Methods for Smooth Constrained Minimization over a Simple Convex Set

INMA 2460:
Nonlinear Programming,
Exercise # 1

March 13, 2020

1 Motivation

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

- **Schedule:** The last day of sending the final report last Friday of May in 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 consists of the following parts:
  1. Description of problems and numerical methods.
  2. Description of the set of test problems and 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 attributing the final points.

2 Problem Formulation

The problem formulation is as follows:

$$\min_{x \in Q} \; f(x), \tag{1}$$

where

- Q is a *simple* convex set (see Section 4),
- f is a continuous strongly convex function with Lipschitz-continuous gradient. For simplicity, we assume, that f is *twice* differentiable and for all $x \in \mathbb{R}^n$ we have:

$$\mu I_n \preceq f''(x) \preceq L I_n, \quad \mu > 0. \tag{2}$$

Assumption 2.1 *For simplicity, we assume that the parameters μ and L of the objective function $f(\cdot)$* are known. *They can be used* directly *in the optimization methods.*

3 Methods

In this exercise it is necessary to test two numerical methods. Both of them are based on the notion of *Gradient Mapping* (see Lecture 5).

Let us choose some $\bar{x} \in Q$. Consider the following auxiliary problem:

$$\min_{x \in Q} \; \left[\, \langle f'(\bar{x}), x - \bar{x} \rangle + \tfrac{L}{2} \|x - \bar{x}\|^2 \, \right], \tag{3}$$

where L is the constant from (2) and the norm is the standard Euclidean:

$$\|x\| \;=\; \left[\sum_{i=1}^{n} (x^{(i)})^2 \right]^{1/2}.$$

Denote by $x_Q(\bar{x})$ the unique solution of this problem and

$$g_Q(\bar{x}) = L(\bar{x} - x_Q(\bar{x})).$$

This vector is called the *Gradient Mapping* of the problem (1), computed at the point $\bar{x}$.

3.1 Gradient Method

$$\begin{aligned} x_0 \;&\in\; Q; \\ x_{k+1} \;&=\; x_Q(x_k), \quad k \geq 0. \end{aligned}$$

3.2 Optimal Method

Choose $y_0 = x_0 \in Q$ and $\beta = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$. Then

$$x_{k+1} = x_Q(y_k), \quad y_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k), \quad k \geq 0.$$

4 Computation of the Gradient Mapping

Objective function of problem (3) can be written in the following form:

$$\langle f'(\bar{x}), x - \bar{x}\rangle + \frac{L}{2} \parallel x - \bar{x} \parallel^2 = -\frac{1}{2L} \parallel f'(\bar{x}) \parallel^2 + \frac{L}{2} \left\| x - \left(\bar{x} - \frac{1}{L} f'(\bar{x})\right)\right\|^2.$$

(Check this!) Therefore, in order to compute point $x_Q(\bar{x})$, we need to solve the problem

$$\min\{\parallel x - y(\bar{x}) \parallel \mid x \in Q\}, \tag{4}$$

where $y(\bar{x}) = \bar{x} - \frac{1}{L} f'(\bar{x})$. Clearly, solution of this problem is exactly $x_Q(\bar{x})$.

Solution of problem (4) is the *projection* of point $y(x_0)$ onto the set Q. If Q is simple enough, this projection can be found analytically.

Denote by $\pi_Q(y)$ the Euclidean projection of point y onto the set Q. Consider several examples of the simple sets.

1. *Box*:
$$Q = \left\{ x \in \mathbb{R}^n \mid a^{(i)} \leq x^{(i)} \leq b^{(i)}, \ i = 1 \ldots n \right\}.$$
Let us introduce the following function of one variable:
$$[t]_\alpha^\beta = \begin{cases} \beta, & \text{if} \quad t \geq \beta, \\ t, & \text{if} \quad \alpha \leq t \leq \beta, \\ \alpha, & \text{if} \quad t \leq \alpha. \end{cases}$$
Then it is easy to see that
$$(\pi_Q(y))^{(i)} = \left[y^{(i)}\right]_{a^{(i)}}^{b^{(i)}}, \quad i = 1 \ldots n.$$

2. *Euclidean Ball*:
$$Q = \left\{ x \in \mathbb{R}^n \mid \parallel x \parallel^2 \equiv \sum_{i=1}^n (x^{(i)})^2 \leq \mathbb{R}^2 \right\}.$$
Then
$$\pi_Q(y) = \begin{cases} y, & \text{if} \quad \parallel y \parallel \leq R, \\ Ry / \parallel y \parallel, & \text{if} \quad \parallel y \parallel > R. \end{cases}$$

