Accelerating the Gradient Projection Iterative Sketch for large scale constrained Least-squares

Junqi Tang, Mohammad Golbabae, Mike Davies
Institutte of Digital Communications, The University of Edinburgh, EH9 3JE, UK

Abstract—This paper proposes an accelerated sketched gradient method [1] which was based on combining a combination of the meta-algorithms Classical Sketch (CS) [2] and Iterative Hessian Sketch (IHS) [3] with the Projected / Proximal Gradient Descent (PGD) algorithm and Nesterov’s acceleration scheme for efficiently solving large scale constrained Least-squares and regularized Least-squares. As a first order solver, the PGD can provide us flexibility in handling the constraints and scalability in computation. The proposed algorithm satisfies a number of our expectations as an efficient large scale constrained/regularized LS solver, which are mainly inherited from the scalability and flexibility of the PGD combined with dimensionality reducing properties of the sketching techniques: (a) computational efficiency, (b) efficiency on high speed storage, and (c) flexibly to incorporate a wide range of constraints and non-smooth regularization.

I. INTRODUCTION

Consider a noisy linear measurement model for a vector $x_{gt}$ (ground truth) which belongs to a convex constrained set $K$, an $n$ by $d$ linear operator matrix $A$, and additive noise denoted by $w \in \mathbb{R}^{n \times 1}$:

$$y = Ax_{gt} + w, \quad x_{gt} \in K, \quad A \in \mathbb{R}^{n \times d}. \tag{1}$$

In the context of imaging applications such as CT or MRI, the vector $y$ denotes a set of $n$ physical measurements collected from an image $x_{gt}$ through the measurement operator $A$, and in the context of machine learning, $A$ is often a training data matrix used for setting the regression parameters $x_{gt}$ from the observations $y$. The Least-square (LS) estimator for $x_{gt}$ is:

$$x^* = \arg \min_x \|y - Ax\|_2^2 + f_K(x). \tag{2}$$

where the convex (could be non-smooth) function $f_K$ enforces the constraint into the Least-squares estimator. If the constraint is exactly known, the $f_K$ can be set as the indicator function of the set $K$, if not, we can set it as a regularizer.

II. SKETCHED GRADIENT WITH NESTEROV’S ACCELERATION SCHEME

A standard first order solver for (2) is the PGD algorithm which can be defined for any convex constrained set $K$, as long as the projection (or proximal operation) onto the set is efficient:

$$x_{j+1} = \text{Prox}_{f_K}(x_j - \eta A^T (Ax_j - y)). \tag{3}$$

The PGD is known to be flexible to various constraint sets, but it faces two major challenges: 1) when the operator $A$ is large, the computational cost of the iterates can be large; 2) when $A$ is ill-conditioned, the PGD may take a very large number of iterations to converge. Moreover when the computational cost of the projection/proximal operator is non-trivial, we also wish to reduce the number of iterations as much as possible (the stochastic gradient algorithms usually demands a small batch size which will lead to a large number of iterations). The proposed algorithm is aimed at tackling both reducing the cost of the gradient calculation and the number of iterations.

Algorithm 1:

Initialization: $p_0 = 1$ for all $t$, $z^0_0 = 0$, $z^0 = 0$ ;
Given $A \in \mathbb{R}^{n \times d}$, sketch size $m \ll n$;
Generate a random sketching matrix $S^0 \in \mathbb{R}^{m \times n}$;
Calculate $S^0 A$, $S^0 y$;
while $i = 0 : k_0 - 1$ do
$$x^0_{i+1} = \text{Prox}_{f_K}(z^0_i - \eta (S^0 A)x^0_i - S^0 y));$$
$$p_{i+1} = (\eta p_i)^2 + \sqrt{(\eta p_i)^4 + 4(\eta p_i)^2};$$
$$\tau_{i+1} = \frac{p_{i+1}}{(\eta p_i)^2 + p_{i+1}^2};$$
$$z^0_{i+1} = x^0_{i+1} + \tau_{i+1}(x^0_i - x^0_i);$$
end
$$x^0_{1} = z^0_{1} = x^0_{0};$$
while $t = 1 : N - 1$ do
Calculate $g = A^T (Ax^0_t - y)$;
Generate a random sketching matrix $S^t \in \mathbb{R}^{n \times m}$;
Calculate $A^t = S^t A$;
while $i = 0 : k_t - 1$ do
$$x^t_{i+1} = \text{Prox}_{f_K}(z^t_i - \eta (A^t x^t_i - x^0_i + mg));$$
$$p_{i+1} = (\eta p_i)^2 + \sqrt{(\eta p_i)^4 + 4(\eta p_i)^2};$$
$$\tau_{i+1} = \frac{p_{i+1}}{(\eta p_i)^2 + p_{i+1}^2};$$
$$z^t_{i+1} = x^t_{i+1} + \tau_{i+1}(x^t_i - x^t_i);$$
end
$$x^t_{1} = z^t_{1} = x^t_{0};$$
end
Return $x^N_t$ ;

As shown in the Algorithm 1, we use the classical sketching [2] and iterative sketching [3] framework as we have done in the [1], and then since the sketched least-square problem is fixed in every outer loop, the Nesterov’s acceleration scheme [4][5] is in theory directly applicable to provide acceleration in both strict constrained setting and proximal setting.

We have also tested its performance through numerical experiments, and observe that the proposed algorithm achieves a further speed-up onto the GPS / GPS-prox algorithm in all the experiments.

ACKNOWLEDGMENT

Junqi Tang would like to acknowledge the support from H2020-MSCA-ITN Machine Sensing Training Network (MacSeNet), project 642685
Fig. 1. Experimental results on a synthetic $l_1$ constrained Least-square regression problem

Fig. 2. Experimental results on a fan-beam CT image reconstruction (Regularized least-squares)

REFERENCES


