Subgradient Methods Applied to LASSO Regression

Kasey Tian
Electrical and Computer Engineering
Rutgers University
New Brunswick, NJ, USA
kasey.tian@rutgers.edu

***Abstract*—Subgradient methods**

***Index Terms*—optimization, subgradients, LASSO, regression**

I. Introduction

Subgradient methods are a way to perform an optimization on an objective function which is not fully differentiable. For non-differentiable objective functions, traditional methods, such as gradient descent and Newton's method, are impossible to execute. They can also be combined with a wide variety of other optimization methods, and have far reaching applications. They were originally developed by Shor and others in the Soviet Union in the 1960s and 70s [1].

II. Mathematical Basis

In this section we will explore the mathematics of subgradient methods and the LASSO Regression method.

A. Subgradient

1) Definitions: A subgradient is defined for some convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ at a point $x \in \operatorname{dom} f$ as a vector $g \in \mathbb{R}^n$ such that $\forall y \in \operatorname{dom} f$

$$f(y) \geq f(x) + g^T(y - x) \tag{1}$$

There can be multiple subgradients at a point x, so we will also define the subdifferential $\partial f(x)$ to be the set of all subgradients at x.

$$\partial f(x) = \left\{ g : f(y) \geq f(x) + g^T(y - x) \right\} \tag{2}$$

2) Example: Absolute Value: If we consider g in (1) to be a slope, we can visualize a subgradient as being some hyperplane intersecting our function at x for which all values of the function are on or above the plane.

Let us employ this intuition to find a subgradient of the function $f(x) = |x|$ at the point $x = 0$. Graphically, we can see in Fig. 1 that many different lines satisfy this criteron. In fact, we can say that any $g \in [-1, 1]$ would be a subgradient, and therefore $\partial f(0) = [-1, 1]$. But what about other points? For a point $x > 0$, we can surmise that the only possible $g = 1$, as any other value will leave some parts of our function beneath the resulting plane. Likewise for $x < 0$, $g = -1$. Thus we can write

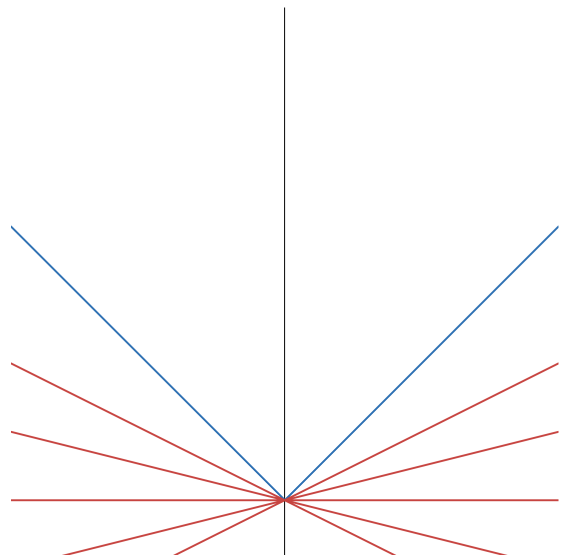

Fig. 1. Subgradients of absolute value function.

$$\partial f(x) = \begin{cases} -1 & x < 0 \\ [-1, 1] & x = 0 \\ 1 & x > 0 \end{cases} \tag{3}$$

This can be compared against the derivative of $f(x)$.

$$f'(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases} \tag{4}$$

We find that where the function is differentiable, the subdifferential contains only the gradient. There is only a difference where the function is not differentiable.

3) Properties: There are a few important properties that we shall take note of. Proofs of these properties can be found in Appendix A.

1) If the function is differentiable at x, the only member of $\partial f(x)$ will be the gradient

2) If x is a global minimum, $\partial f(x)$ must contain the zero vector

B. LASSO Regression

III. Code Implementation

IV. Numerical Results

V. Conclusion

Appendix

A. Subgradient Properties

1) :

B. Second Appendix

ha ha ha

References

[1] S. Boyd and J. Park, "Subgradient methods," May 2014. [Online]. Available: https://web.stanford.edu/class/ee364b/lectures/subgrad_method_notes.pdf