Subsampled cubic regularization method for finite-sum minimization

M.L.N. Gonçalves *

November 16, 2022

Abstract

This paper proposes and analyzes a subsampled Cubic Regularization Method (CRM) for solving finite-sum optimization problems. The new method uses random subsampling techniques to approximate the functions, gradients and Hessians in order to reduce the overall computational cost of the CRM. Under suitable hypotheses, first- and second-order iteration-complexity bounds and global convergence analyses are presented. We also discuss the local convergence properties of the method. Numerical experiments are presented to illustrate the performance of the proposed scheme.

1 Introduction

Consider the following finite-sum optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{d}\sum_{i=1}^{d} f_i(x), \tag{1}$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a twice-differentiable function, potentially nonconvex. We are interested in the case of very large dimension d. We denote the optimal value of (1) by f^* and assume that $f^* > -\infty$. Problem (1) has been the object of intense research in the last decades since many important machine learning and statistics applications can be modeled in this form.

A plethora of methods has been proposed for solving the aforementioned optimization problem, including first- and second-order procedures. Since the evaluations of the objective function f and its first- and second-order derivatives may be computationally expensive for large value of d, the most efficient algorithms for solving (1) are those that take advantages of the special structure of (1) by considering approximations of f and/or its gradient and/or Hessian formed by a subset of the functions $\{f_1, \ldots, f_d\}$. The last approach is called subsampling strategy and it has been used, for example, in the deterministic and probabilistic methods of [5, 6, 20] and [2, 3, 4, 9, 11, 17, 19, 22], respectively.

In this work, we are interested in the Cubic Regularization Method (CRM), see [15, 18], which is a globally convergent variant of the Newton method for unconstrained minimization of a twice continuously differentiable function h. Basically, at the t-th iteration, the next iterate x_{t+1} of the

*IME, Universidade Federal de Goiás, Campus II- Cx. Postal 131, CEP 74001-970, Goiânia-GO, Brazil (maxlng@ufg.br). The work was supported in part by CNPq Grants 405349/2021-1 and 304133/2021-3.

CRM is obtained by minimizing a *cubic model* that consists of a third-order regularization of the second-order Taylor approximation of the objective function $h(\cdot)$ around x_t, i.e.,

$$\min_{y\in\mathbb{R}^d} h(x_t) + \langle \nabla h(x_t), y - x_t\rangle + \frac{1}{2}\langle \nabla^2 h(x_t)(y-x_t), y-x_t\rangle + \frac{\sigma_t}{6}\|y-x_t\|^3, \tag{2}$$

where σ_t is the regularization parameter. An attractive feature of the cubic regularization methods is that, for a given tolerance $\epsilon > 0$, it takes at most $\mathcal{O}\left(\epsilon^{-3/2}\right)$ iterations to generate an ϵ-approximate stationary point of the objective function h (i.e., an iterate x_t such that $\|\nabla h(x_t)\|_2 \leq \epsilon$), when $h(\cdot)$ is a nonconvex function with Lipschitz continuous Hessian; see, for example, [18]. Remarkably, work [8] proved that in the same problem class the standard Newton method (without regularization) may need a number of iterations arbitrarily close to $\mathcal{O}\left(\epsilon^{-2}\right)$ to generate an ϵ-approximate stationary point of the objective function h.

Here, in order to reduce the overall computational cost of the CRM for solving (1), we propose a variant of it in which the function f and its gradient and Hessian are approximated by random subsampling techniques. Essentially, the cubic model in (2) with $h = f$ is replaced by

$$\min_{y\in\mathbb{R}^d} f_{\mathcal{G}}(x_t) + \langle \nabla f_{\mathcal{G}}(x_t), y - x_t\rangle + \frac{1}{2}\langle \nabla^2 f_{\mathcal{H}}(x_t)(y-x_t), y-x_t\rangle + \frac{\sigma_t}{6}\|y-x_t\|^3$$

where

$$f_{\mathcal{G}}(x) := \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} f_i(x), \quad \nabla f_{\mathcal{G}}(x) := \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \nabla f_i(x), \quad \nabla^2 f_{\mathcal{H}}(x) := \frac{1}{|\mathcal{H}|}\sum_{i\in\mathcal{H}} \nabla^2 f_i(x), \tag{3}$$

