university of groningen / faculty of science and engineering

INTRODUCTION TO OPTIMIZATION

ASSIGNMENT REPORT

HOMEWORK 3

Author:
Bharath Syamala (S5303028)
Hrishikesh Sawant (S5292670)
Zhou Zi Hui (S5778263)

Lecturer:
Prof. Dr. J.G. Juan Peypouquet

Groningen, October 29, 2023

MSc. Industrial Engineering and Management
Faculty of Science and Engineering
University of Groningen

Contents

1 Question 1

(a) Fenchel conjugate of $g(x)$:
The Fenchel conjugate of a function $g(x)$ is defined as follows:

$$g^*(y) = \sup_{x \in \mathrm{dom}(g)} \{\langle y, x\rangle - g(x)\}$$

Here, y is in the dual space $(\mathbb{R}^N)^*$ and $\langle y, x\rangle$ denotes the inner product of y and x.
Given that $g(x) = \iota_Q(x)$, where $Q = [-\rho, \rho]^N$, and $\iota_Q(x)$ is the indicator function of the set Q (i.e., $\iota_Q(x) = 0$ if x is in Q, and $\iota_Q(x) = +\infty$ if x is outside Q, we need to find the Fenchel conjugate of this function.
For the indicator function, the Fenchel conjugate is given by:

$$g^*(y) = \sup_{x \in Q} \{\langle y, x\}$$

Since Q is a bounded set $[-\rho, \rho]^N$, the supremum will be attained at the boundary of the set. The boundary points of Q are $x = \pm\rho$ in each dimension. So, we can write:

$$g^*(y) = \max\{\langle y, \rho\rangle, \langle y, -\rho\rangle\}$$

Now, we can simplify this expression:

$$g^*(y) = \rho \max\{y, -y\}$$

Depending on the sign of y, we can see that the maximum is either y or $-y$:

$$g^*(y) = \begin{cases} \rho y, & \text{if } y \geq 0 \\ -\rho y, & \text{if } y < 0 \end{cases}$$

So, the Fenchel conjugate of $g(x)$ is a piecewise function.
(b) Fenchel conjugate of $h(y)$:
The Fenchel conjugate of $h(y)$ is also defined using the same formula as above:

$$h^*(x) = \sup_{y \in \mathrm{dom}(h)} \{\langle x, y\rangle - h(y)\}$$

Given that $h(y) = \frac{1}{2}\|y\|^2 + b \cdot y$, where $b \in \mathbb{R}^M$, we need to find the Fenchel conjugate.
First, let's calculate the gradient of $h(y)$ with respect to y:
$\nabla h(y) = y + b$
Now, to find the Fenchel conjugate, we need to solve for $h^*(x)$:

$$h^*(x) = \sup_y \{\langle x, y\rangle - \left(\frac{1}{2}\|y\|^2 + b \cdot y\right)\}$$

To find the supremum, we take the derivative of the expression inside the sup with respect to y and set it equal to zero:
$\nabla[\langle x, y\rangle - \frac{1}{2}\|y\|^2 - b \cdot y] = 0$

$$x - y - b = 0$$

Solving for y:

$$y = x - b$$

Now, we can substitute this value of y back into the original expression:

$$h^*(x) = \langle x, x - b\rangle - \left(\frac{1}{2}\|x - b\|^2 + b \cdot (x - b)\right)$$

Expanding and simplifying:

$$h^*(x) = \langle x, x\rangle - \langle x, b\rangle - \frac{1}{2}\|x - b\|^2 - \langle b, x - b\rangle$$

$$h^*(x) = \frac{1}{2}\|x\|^2 - \langle x, b\rangle - \frac{1}{2}\|x - b\|^2 + \langle b, b$$

So, the Fenchel conjugate of $h(y)$ is given by this expression.

2 Question 2

Primal Problem (P):

$$\min_{x \in \mathbb{R}^N} \{g(A^T x) + h(x)\}$$

where A is a real matrix of size $M \times N$.

The dual problem (D) of the primal problem (P) is typically derived by finding the convex conjugates of the individual functions in the primal problem.

Let's denote the convex conjugate of $g(A^T x)$ as $g^*(y)$. By the definition of the convex conjugate, we have:

$$g^*(y) = \sup_x \{\langle x, y\rangle - g(A^T x)\}$$

Similarly, let's denote the convex conjugate of $h(x)$ as $h^*(z)$. By the definition of the convex conjugate, we have:

$$h^*(z) = \sup_x \{\langle x, z\rangle - h(x)\}$$

The dual problem (D) is then typically formulated as the maximum of the convex conjugates of the individual functions:

$$\text{(D)} \;\; \max_y \{g^*(y) + h^*(z)\}$$

From the above equations, it can be observed that in (P), the objective function combines a linear term ($A^T x$) with a non-smooth term h(x), and the proximal-gradient method is commonly employed for optimizing problems with non-smooth components like h(x). Similarly, in (D), the dual problem includes the convex conjugates of the terms g($A^T x$) and h(x). The proximal-gradient method is well-suited for solving convex optimization problems involving the convex conjugates [2].

3 Question 3

In terms of expected convergence rate, the proximal-gradient method has a better performance on the dual problem(D).

D is strongly convex so a linear convergence rate is expected.

P is non strongly convex so the sublinear convergence rate is expercted.

In terms of ease of implementation, the proximal-gradient method is easier to implement on the primal problem(P).

For P, we need to deal with non-smooth component h(x), but it is still manageable with method metioned in *Inexact Proximal Gradient Methods for Non-Convex and Non-Smooth Optimization*[1] .

For D, we need to work with the dual variables, compute the dual objective function, and find the proximal operator of the conjugate of the dual problem's objective.

References

[1] B. Gu, D. Wang, Z. Huo, and H. Huang. Inexact proximal gradient methods for non-convex and non-smooth optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018.

[2] N. Parikh, S. Boyd, et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.