Advanced Topics in Control 2020: Large-Scale Convex Optimization

Exercise 6: Coordinate Descent Methods

Goran Banjac, Mathias Hudoba de Badyn, Andrea Iannelli,
Angeliki Kamoutsi, Ilnura Usmanova

April 10, 2020

Date due: April 23, 2020 at 23:59.

Please submit your solutions via Moodle as a PDF with filename `Ex06_Surname.pdf`, replacing `Surname` with your surname.

---

1 Problem 1

(a) Consider the linear system of equations $Ax = b$, where $A \in \mathbb{S}^n_{++}$ is a symmetric positive definite matrix, $b \in \mathbb{R}^n$ and $x^\star$ denotes the solution.
Starting from an initial guess x^0, the following modified version of the Gauss-Seidel algorithm is proposed:

$$x_i^{k+1} = (1-\gamma)x_i^k - \frac{\gamma}{a_{ii}}\left[\left(\sum_{j<i} a_{ij}x_j^{k+1}\right) + \sum_{j>i} a_{ij}x_j^k - b_i\right] \tag{1}$$

where a_{ij} is the (i,j)-th entry of A and b_i is the i-th entry of b. The original Gauss-Seidel algorithm is retrieved for γ=1.

Show that, if $\gamma \notin (0,2)$, then for every $x^0 \neq x^\star$, the sequence (1) does not converge to $x^\star$.
Hint: You can try to use the analogy between solutions of linear systems and optimization of the associated quadratic function f and show that $f(x^{k+1}) \geq f(x^0)$ for all k.

(b) Consider now the function f defined as:

$$f(x_1, x_2) = \max\left((x_1-1)^2 + (x_2+1)^2, (x_1+1)^2 + (x_2-1)^2\right) \tag{2}$$

Find the minimizer $x^\star = (x_1^\star, x_2^\star)$ of f, and show that, by applying the Gauss-Seidel (or coordinate minimization) algorithm to f from $x^0 = (1,1)$, this does not converge to $x^\star$.
Hint: To find the minimizer, use the fact that $f(x_1, x_2) = f(-x_1, -x_2)$. To answer the second part of the question, you can show that $x^0 = (1,1)$ is a fixed point of the Gauss-Seidel algorithm.

2 Problem 2

Consider the optimization problem:

$$\begin{aligned} \min_x \quad & \frac{1}{2}||x||_2^2, \\ \text{s.t.} \quad & Ax = b, \end{aligned} \tag{3}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and the rows of A (denoted by $a_i^\top$) are normalized such that:

$$||a_i||_2 = 1, \quad i = 1, ..., n. \tag{4}$$

The dual problem of (3) is:

$$\max_z \quad b^\top z - \frac{1}{2}||A^\top z||_2^2. \tag{5}$$

The algorithms in this problem, unless specified otherwise, make use of a cyclic choice of coordinates. Moreover, the stepsize can always be assumed equal to 1.

(a) Write the explicit update rule for the coordinate gradient descent algorithm applied to the dual problem.

(b) Leveraging the previously derived rule, compute the update rule for recovering the primal variables and show that at each iteration the i-th equation in the system $Ax = b$ holds.
Hint: Find the relationship between the primal and dual variables, and use it to convert the update rule find in part (a) to the update rule for the primal variables.

(c) Consider the problem:

$$\min_y \quad \frac{1}{2}||Ay - b||_2^2 = \frac{1}{2}\sum_{j=1}^{m}(a_j^\top y - b_j)^2. \tag{6}$$

where A, b, are the same used in the previous parts. Consider a version of the stochastic gradient descent (SGD) method where an estimate of the gradient g^k of the *full summation* (6) is obtained by taking a *single term* j in the summation, where j is drawn from a uniform probability distribution, that is:

$$\begin{aligned} g^k &= \nabla q_t(y^k), \\ \text{where} \quad q_t(y^k) &= \frac{1}{2}(a_t^\top y^k - b_t)^2, \\ p_t = p(I = t) &= \frac{1}{m}, \end{aligned}$$

Write the explicit update rule for the iterates obtained applying the SGD method.
Hint: If t was chosen deterministically according to the following rule $t = \text{mod}(k, m) + 1$, where $mod(k, m)$ is the remainder after division of k by m (e.g. mod(3,3)=0, mod(5,3)=2), then you should obtain the same update rule of the primal variables derived in part (b).