$\mathcal{H} \subset \mathcal{G} \subset \{1,\ldots,d\}$ are the subsamples, and $|\mathcal{H}|$ and $|\mathcal{G}|$ are their cardinality, respectively. By assuming that the Hessian $\nabla^2 f_{\mathcal{H}}(x)$ is Lipschitz continuous for every $\mathcal{H} \subset \{1,\ldots,d\}$ and that the sequence $\{\nu_t\}$ defined by

$$\nu_t := \max\{|f_{\mathcal{G}_t}(x_{t+1}) - f(x_{t+1})|, |f_{\mathcal{G}_t}(x_t) - f(x_t)|\}, \quad \forall t \geq 0,$$

is summable, it is shown that the new algorithm needs at most $\mathcal{O}\left(\epsilon^{-3/2}\right)$ calls of the oracle [1] to generate an iterate x_t such that $\|\nabla f_{\mathcal{G}_t}(x_t)\| \leq \epsilon$. Moreover, it is proven that an iterate x_t satisfying

$$\|\nabla f_{\mathcal{G}_t}(x_t)\| < \epsilon_g \quad \text{and} \quad \lambda_{min}\left(\nabla^2 f_{\mathcal{H}_t}(x_t)\right) > -\epsilon_H,$$

is generated in at most $\mathcal{O}(\max\{\epsilon_g^{-\frac{3}{2}}, \epsilon_g^{-3}\})$ calls of the oracle. We remark that the same order of iteration-complexity bounds of the full CRM (i.e., $\mathcal{G}_t = \mathcal{H}_t = \{1,\ldots,d\}$ for all $t \geq 0$) are obtained for the new method, in spite of inaccuracy in the functions, gradients and Hessians. Global convergence properties for finding approximate first- and second-order stationary points are also discussed under the aforementioned assumptions. We further discuss complexity estimates as well as global convergence results of the proposed algorithm under a condition weaker than the summability of the sequence $\{\nu_t\}$. The local quadratic convergence rate of the method is also established under standard hypotheses. Some numerical experiments, including comparisons with the subsampled adaptive cubic regularization method in [3], are presented in order to illustrate the performance of the method. In particular, it is verified that the use of subsampling techniques to approximate the

[1] Throughout this work, *a call of the oracle* means the partial or total evaluation of one of following terms $f(\cdot)$, $\nabla f(\cdot)$ and $\nabla^2 f(\cdot)$.

function f as well as its gradient and Hessian, improve considerably the numerical behavior of the cubic regularization method for solving (1).

This work is organized as follows. Section 2 describes the subsampled cubic regularization method and presents its first-order iteration-complexity bounds and global convergence analyses. Subsection 2.1 is devoted to the second-order results, while the local convergence is established in Subsection 2.2. Section 3 presents some numerical experiments of the proposed method and some concluding remarks are given in Section 4.

2 Subsampled cubic regularization method

In this section, we formally state the subsampled cubic regularization method for computing approximate solutions of (1) and present its first- and second-order iteration-complexity bounds. Global and local convergence properties of the method are also discussed.

We start by describing the new method.

Algorithm 1. ***(Subsampled-CRM)***

Step 0. Choose $x_0 \in \mathbb{R}^n$, $\theta \geq 0$, $\sigma_0 > 0$, subsamples $\mathcal{H}_0 \subset \mathcal{G}_0 \subset \{1, \ldots, d\}$, and set $t := 0$.

Step 1. Construct $f_{\mathcal{G}_t}(x_t)$ and $\nabla f_{\mathcal{G}_t}(x_t)$ as in (3).

