Homework assignment 5 and 6

Exercise 1 and 2 or 3 are a Hand-in exercises. That is, you can choose if you want to hand in exercise 2 or 3 along with exercise 1.

1. Consider the problem

$$\text{minimize } \tfrac{1}{2}\|Ax-b\|^2 + \lambda\|Dx\|_1 \tag{P}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $D \in \mathbb{R}^{p\times n}$, $x \in \mathbb{R}^n$, and $\lambda \in (0,\infty)$. Assume that A and D always have full rank. We will consider two different dual problems for (P). The primal problem can be written on the following general form

$$\text{minimize } f(x) + g(Lx).$$

Let $f(x) = \frac{1}{2}\|Ax-b\|^2$, $g_1(x) = \lambda\|Dx\|_1$, and $L = I$ to get the first dual problem

$$\text{minimize } f^*(-\mu) + g_1^*(\mu). \tag{D1}$$

Let $f(x) = \frac{1}{2}\|Ax-b\|^2$, $g_2(y) = \lambda\|y\|_1$, and $L = D$ to get the second dual problem

$$\text{minimize } f^*(-L^*\mu) + g_2^*(\mu). \tag{D2}$$

Consider the following cases:

(i) $m \geq n$ and $D = I$
(ii) $m \geq n$ and $p < n$ and D general structure
(iii) $m \geq n$ and $p \geq n$ and D general structure
(iv) $m < n$ and $D = I$
(v) $m < n$ and $p < n$ and D general structure
(vi) $m < n$ and $p \geq n$ and D general structure

a. For each of these cases, motivate if problems (P), (D1), and (D2) can be solved by

(1) forward-backward splitting,
(2) Douglas-Rachford splitting,
(3) coordinate descent,
(4) coordinate gradient descent.

Also motivate if the problem (P) can be solved by linearized Douglas-Rachford splitting applied to (D1) and (D2). (Recall that (P) is solved by linearized Douglas-Rachford by solving a dual problem).

b. For each of these cases (i)-(vi) and each algorithm (1)-(4), motivate which problem formulation (P), (D1) and (D2) (if applicable) that gets cheapest iterations. Also, motivate which of the dual problems (D1) and (D2) that gets cheapest iterations for linearized Douglas-Rachford splitting (you can linearize L^*L only or and L^*L and A^TA).

c. For each case (i)-(vi), algorithms (1)-(2), and problem (P), (D1), and (D2), motivate if the convergence is linear or sublinear.

d. Implement algorithms (1)-(5) for case (iv) with random sparse A with full row rank. Choose λ such that some elements of the solution $x^\star$ are non-zero while the rest are 0.

2. Consider the quadratic program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Hx + h^T x \\ \text{subject to} & Ax = b \\ & l \leq Cx \leq d \end{array}$$

Assume that there exists a feasible point, that $A \in \mathbb{R}^{m\times n}$ with $m < n$, and that A has full row rank. Consider the following cases

(i) H is positive semi-definite, C is sparse,

(ii) H is positive semi-definite, $C = I$,

(iii) H is positive definite, C is sparse,

(iv) H is positive definite, $C = I$.

Formulate the (primal) problem on the form

$$\text{minimize } f(x) + g(Lx)$$

where $f(x) = \frac{1}{2}x^T Hx + h^T x + \iota_{Ax=b}(x)$. To formulate the dual problem, L and g must be chosen. In cases (i) and (iii) consider $L = C$ and $L = I$, and in cases (ii) and (iv) consider $L = I$. That is, in case (i) and (iii), we get two dual problems and in cases (ii) and (iv) we get one dual problem.

For all cases (i)-(iv) and all choices of L, motivate if the primal and dual problems can be solved using forward-backward splitting and Douglas-Rachford splitting.

For each case (i)-(iv), if the algorithm can solve more than one of the considered problems, motivate which formulation (primal, dual with $L = I$, or dual with $L = C$ if applicable) that has the cheapest iterations.

3. Consider the problem

$$\text{minimize } \frac{1}{2}\|Ax - b\|^2 + \lambda\|Dx\|_1$$

with $D = \text{diag}(d) \in \mathbb{R}^{n\times n}$ where $d \in \mathbb{R}^n$ is created from a uniform distribution over $[0, 1]$, and then scaled such that $d_{\max} = 1/d_{\min}$. (Note that $\|Dx\|_1$ is still separable since D is diagonal.) Also let $A \in \mathbb{R}^{m\times n}$ with $m > n$ be a random matrix with full column rank (that is $A^T A$ is positive definite). Let λ be such that some elements of the solution $x^\star$ are non-zero while the others are 0. Solve the problem numerically (without going to the dual) using forward-backward splitting, Douglas-Rachford splitting, coordinate descent, and coordinate gradient descent. Compare the convergence with solving the equivalent problem after a change of variables $q = Dx$

$$\text{minimize } \tfrac{1}{2}\|AD^{-1}q - b\|^2 + \lambda\|q\|_1.$$

Explain the difference in convergence speed (if such a difference exists) between the two problem formulations. Hint: look at the condition numbers of $A^T A$ and $(AD^{-1})^T(AD^{-1})$.