Group 12: Quadratic ℓ_1 Regularized Optimization with a Flexible Active-Set Strategy

Sreekar Sai Ranganathan (EE18B154), Anirudh R (EE18B103), Chandan Bhat (EE16D209)

I. Introduction

In areas like inverse problems, statistics and machine learning which are highly ill-posed, regularization plays an important role in mitigating the ill-posedness and avoiding over-fitting of the data. The problem that we explore here is the ℓ_1 regularised quadratic minimization problem (1).

$$\min_{x\in\mathbb{R}^n} F(x) \stackrel{\text{def}}{=} \frac{1}{2}x^{\mathrm{T}}Ax - b^{\mathrm{T}}x + \tau\|x\|_1 \tag{1}$$

In the literature, the quadratic - ℓ_1 problem has been widely discussed, especially for its ability to promote sparse solutions. Some of the methods which are very well known to solve the quadratic - ℓ_1 problem are the class of Iterative Soft-Thresholding Algorithms (ISTA) [1] and its accelerated variants [2].

Solution Strategy - The paper [3] proposes a new method that *flexibly* switches between the following two steps, leading to an efficient implementation.

1) Improving the active-set (non-zero elements) prediction using first order steps (ISTA); and
2) Exploring the current active-set through an inner conjugate gradient (CG) approach.

The active-set H, is a subspace of $\mathbb{R}^n$ with the same non-zero variables as x^{cg} is defined as:

$$H = \{x \mid x_i = 0,\ \forall\ i \text{ such that } x_i^{\mathrm{cg}} = 0\} \tag{2}$$

where x_i^{cg} is the point at which the CG step is started. The choice between these two steps is controlled by a *gradient balance condition* described below.

The minimum norm subgradient for the ISTA update at x (denoted by $v(x)$) is given by,

$$v(x) = \omega(x) + \psi(x) \tag{3}$$

where $\omega(x)$ and $\psi(x)$ contain the components of $v(x)$ corresponding to zero variables and non-zero variables of x respectively. The *gradient balance condition* checks the inequality $\|\omega(x)\|_2 \leq \|\psi(x)\|_2$ to flexibly switch between steps 1 and 2, which indicates whether adding more elements to the active set would be helpful or not. This approach implicitly incorporates second-order information of the function $F(x)$ in the optimization iterations.

Abstract for EE5121: Convex Optimization Course Project (July-Nov 2021), Electrical Engineering, IIT Madras

A. Objective

Through this project, our main objective is to get a good intuition for how the proposed algorithm works, and *why* it works, from both a theoretical & visual perspective. We also look to evaluate its improvements over a few related algorithms for solving (1).

1) Proposed Analyses & Visualizations: Here are some of the specific tasks we hope to achieve during the course of this project:

- Briefly explain the principle behind ISTAs, and how ISTA was accelerated in FISTA [2], creating relevant 2D/3D visualizations to aid in this.
- Visualizing how the *gradient balance condition*, which is the main contribution of this paper, works in different scenarios, in terms of, e.g. the number of CG cycles in each subspace phase, variation of the quantities $\|\psi\|$ and $\|\omega\|$, and the number of non-zero elements in x as the iterations progress.
- Compare the performance of the proposed technique with one of the common algorithms for ℓ_1 regularised quadratic problems, e.g. FISTA [2] in terms of accuracy achieved vs. number of matrix-vector (MV) computations.

B. Optimization concepts

In our project, we would cover the following general concepts in optimization:

- Iterative soft-thresholding algorithms (ISTA) and its variants (for ℓ_1 regularised optimization).
- Subspace optimization performed using conjugate gradient (CG) for a quadratic model of a function.
- Proximal operators

References

[1] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 11 2004.

[2] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems *," *Society for Industrial and Applied Mathematics*, vol. 2, no. 1, pp. 183–202, 2009.

[3] S. Solntsev, J. Nocedal, and R. H. Byrd, "An algorithm for quadratic 1-regularized optimization with a flexible active-set strategy," *http://dx.doi.org/10.1080/10556788.2015.1028062*, vol. 30, no. 6, pp. 1213–1237, 11 2015.