Step 2. Find the smallest integer $i \geq 0$ such that $2^i \sigma_t \geq 2\sigma_0$. Choose subsample $\mathcal{H}_t$ so that $|\mathcal{H}_t| \geq |\mathcal{H}_0|$ and $\mathcal{H}_t \subset \mathcal{G}_t$, and set $\mathcal{H}_{t,i} \leftarrow \mathcal{H}_t$.

Step 2.1. Construct $\nabla f^2_{\mathcal{H}_{t,i}}(x_t)$ as in (3) and compute an approximate solution $x^+_{t,i}$ of the subproblem

$$\min_{y \in \mathbb{R}^d} M^{\mathcal{G}_t, \mathcal{H}_{t,i}}_{x_t, 2^i\sigma_t}(y) := f_{\mathcal{G}_t}(x_t) + \langle \nabla f_{\mathcal{G}_t}(x_t), y - x_t \rangle + \frac{1}{2}\langle \nabla^2 f_{\mathcal{H}_{t,i}}(x_t)(y - x_t), y - x_t \rangle + \frac{2^i \sigma_t}{6}\|y - x_t\|^3 \quad (4)$$

such that

$$M^{\mathcal{G}_t, \mathcal{H}_{t,i}}_{x_t, 2^i\sigma_t}(x^+_{t,i}) \leq f_{\mathcal{G}_t}(x_t) \text{ and } \|\nabla M^{\mathcal{G}_t, \mathcal{H}_{t,i}}_{x_t, 2^i\sigma_t}(x^+_{t,i})\| \leq \theta \min\left\{\|x^+_{t,i} - x_t\|^2, \|\nabla f_{\mathcal{G}_t}(x_t)\|\right\}. \quad (5)$$

Step 2.2. Compute $f_{\mathcal{G}_t}(x^+_{t,i})$ and $\nabla f_{\mathcal{G}_t}(x^+_{t,i})$. If

$$f_{\mathcal{G}_t}(x_t) - f_{\mathcal{G}_t}(x^+_{t,i}) \geq \frac{2^i \sigma_t}{12}\|x^+_{t,i} - x_t\|^3 \quad (6)$$

and

$$\|\nabla f_{\mathcal{G}_t}(x^+_{t,i})\| \leq \left(\frac{3(2^i \sigma_t)}{4} + \sigma_0 + \theta\right)\|x^+_{t,i} - x_t\|^2 \quad (7)$$

hold, set $i_t = i$, and go to Step 3. Otherwise, choose $\mathcal{H}_{t,i+1}$ so that

$$|\mathcal{H}_{t,i+1}| = \min\{\lceil 2^i \sigma_t \rceil |\mathcal{H}_t|, |\mathcal{G}_t|\}, \quad \mathcal{H}_{t,i+1} \subset \mathcal{G}_t$$

and set $i := i + 1$ and go to Step 2.1.

Step 3. Choose subsample $\mathcal{G}_t^+$ such that $|\mathcal{G}_t^+| \geq |\mathcal{G}_0|$. Set $x_{t+1} := x_{t,i_t}^+$, $\mathcal{G}_{t+1} := \mathcal{G}_t^+$, $\sigma_{t+1} := 2^{i_t-1}\sigma_t$, $t := t+1$, and go to Step 1.

Some comments are in order. (i) As will be shown, the well-definedness of inner loop in Step 2 will be guaranteed, in particular, by the fact that $\mathcal{H}_{t,i} = \mathcal{G}_t$ at some inner iteration i. In Step 3, the sample size $|\mathcal{G}_t|$ does not necessary increase in the next iteration although it is expected that the full sample size is reached at some iteration t. We refer the reader to [5, 6, 20], where some updates rules for the subsamples sequences are discussed. (ii) Since the first condition in (5) is equivalent to $M_{x_t,2^i\sigma_t}^{\mathcal{G}_t,\mathcal{H}_t}(x_{t,i}^+) \leq M_{x_t,2^i\sigma_t}^{\mathcal{G}_t,\mathcal{H}_t}(x_t)$, we obtain that $x_{t,i}^+$ gives a decrease of the subsampled model in (4). While, the second condition in (5) means that $x_{t,i}^+$ is an approximate first-order stationary point for (4). We mention that conditions in (5) are similar to the ones used in [7, 14].

