AM8211 Non-Linear Optimization Final Report: Subdifferential Calculus & Optimization

Mariam Walaa

November 29, 2023

Contents

1 Overview

When working with non-linear programming problems, it is possible to encounter objective functions that are not necessarily possible to differentiate, which is commonly done to objective functions in non-linear optimization. One example is any objective function that uses the absolute value function. The theory of convex analysis gives a characterization of continuously differentiable convex functions, but not non-differentiable convex functions. The theory of subdifferential calculus introduces a new characterization of non-differentiable convex functions, which generalizes the concept of gradient to all kinds of convex functions.

An important result that stems from the theory of subdifferentials is the subdifferential optimality condition, which is a necessary and sufficient condition for optimality in convex optimization problems involving non-differentiable convex functions. There are numerous applications of the theory of subdifferentials, one of the most prominent being the L1 norm which is typically used in L1 regularization in machine learning and data analysis. Another example in machine learning is the hinge loss of support vector machines (SVMs). This report will aim to summarize the topic of subdifferential calculus for non-differential convex functions by providing the main ideas, theorems, and applications of subdifferential calculus.

2 Introduction to subdifferentials

The following theorem characterizes convexity for smooth functions.

Theorem 2.1 *A smooth, differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for all x, y in its domain, the following inequality holds:*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

We seek to find a similar characterization of convexity for non-smooth functions.

Definition 2.1 *For a convex function $f : \mathbb{R}^n \to \mathbb{R}$ defined on a convex domain, the subdifferential* *or subgradient at a point x is denoted as $\partial f(x)$ and is defined as:*

$$\partial f(x) = \{g \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle g, y - x \rangle, \quad \forall y \in \mathbb{R}^n\}$$

3 Optimization for subdifferentials

The theory of subdifferentials gives a criteria that is analogous to the optimality conditions for smooth convex functions.

Theorem 3.1 *Suppose f is a convex, subdifferentiable function on Ω $x_0 \in \Omega$ is a minimizer of f if and only if $\vec{0} \in \partial f(x_0)$*

3.1 Geometric Sense

For a differentiable convex function, the gradient at a point provides a unique direction that points uphill, and the function's local behavior is fully characterized by this slope. However, when dealing with non-differentiable convex functions, there may be points where a unique gradient doesn't exist. The subgradient serves as a generalization of the gradient for non-differentiable convex functions. At a point where the function is not differentiable, the subgradient captures a range of possible slopes or directions that still ensure the function lies below any tangent line. It represents a set of valid gradients at that point.

In a geometric sense, the subgradient can be thought of as a set of vectors that, when added to the function value at a given point, provides a lower bound for the function values at nearby points. The subgradient allows us to define a linear approximation that underestimates the function.

3.2 Applications

There are numerous applications of the theory of subdifferentials, one of the most prominent being the L1 norm which is typically used in L1 regularization in machine learning and data analysis. Another example in machine learning is the hinge loss of support vector machines (SVMs).

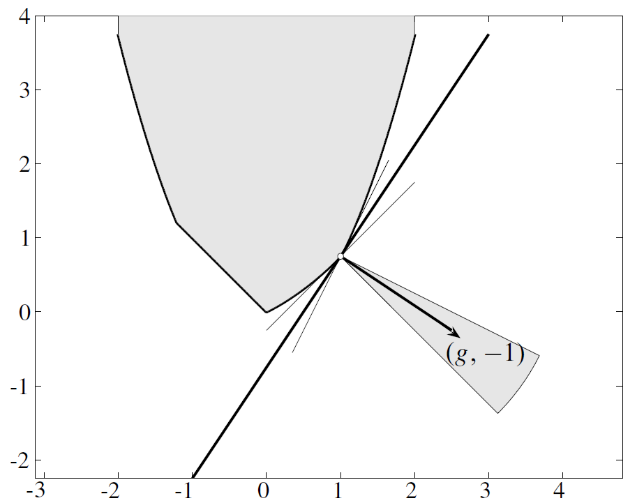

Figure 1: Figure 2.12 from Chapter 2.5 in [3].

3.3 Examples

Example 3.1 *Consider the function $f(x) = |x|$, which is not differentiable at $x = 0$. The subdifferential at $x = 0$ is the interval $[-1, 1]$.*

$$f(x) = |x|$$

$$\partial f(x) = \begin{cases} -1 & \textit{if } x < 0 \\ [-1, 1] & \textit{if } x = 0 \\ 1 & \textit{if } x > 0 \end{cases}$$

Example 3.2 *Consider the function $f(x) = \max(0, x)$, which is not differentiable at $x = 0$. Its subdifferential at $x = 0$ is $[0, \infty)$.*

$$f(x) = \max(0, x)$$

$$\partial f(x) = \begin{cases} [0, \infty) & \textit{if } x = 0 \\ 1 & \textit{if } x > 0 \end{cases}$$

4 Calculus of Subgradients

The "calculus" of subgradients [2, 1] is a mathematical framework, or a set of rules, that extends the concept of derivatives to non-smooth functions, and helps to manipulate subdifferentials when working with them. Subgradients play a crucial role in the optimization of convex functions, where traditional derivatives may not exist.

4.1 Non-Negative Scaling

The subdifferential of a non-smooth convex function $f : \mathbb{R}^n \to \mathbb{R}$ at a point x is denoted as $\partial f(x)$. If $\alpha \geq 0$, then the subdifferential of αf at x is given by:

$$\partial f(\alpha)(x) = \alpha \partial f(x)$$

4.2 Sum and Integral

For the sum of two convex functions f_1 and f_2, the subdifferential at x is the Minkowski sum of their individual subdifferentials:

$$\partial(f_1 + f_2 + ... + f_m)(x) = \partial f_1(x) + \partial f_2(x) + ... + \partial f_m(x)$$

Similarly, for integral of a convex function g over a set A, the subdifferential at x is given by:

$$\partial \left(\int_A g(y)\, dy \right)(x) = \int_A \partial g(y)\, dy$$

4.3 Affine Transformation of Domain

If A is an affine transformation $h := Ax + b$, then the subdifferential of f after transformation is:

$$\partial h(x) = A^T \partial f(Ax + b)$$

4.4 Pointwise Maximum

For the pointwise maximum of two convex functions f_1 and f_2, the subdifferential at x is the convex hull of the individual subdifferentials at x:

$$\partial(\max(f_1, f_2, ..., f_m))(x) = \max_{i=1,2,...,m} f_i$$

4.5 Supremum

For the supremum of a family of functions $\{f_i\}$, the subdifferential at x is the convex hull of the union of the individual subdifferentials:

$$\partial\left(\sup_i f_i\right)(x) = \text{conv}\left(\bigcup_i \partial f_i(x)\right)$$

4.6 Minimization Over Some Variables

If f is minimized over a subset S of the variables, the subdifferential with respect to the remaining variables is obtained by setting the partial derivatives to zero:

$$\partial\left(\min_{x \in S} f(x)\right)(x) = \partial f(x) \cap T_S$$

Here, T_S is the tangent cone to S at x. These expressions provide a glimpse into the structure of subgradients and their behavior under various operations. The calculus of subgradients is a powerful tool for analyzing and optimizing non-smooth convex functions.

References

[1] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004. MR 2061575

[2] F. H. Clarke, *Optimization and nonsmooth analysis*, second ed., Classics in Applied Mathematics, vol. 5, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[3] Andrzej Ruszczynski, *Non-linear optimization*, Princeton University Press, Princeton, NJ, 2006.