(x_0)$, $g_2 \in \partial f_2(x_0)$ and $\alpha \in [0, 1]$, we have

$$\alpha g_1 + (1-\alpha) g_2 \in \partial f(x_0).$$

5 Test Problems

Any test problems is defined by the choice of the following objects:

- objective function $f(x)$,
- starting point x_0 for the minimization process,
- accuracy of the approximate solution $\epsilon > 0$.

In order to have a complete information about the behavior of numerical method, it is reasonable to generate the test problems with known optimal solutions. Therefore, we suggest to use the following strategy.

1. Choose the dimension $n \geq 2$ of the space of variables.
2. Fix the optimal solution of the problem as $x^* = 0 \in R^n$.
3. Choose the objective function. We suggest to use the following family of objective functions:
$$f(x) = \alpha f_1(x) + \beta f_2(x),$$
where the parameters α and β are nonnegative and
$$f_1(x) = \sum_{i=1}^{n-1} \mid x^{(i)} \mid,$$
$$f_2(x) = \max_{1 \leq i \leq n} \mid x^{(i)} \mid - x^{(1)}.$$
Then the Lipschitz constant L for the objective function can be estimated as follows:
$$L = \alpha\sqrt{n} + 2\beta.$$
4. Choose the starting point $x_0 \in Q$. The important characteristic of the problem is $\rho = \parallel x_0 - x^* \parallel$.

5. Choose the desired accuracy $\epsilon > 0$. If you implement the above mentioned strategy for generating the objective function, then the optimal value of the problem is always zero. Therefore it is reasonable to introduce in the minimization scheme a termination criterion $f(x_k) \leq \epsilon$. Then the number of iterations, which is necessary to achieve the desired accuracy, can be easily fixed out.

6 Testing Strategy

In the final report, it is necessary to justify a conclusion on the performance and the *sensitivity* of the methods to the following characteristics:

- Desired accuracy ϵ.
- Lipschitz constant L.
- Initial distance to the minimum ρ.
- Dimension of the space n.

The typical values of these parameters are presented in the following table.

	ϵ	L	ρ	n
Low/Small	10^{-2}	10	10	10
Moderate	10^{-4}	100	100	100
High/Large	10^{-6}	1000	1000	1000