In what follows, we will present some convergence properties of Algorithm 1. The following assumption is made throughout this work:

(A1) The Hessian $\nabla^2 f_{\mathcal{H}}$ is L-Lipschitz continuous for every $\mathcal{H} \subset \{1, \ldots, d\}$, i.e.,

$$\|\nabla^2 f_{\mathcal{H}}(y) - \nabla^2 f_{\mathcal{H}}(x)\| \leq L\|y - x\|, \quad \forall x, y \in \mathbb{R}^n.$$

We begin by proving an auxiliar result.

Lemma 2.1. *For given $x \in \mathbb{R}^n$, $\sigma > 0$ and $\mathcal{H} \subset \mathcal{G} \subset \{1, \ldots, d\}$, consider*

$$M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(y) = f_{\mathcal{G}}(x) + \langle \nabla f_{\mathcal{G}}(x), y - x\rangle + \frac{1}{2}\langle \nabla f_{\mathcal{H}}^2(x)(y - x), y - x\rangle + \frac{\sigma}{6}\|y - x\|^3. \tag{8}$$

Assume that $x^+ \in \mathbb{R}^n$ satisfies

$$M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+) \leq f_{\mathcal{G}}(x) \quad \textit{and} \quad \|\nabla M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+)\| \leq \theta\|x^+ - x\|^2. \tag{9}$$

If

$$\|\nabla^2 f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{H}}(x)\| \leq \sigma_0\|x^+ - x\|, \qquad \sigma \geq 2(L + 3\sigma_0), \tag{10}$$

then

$$f_{\mathcal{G}}(x) - f_{\mathcal{G}}(x^+) \geq \frac{\sigma}{12}\|x^+ - x\|^3, \quad \|\nabla f_{\mathcal{G}}(x^+)\| \leq \left(\frac{3\sigma}{4} + \sigma_0 + \theta\right)\|x^+ - x\|^2. \tag{11}$$

If additionally

$$\nabla f_{\mathcal{H}}^2(x) + \frac{\sigma}{2}\|x^+ - x\|I \succeq -\theta\|x^+ - x\|I, \tag{12}$$

then

$$-\lambda_{min}\left(\nabla^2 f_{\mathcal{G}}(x^+)\right) \leq \frac{\sigma + 2(\theta + \sigma_0 + L)}{2}\|x^+ - x\|. \tag{13}$$

Proof. It follows from **A1** that

$$f_{\mathcal{G}}(x^+) \leq f_{\mathcal{G}}(x) + \langle \nabla f_{\mathcal{G}}(x), x^+ - x\rangle + \frac{1}{2}\langle \nabla^2 f_{\mathcal{G}}(x)(x^+ - x), x^+ - x\rangle + \frac{L}{6}\|x^+ - x\|^3 \tag{14}$$

and

$$\left\|\nabla f_{\mathcal{G}}(x^+) - \nabla f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{G}}(x)(x^+ - x)\right\| \leq \frac{L}{2}\|x^+ - x\|^2. \tag{15}$$

Using (14), the definition of $M_{x,\sigma}^{\mathcal{G},\mathcal{H}}$ in (8), and the first inequalities in (9) and (10), we have

$$\begin{aligned} f_{\mathcal{G}}(x^+) &\leq M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+) + \frac{1}{2}\langle(\nabla^2 f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{H}}(x))(x^+ - x), x^+ - x\rangle + \frac{L-\sigma}{6}\|x^+ - x\|^3 \\ &\leq f_{\mathcal{G}}(x) + \frac{1}{2}\|\nabla^2 f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{H}}(x)\|\|x^+ - x\|^2 + \frac{L-\sigma}{6}\|x^+ - x\|^3 \\ &\leq f_{\mathcal{G}}(x) + \left(\frac{\sigma_0}{2} + \frac{L-\sigma}{6}\right)\|x^+ - x\|^3, \end{aligned}$$

which, combined with the second inequality in (10), proves the first inequality in (11).