3. *Standard simplex*: $Q = \left\{ x \in \mathbb{R}^n : \sum\limits_{i=1}^n x^{(i)} = 1, \; x^{(i)} \geq 0, i = 1, \dots, n \right\}$. In order to find this projection, we need to apply the following transformation of the problem:

$$\begin{aligned} \min_{x \in Q} \tfrac{1}{2}\|x - y\|^2 &= \min_{x \geq 0} \max_{\lambda \in \mathbb{R}} \left\{ \tfrac{1}{2} \sum_{i=1}^n (x^{(i)} - y^{(i)})^2 + \lambda \cdot (1 - \sum_{i=1}^n x^{(i)}) \right) \\ &= \max_{\lambda \in \mathbb{R}} \left[\lambda + \sum_{i=1}^n \min_{x^{(i)} \geq 0} \left\{ \tfrac{1}{2}(x^{(i)} - y^{(i)})^2 - \lambda x^{(i)} \right\} \right] \end{aligned}$$

For the internal minimization problem, the optimal solution is

$$x^{(i)}(\lambda) = \max\{y^{(i)} + \lambda, 0\} \stackrel{\mathrm{def}}{=} \left(y^{(i)} + \lambda\right)_+ .$$

Substituting it in the objective function, we get a concave piece-wise quadratic function of $\lambda \in \mathbb{R}$. Indeed

$$\begin{aligned} \tfrac{1}{2}(x^{(i)}(\lambda) - y^{(i)})^2 - \lambda x^{(i)}(\lambda) &= \tfrac{1}{2}(x^{(i)}(\lambda))^2 - x^{(i)}(\lambda) y^{(i)} + \tfrac{1}{2}(y^{(i)})^2 - \lambda x^{(i)}(\lambda) \\ &= \tfrac{1}{2}\left(y^{(i)} + \lambda\right)_+ + \tfrac{1}{2}(y^{(i)})^2 - (y^{(i)} + \lambda)\left(y^{(i)} + \lambda\right)_+ \\ &= \tfrac{1}{2}(y^{(i)})^2 - \tfrac{1}{2}\left(y^{(i)} + \lambda\right)_+^2 . \end{aligned}$$

Therefore, the objective function

$$\lambda + \sum_{i=1}^n \left\{ \tfrac{1}{2}(x^{(i)}(\lambda) - y^{(i)})^2 - \lambda x^{(i)}(\lambda) \right\} = \lambda + \sum_{i=1}^n \tfrac{1}{2} \left\{ (y^{(i)})^2 - \left(y^{(i)} + \lambda\right)_+^2 \right\}$$

can be easily maximized in λ either by binary search or by a direct analysis of all switching points of this functions $\lambda_i = -y^{(i)}$, $i = 1, \dots, n$.

5 Test Problems

All test problems below are defined by the choice of the following objects:

- the feasible set Q,
- the objective function $f(\cdot)$,
- the starting point x_0 for the minimization process,
- the 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 solutions. Therefore, we suggest to use the following strategy.

1. Choose the type of feasible set (see the previous section).

2. Choose a optimal point $x^* \in Q$. For that you can use a random generator. Note, that for the positive orthant and for the box constraints it may be important how many linear constraints are *active* at the solution (in the case of the positive orthant, this is the number of zeros among the components of x^*).

3. Choose the objective function. We suggest to use the following family of objective functions:

$$f(x) \quad = \quad \tfrac{\alpha}{2} \parallel x - x^* \parallel^2 + \beta f_1(x - x^*) + \gamma[f_2(x) - f_2(x^*) - \langle f_2'(x^*), x - x^* \rangle], \qquad (5)$$

where the parameters α, β and γ are positive and

$$f_1(x) = \tfrac{1}{2}[(x^{(1)})^2 + \sum_{i=1}^{n-1} (x^{(i)} - x^{(i+1)})^2 + (x^{(n)})^2],$$

$$f_2(x) = \ln \left(\sum_{i=1}^{n} e^{x^{(i)}} \right).$$

For stability of computations, it is reasonable to keep f_2 in the form

$$f_2(x) \quad = \quad \delta + \ln \left(\sum_{i=1}^{n} e^{x^{(i)} - \delta} \right),$$

where $\delta \geq \max\limits_{1 \leq i \leq n} x^{(i)}$. Then the parameters μ and L of the objective function can be estimated as follows:

$$\mu = \alpha, \quad L = \alpha + 4\beta + \gamma.$$

4. Choose a starting point $x_0 \in Q$. The important characteristic of the problem is the distance to the minimum: $R = \parallel x_0 - x^* \parallel$.

5. Choose the desired accuracy $\epsilon > 0$. For all test functions of the form (5), the optimal value of the problem is zero. Therefore it is reasonable to introduce in the minimization scheme the termination criterion $f(x_k) \leq \epsilon$. Then you will know exactly the number of iterations, which are necessary for achieving the desired accuracy.

6 Testing Strategy

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

- Desired accuracy ϵ.
- Condition number $\kappa = \frac{L}{\mu}$.
- Initial distance to the minimum R.
- Dimension of the problem n.
- Number of active constraints m.

The typical values of these parameters you can find in the following table.

	ϵ	κ	R	n	m
Low (Small)	10^{-2}	10	10	10	0
Moderate	10^{-4}	10^3	10^2	10^2	$\min\{n/10, 10\}$
High (Large)	10^{-6}	10^6	10^3	10^3	$n/2$

Justification can be done by comparing your observations with the facts known from the theory.