Computer Science Department, Ben Gurion University of the Negev

Optimization methods with Applications - Spring 2022

Home Assignment 4

Due on July 1st.

1. **Equality constrained optimization**
   In this section we will find the maximal surface area of a box given the sum of its edges' length. The optimization problem is given by
   $$\max_{\mathbf{x}\in\mathbb{R}^3} \{x_1x_2 + x_2x_3 + x_1x_3\} \quad s.t. \quad \{\ x_1 + x_2 + x_3 = 3\ . \tag{1}$$
   (a) Find a critical point for the problem (1) using the Lagrange multiplier method.
   (b) Show that this critical point is a maximum point. For this show that the Hessian of the Lagrangian is negative ($\mathbf{y}^\top\nabla^2\mathcal{L}\mathbf{y} < 0$) for vectors $\mathbf{y} \neq 0$ who satisfy $\mathbf{y}^\top\mathbf{1} = y_1 + y_2 + y_3 = 0$.

2. **General constrained optimization**
   Assume that we have the following problem.
   $$\min_{\mathbf{x}\in\mathbb{R}^2} \left\{(x_1 + x_2)^2 - 10(x_1 + x_2)\right\} \quad s.t. \quad \left\{ \begin{array}{c} 3x_1 + x_2 = 6 \\ x_1^2 + x_2^2 \leq 5 \\ -x_1 \leq 0 \end{array} \right. . \tag{2}$$
   (a) Find a critical point $\mathbf{x}^*$ using the Lagrange multipliers method for which the fewest inequality constraints are active. Show that the KKT conditions hold.
   (b) Using second order conditions, determine whether the point $\mathbf{x}^*$ you found in the previous section is a minimum or maximum.
   (c) Write the unconstrained minimization problem that corresponds to the problem (2) above using the penalty method with $\rho(x) = x^2$ as a penalty function.
   (d) Use the penalty method to get the minimizer $\mathbf{x}^*$ up to two digits of accuracy. Solve the optimization problems using steepest descent with Armijo linesearch. Use penalty parameters $\mu = 0.01, 0.1, 1, 10, 100$, and plot the minimization graphs of the penalized function throughout the iterations.

3. **Box-constrained optimization**
In this question we will write the **coordinate descent** method for quadratic box-constrained minimization. Assume the following problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{\frac{1}{2}\mathbf{x}^\top H\mathbf{x} - \mathbf{x}^\top \mathbf{g}\right\} \quad s.t. \quad \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}, \tag{3}$$

where $H \in \mathbb{R}^{n\times n}$ is symmetric positive definite, $\mathbf{g} \in \mathbb{R}^n$, and $\mathbf{a} < \mathbf{b} \in \mathbb{R}^n$ are the lower and upper bounds on the solution $\mathbf{x}$.

(a) Give a closed form solution for the scalar box constrained minimization problem

$$\min_{x\in\mathbb{R}} \left\{\frac{1}{2}hx^2 - gx\right\} \quad s.t. \quad a \leq x \leq b, \qquad \text{for } a < b, h > 0.$$

(b) In the coordinate descent algorithm we sweep over all the variables x_i one by one, and for each solve the box-constrained minimization problem for the scalar variable x_i, given that the rest are known. Show that the minimization for each scalar x_i is given by

$$\min_{x_i\in\mathbb{R}} \left\{\frac{1}{2}h_{ii}x_i^2 + \left[\left(\sum_{j\neq i} h_{ij}x_j\right) - g_i\right] x_i\right\} \quad s.t. \quad a_i \leq x_i \leq b_i.$$

Use the previous section to show the expression for the update of the coordinate descent method for this problem.

(c) Use the previous section to write a program for solving (3) using the coordinate descent algorithm.

(d) Solve the problem (3) for the following parameters using the CD method up to three digits of accuracy:

$$H = \begin{bmatrix} 5 & -1 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 \\ -1 & -1 & 5 & -1 & -1 \\ -1 & -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & -1 & 5 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} 18 \\ 6 \\ -12 \\ -6 \\ 18 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \\ 5 \end{bmatrix}.$$

4. **(Not for Submission) Projected Gradient Descent for LASSO regression**

In this section we will solve the minimization

$$\arg\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1 \tag{4}$$

where $\lambda > 0$ which is called the LASSO (least absolute shrinkage and selection operator) problem, and leads to a sparse solution $\mathbf{x}$ - we'll have less non-zeros entries as we increase λ. We did not learn to solve this non-smooth optimization problem yet, so we will use an alternative problem definition.

(a) Show that the following problem is equivalent (that is, it has the same solution $\mathbf{x} = \mathbf{u} - \mathbf{v}$) to the problem (4):

$$\hat{\mathbf{u}}, \hat{\mathbf{v}} = \arg\min_{\mathbf{u},\mathbf{v}\in\mathbb{R}^n} \|A(\mathbf{u} - \mathbf{v}) - \mathbf{b}\|_2^2 + \lambda\left(\mathbf{1}^\top(\mathbf{u} + \mathbf{v})\right), \quad s.t \quad \mathbf{u} \geq 0, \mathbf{v} \geq 0.$$

and the final solution is $\hat{\mathbf{x}} = \hat{\mathbf{u}} - \hat{\mathbf{v}}$.

(b) Write a program for solving the problem using projected Steepest Descent (similarly to the box constraints example in the lecture notes), with Armijo linesearch. Note that:

- To apply steepest descent you should define the gradient w.r.t. the unknowns, which was previously generally described as $\mathbf{x} \in \mathbb{R}^n$. Here, you have two unknown vectors $\mathbf{u}, \mathbf{v}$ and you need to compute derivatives w.r.t both (when you solve the problem you can forget about the relation of them with $\mathbf{x}$ - this is not relevant for this stage). Another way to look at it: think of the unknowns as a large vector $\mathbf{w} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$ and apply projected SD w.r.t. $\mathbf{w}$.
- the projection is done also inside the linesearch procedure.

(c) Demonstrate your program using a synthetic example:

- Choose a random matrix $A \in \mathbb{R}^{100\times200}$ (using Gaussian distribution), and a sparse vector $\mathbf{x} \in \mathbb{R}^{200}$ (uniformly choose 10% of the entries to be non-zeros, and randomly choose their values).
- Define $\mathbf{b} = A\mathbf{x} + \eta$, where $\eta \in \mathbb{R}^{100}$ is a random vector of white Gaussian noise of standard deviation 0.1 (you may try other values as well).

Now try to reconstruct $\mathbf{x}$ given $A, \mathbf{b}$, by solving the problem in section (a). Try a few values of λ that lead to a solution with approximately 10% non-zero entries, and (most importantly) show that the objective is minimized. Out of your try-outs, see that the solution $\hat{\mathbf{x}}$ more or less approximates your original $\mathbf{x}$.