Now, using the definition of $M_{x,\sigma}^{\mathcal{G},\mathcal{H}}$ in (8), the second inequality in (9), the first inequality in (10), and (15), we get

$$\begin{aligned} \|\nabla f_{\mathcal{G}}(x^+)\| &\leq \|\nabla f_{\mathcal{G}}(x^+) - \nabla M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+)\| + \|\nabla M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+)\| \\ &= \left\|\nabla f_{\mathcal{G}}(x^+) - \nabla f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{H}}(x)(x^+ - x) - \frac{\sigma}{2}\|x^+ - x\|(x^+ - x)\right\| + \|\nabla M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+)\| \\ &\leq \left\|\nabla f_{\mathcal{G}}(x^+) - \nabla f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{H}}(x)(x^+ - x)\right\| + \frac{\sigma}{2}\|x^+ - x\|^2 + \|\nabla M_{x,\sigma}^{\mathcal{G},\mathcal{H}}(x^+)\| \\ &\leq \left\|\nabla f_{\mathcal{G}}(x^+) - \nabla f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{G}}(x)(x^+ - x)\right\| + \|\nabla^2 f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{H}}(x)\|\|x^+ - x\| \\ &+ \left(\frac{\sigma}{2} + \theta\right)\|x^+ - x\|^2 \\ &\leq \left(\frac{L+\sigma}{2} + \sigma_0 + \theta\right)\|x^+ - x\|^2, \end{aligned}$$

which, combined with the second inequality in (10), proves the second inequality in (11).

On the other hand, by **A1** and the first inequality in (10), for any $d \in \mathbb{R}^n$, we have

$$\begin{aligned} \langle\left(\nabla f_{\mathcal{H}}^2(x) - \nabla^2 f_{\mathcal{G}}(x^+)\right)d, d\rangle &= \langle\left(\nabla f_{\mathcal{H}}^2(x) - \nabla^2 f_{\mathcal{G}}(x)\right)d, d\rangle + \langle\left(\nabla^2 f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{G}}(x^+)\right)d, d\rangle \\ &\leq \|\nabla f_{\mathcal{H}}^2(x) - \nabla^2 f_{\mathcal{G}}(x)\|\|d\|^2 + \|\nabla^2 f_{\mathcal{G}}(x) - \nabla^2 f_{\mathcal{G}}(x^+)\|\|d\|^2 \\ &\leq \langle(\sigma_0 + L)\,\|x^+ - x\| I\, d, d\rangle. \end{aligned}$$

Since the inequality above holds for all $d \in \mathbb{R}^n$, it follows that

$$\nabla f_{\mathcal{H}}^2(x) \preceq \nabla^2 f_{\mathcal{G}}(x^+) + (\sigma_0 + L)\,\|x^+ - x\| I,$$

which, using the Weyl's inequality [10], yields

$$\lambda_{min}\left(\nabla f_{\mathcal{H}}^2(x)\right) \leq \lambda_{min}\left(\nabla^2 f_{\mathcal{G}}(x^+)\right) + (\sigma_0 + L)\,\|x^+ - x\|. \tag{16}$$

Now, assuming that (12) is true, we also have

$$\lambda_{min}\left(\nabla f_{\mathcal{H}}^2(x)\right) \geq -\left(\frac{\sigma}{2} + \theta\right)\|x^+ - x\|. \tag{17}$$

Finally, combining (16) and (17), we obtain (13).

□

The next lemma shows that the inner procedure in Step 1 stops in a finite number of trials. Moreover, it provides an estimate of the number of function, gradient and Hessian evaluations after a certain number of iterations.

