On Some Versions of Subspace Optimization Methods with Inexact Gradient Information

Ilya Kuruzov [1,2] Fedor Stonyakin [1]
Moscow Institute of Physics and Technology[1] Innopolis University[2]
kuruzov.ia@phystech.edu fedyor@mail.ru

Introduction

It is well-known that accelerated gradient first order methods possess optimal complexity estimates for the class of convex smooth minimization problems. In many practical situations, it makes sense to work with inexact gradients. However, this can lead to the accumulation of corresponding inexactness in the theoretical estimates of the rate of convergence. We propose some modification of the methods for convex optimization with inexact gradient based on the subspace optimization sush as Nemirovski's Conjugate Gradients and Sequential Subspace Optimization. We research the convergence for different condition of inexactness both in gradient value and accuracy of subspace optimization problems. Besides this, we investigate generalization of this result to the class of quasar-convex (weakly-quasi-convex) functions.

Contributions

1. Linear convergence for inexact CG method in non-convex case with.
2. Complexity of auxiliary problems for SESOP and CG methods in convex case using Ellipsoid Method and Multidimensional dichotomy.

1 Problem Statement

We consider an optimization problem

$$\min f(x)$$

where f is L-smooth function ($\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x-y\| \quad \forall x, y \in \mathbb{R}^n$) and gamma-quasar function

$$f(x^*) \geq f(x) + \frac{1}{\gamma}\langle \nabla f(x), x^* - x\rangle. \tag{1}$$

where $\gamma \in (0, 1]$ and x^* is a minimizer.

In our paper we consider methods that can work with inexact gradient $g : \mathbb{R}^n \to \mathbb{R}^n$:

$$\|g(x) - \nabla f(x)\| \leqslant \delta. \tag{2}$$

Besides, we will consider the following conditions:

- PL-condition $f(x) - f^* \leqslant \frac{1}{2\mu}\|\nabla f(x)\|^2 \quad \forall x \in \mathbb{R}^n$,
- Quadratic Growth Condition $f(x) - f^* \geqslant \frac{\mu}{2}\|x - x^*\|^2 \quad \forall x \in \mathbb{R}^n$.

2 Subspace Optimization Method

2.1 A modification of the SESOP method with an inexact gradient

Let us define $\mathbf{d}_k^0 = g(x_k)$, $\mathbf{d}_k^1 = x_k - x_0$, $\mathbf{d}_k^2 = \sum_{i=0}^{k} \omega_i g(x_i)$ and $w_{k+1} = \frac{1}{2} + \sqrt{\frac{1}{4} + w_k^2}$ with $w_0 = 1$. Then update variable in SESOP method is given by the following expression:

$$\tau_k \leftarrow \arg\min_{\tau \in \mathbb{R}^3} f\left(x_k + \sum_{i=1}^{3} \tau_1 d_k^{i-1}\right)$$
$$x_{k+1} \leftarrow x_k + \sum_{i=1}^{3} \tau_1 d_k^{i-1}$$

2.2 A modification of Nemirovski's Conjugate Gradient method with an inexact gradient

Let us define $\mathbf{q}_k = \mathbf{q}_{k-1} + g(\hat{x}_k)$. Then one iteration of CG method is given by the following form:

$$\hat{x}_k \leftarrow \arg\min_{x \in \mathcal{X}_k} f(x), \quad \text{where } \mathcal{X}_k = x_0 + \mathrm{Lin}(x_k - x_0, \mathbf{q}_k)$$
$$x_k = \hat{x}_k - \frac{1}{2L} g(\hat{x}_k)$$

3 Main Results

Previous paper states that SESOP method does not accumulate error. At this work, we provide estimation for number of calculations low-dimensional gradient.

Theorem 1. *To approach quality ε on initial problem by SESOP method one requires not more than $N = \left\lceil \sqrt{\frac{40LR^2}{\gamma^2\varepsilon}} \right\rceil$ of inexact gradient calculations with respect to x and not more than $M = \left\lceil 18N \ln \frac{12800LBC_N}{\varepsilon^4} \right\rceil$ of inexact gradient calculations with respect to τ.*

It gives us a method that has convergence rate similar to accelerated methods but this method does not accumulate error. This is achieved by solving an auxiliary low-dimensional problem. The second considered method is Conjugate Gradient with restarts and stopping rule.

Theorem 2. *Let assumptions of Theorem **??** hold and all subproblems are convex. Besides, there is R such that $\hat{x}_k - x_k \in B_{R_x}$ for all k. Each point $\hat{x}_k$ is output of two-dimensional dichotomy algorithm (see [**?**]) after M steps, where M is given by $M = \left\lceil 16\left(\ln \frac{CR_x}{\varepsilon^4}\right)^2 \right\rceil$*

Let one of the following alternatives hold:

1. *CG method makes $K = \left\lceil \frac{2}{1-\alpha} \log \frac{1}{\varepsilon} \right\rceil$ restarts and $T = \left\lceil \frac{8}{\gamma}\sqrt{\frac{L}{\mu}\frac{\sqrt{1+\alpha}}{1-\alpha}} \right\rceil$, iterations per each restart, where $\varepsilon = \frac{64}{\gamma^2\mu}\delta_1^2$*
2. *For some iteration $N \leq N^*$, at the N-th iteration of Nemirovski's Conjugate Gradient method, the stopping criterion $\|g(x_N)\| \leq \frac{8}{\gamma}\delta_1$ is satisfied for the first time.*

Then for the output point $\widehat{x}$ ($\widehat{x} = x_N$ or $\widehat{x} = x_{N_*}$) of Nemirovski's Conjugate Gradient method, the following inequalities hold: $f(\widehat{x}) - f^* \leqslant \frac{64\delta_1^2}{\gamma^2\mu}$.

As the result the algorithm requires not more N of calculations of inexact gradient with respect to x and $MN = O(\ln^3(1/\varepsilon))$ of low-dimensional inexact gradient calculations.

4 Numerical Experiments

We consider the problem of logistic regression with ℓ_2-regularization:

$$f(x) = (1/m)\sum_{j=1}^{m} \log(1 + \exp(-y_j\langle f_j, x\rangle)) + \mu\|x\|^2 \tag{3}$$

δ_1	SESOP	CG+Ellipsoids	CG+Dichotomy	STM
10^{-3}	1	1.4	0.9	1.7
10^{-5}	10.1	15.3	9.5	13.8
10^{-7}	35.3	60.9	36.8	42.1

Time comparison (s) for problem (3)

Subspace methods outperform Similar Triangle Method.