Lemma 2.2. *The sequence of regularization parameters $\{\sigma_t\}$ in Algorithm 1 satisfies*

$$\sigma_0 \leq \sigma_t \leq \max\{2(L+3\sigma_0), 1/\alpha\} := \sigma_{max}, \quad \forall t \geq 0, \tag{18}$$

where α is such that $|\mathcal{H}_0| = \alpha d$. Moreover, the number O_T of calls of the oracle after T iterations is bounded as follows:

$$O_T \leq 3[3T + \log_2(\sigma_{max}) - \log_2(\sigma_0)]. \tag{19}$$

Proof. Clearly, (18) is true for $t = 0$. Suppose that (18) is true for some $t \geq 0$. If $i_t = 0$, it follows from Step 1 and the induction assumption that

$$\sigma_0 \leq \sigma_{t+1} = \frac{1}{2}\sigma_t \leq \sigma_t \leq \max\{2(L+3\sigma_0), 1/\alpha\},$$

which proves that (18) holds for $t+1$. Now, if $i_t \geq 1$, then we must have

$$2^{i_t-1}\sigma_t \leq \max\{2(L+3\sigma_0), 1/\alpha\}. \tag{20}$$

Indeed, assuming by contradiction that (20) is not true, that is

$$2^{i_t-1}\sigma_t > 2(L+3\sigma_0), \quad 2^{i_t-1}\sigma_t > 1/\alpha. \tag{21}$$

Hence, as $|\mathcal{H}_0| = \alpha d$ and $|\mathcal{H}_t| \geq |\mathcal{H}_0|$, the second inequality in (21) yields that

$$2^{i_t-1}\sigma_t|\mathcal{H}_t| \geq 2^{i_t-1}\sigma_t|\mathcal{H}_0| > d \geq |\mathcal{G}_t|,$$

which in turn implies that $|\mathcal{H}_{t,i_t-1}| = |\mathcal{G}_t|$. So, $\|\nabla^2 f_{\mathcal{G}_t}(x_{t,i}^+) - \nabla^2 f_{\mathcal{H}_{t,i_t-1}}(x_{t,i}^+)\| = 0$. Therefore, by combining the last equality, the first equality in (21) and Lemmas 2.1 with $x := x_t$, $x^+ := x_{t,i}^+$, $\mathcal{G} := \mathcal{G}_t$, $\mathcal{H} := \mathcal{H}_{t,i_t-1}$ and $\sigma := 2^{i_t-1}\sigma_t$, we obtain that the inequalities in (6) and (7) would have been satisfied for $i = i_t - 1$, contradicting the minimality of i_t. Therefore, (20) is true.

Finally, note that at the t-th iteration of Algorithm 1 the number of calls of the oracle is bounded by $2 + 3(i_t + 1)$ times. On the other hand,

$$\sigma_{t+1} = 2^{i_t-1}\sigma_t \Longrightarrow 3(i_t+1) + 2 = 3\left[2 + \log_2(\sigma_{t+1}) - \log_2(\sigma_t)\right] + 2.$$

Thus,

$$\begin{aligned} O_T &\leq \sum_{t=0}^{T}[3(i_t+1)+2] \leq \sum_{t=0}^{T} 9 + 3\log_2(\sigma_{T+1}) - 3\log_2(\sigma_0) \\ &\leq 3[3T + \log_2(\sigma_{max}) - \log_2(\sigma_0)], \end{aligned}$$

where the last inequality is due to (18). $\square$

It follows from (19) that

$$\frac{O_T}{T} \leq 9 + \frac{3}{T}[\log_2(\sigma_{max}) - \log_2(\sigma_0)],$$

which implies that the average number of oracle calls per inner iteration, up to the T-th outer iteration, is asymptotically bounded by 9.

We next present a key result for our analysis.

Lemma 2.3. *Let $\{x_t\}_{t=1}^T$ be generated by Algorithm 1 and define*

$$\nu_t := \max\{|f_{\mathcal{G}_t}(x_{t+1}) - f(x_{t+1})|,\, |f_{\mathcal{G}_t}(x_t) - f(x_t)|\}, \quad \forall t \geq 0. \tag{22}$$

Then,

$$f(x_t) - f(x_{t+1}) \geq \frac{\sigma_{t+1}}{6}\|x_{t+1} - x_t\|^3 - 2\nu_t, \quad t = 0, \ldots, T-1, \tag{23}$$

and

$$\sum_{t=0}^{T-1} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\|^{\frac{3}{2}} \leq \frac{6(f(x_0) - f^* + 2\sum_{t=0}^{T-1}\nu_t)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{\sigma_0}. \tag{24}$$

Proof. By (6) and $\sigma_{t+1} = 2^{i_t - 1}\sigma_t$, we have

$$f_{\mathcal{G}_t}(x_t) - f_{\mathcal{G}_t}(x_{t+1}) \geq \frac{\sigma_{t+1}}{6}\|x_{t+1} - x_t\|^3, \quad t = 0, \ldots, T-1.$$

On the other hand, it follows from (22) that

$$\begin{aligned} f_{\mathcal{G}_t}(x_t) - f_{\mathcal{G}_t}(x_{t+1}) &= f(x_t) - f(x_{t+1}) + f_{\mathcal{G}_t}(x_t) - f(x_t) - f_{\mathcal{G}_t}(x_{t+1}) + f(x_{t+1}) \\ &\leq f(x_t) - f(x_{t+1}) + |f_{\mathcal{G}_t}(x_t) - f(x_t)| + |f_{\mathcal{G}_t}(x_{t+1}) + f(x_{t+1})| \\ &\leq f(x_t) - f(x_{t+1}) + 2\nu_t. \end{aligned}$$

By combining the above inequalities, we find

$$f(x_t) - f(x_{t+1}) + 2\nu_t \geq \frac{\sigma_{t+1}}{6}\|x_{t+1} - x_t\|^3, \quad t = 0, \ldots, T-1,$$

which is equivalent to the inequality in (23).

Summing up the inequalities in (23) and using the definition of f^* and (18), we get

$$\begin{aligned} f(x_0) - f^* + 2\sum_{t=0}^{T-1}\nu_t &\geq f(x_0) - f(x_T) + 2\sum_{t=0}^{T-1}\nu_t \\ &\geq \sum_{t=0}^{T-1}\frac{\sigma_{t+1}}{6}\|x_{t+1} - x_t\|^3 \\ &\geq \frac{\sigma_0}{6}\sum_{t=0}^{T-1}\|x_{t+1} - x_t\|^3, \end{aligned}$$

and so

$$\sum_{t=0}^{T-1}\|x_{t+1} - x_t\|^3 \leq \frac{6(f(x_0) - f^* + 2\sum_{t=0}^{T-1}\nu_t)}{\sigma_0}. \tag{25}$$

On the other hand, it follows from (7), the fact that $\sigma_{t+1} = 2^{i_t - 1}\sigma_t$ and (18) that

$$\|\nabla f_{\mathcal{G}_t}(x_{t+1})\| \leq \left(\frac{3\sigma_{t+1}}{2} + \sigma_0 + \theta\right)\|x_{t+1} - x_t\|^2 \leq \left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)\|x_{t+1} - x_t\|^2,$$

which, combined with (25), proves (24). ∎

We next derive an iteration-complexity bound for Algorithm 1 to obtain approximate stationary points of problem (1).

Theorem 2.4 (Iteration-complexity bound for Algorithm 1)**.** *Let $\{x_t\}_{t=1}^T$ be generated by Algorithm 1. Assume that*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} \nu_t = 0, \tag{26}$$

where ν_t is as in (22). Given $\epsilon \in (0,1)$, define $T_0(\epsilon)$ any non-negative integer such that:

$$T \geq T_0(\epsilon) \Longrightarrow \frac{1}{T}\sum_{t=0}^{T-1} \nu_t \leq \frac{\sigma_0 \epsilon^{\frac{3}{2}}}{24\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}. \tag{27}$$

If

$$T \geq \max\left\{ \frac{12(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{\sigma_0}\epsilon^{-\frac{3}{2}}, T_0(\epsilon)\right\}, \tag{28}$$

then

$$\min_{t=0,\ldots,T-1} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| \leq \epsilon. \tag{29}$$

Proof. First of all, it follows from the assumption in (26) that $T_0(\epsilon)$ is well-defined for any given ϵ. Define $t_* := \arg\min_{j\in\{0,\ldots,T-1\}} \|\nabla f_{\mathcal{G}_j}(x_{j+1})\|$. Hence, it follows from (24) that

$$\begin{aligned}
\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*+1})\|^{\frac{3}{2}} &\leq \frac{6(f(x_0) - f^* + 2\sum_{t=0}^{T-1}\nu_t)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0} \\
&= \frac{6(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0} + \frac{12\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0}\sum_{t=0}^{T-1}\nu_t.
\end{aligned}$$

Since (28) holds, we have $T \geq T_0(\epsilon)$, and so it follows from (27) that

$$\frac{12\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0}\sum_{t=0}^{T-1}\nu_t \leq \frac{\epsilon^{\frac{3}{2}}}{2}.$$

On the other hand, also by (28), we have

$$\frac{6(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0} \leq \frac{\epsilon^{\frac{3}{2}}}{2}.$$

By combining the last three inequalities, we obtain

$$\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*+1})\|^{\frac{3}{2}} \leq \epsilon^{\frac{3}{2}},$$

which proves (29). $\square$

Remark 2.5. *Assumption in* (26) *is weaker than the condition of summability of the sequence* $\{\nu_t\}$ *required in the iteration-complexity analyses of the subsampled inexact Newton and subsampled spectral gradient methods in [5, 20]. Indeed, if* $\{\nu_t\}$ *is summable, then*

$$0 \leq \frac{1}{T}\sum_{t=0}^{T} \nu_t \leq \frac{1}{T}\sum_{t=0}^{\infty} \nu_t,$$

which, by taking the limit as T *goes to infinity, implies* (26). *On the other hand, as pointed out in [13], there exist sequences* $\{\nu_t\}$ *satisfying* (26) *that are not summable; an important class of examples is given by the sequences* $\{\nu_t\}$ *such that* $\nu_t \to 0$ *as* $k \to \infty$ *(see [13, Corollary 2] or the proof of Theorem 2.7(a)).*

Under the assumption that $\{\nu_t\}$ is summable, Theorem 2.3 implies, in particular, an iteration-complexity bound of $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ for Algorithm 1 to generate an approximate stationary points of problem (1).

Corollary 2.6. *Let* $\{x_t\}_{t=1}^{T}$ *be generated by Algorithm 1. Assume that* $\sum_{t=0}^{\infty} \nu_t < +\infty$, *where* ν_t *is as in* (22). *Given* $\epsilon \in (0,1)$, *if*

$$T \geq \frac{12}{\sigma_0}\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}} \max\left\{f(x_0) - f^*, 2\sum_{t=0}^{\infty} \nu_t\right\} \epsilon^{-\frac{3}{2}}, \tag{30}$$

then

$$\min_{t=0,\ldots,T-1} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| \leq \epsilon. \tag{31}$$

As a consequence, Algorithm 1 needs at most $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ *calls of the oracle to generate an iterate* x_t *such that* $\|\nabla f_{\mathcal{G}_t}(x_t)\| \leq \epsilon$.

Proof. As discussed in Remark 2.5, if $\{\nu_t\}$ is summable, then (26) holds. Moreover, by defining

$$T_0(\epsilon) := \frac{24}{\sigma_0}\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}} \epsilon^{-\frac{3}{2}} \sum_{t=0}^{\infty} \nu_t < \infty,$$

we obtain

$$T \geq T_0(\epsilon) \Longrightarrow \frac{1}{T}\sum_{t=0}^{T-1} \nu_t \leq \frac{1}{T}\sum_{t=0}^{\infty} \nu_t \leq \frac{\sigma_0 \epsilon^{\frac{3}{2}}}{24\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}},$$

that is, (27) holds. Thus, (30) can be rewritten as

$$T \geq \max\left\{\frac{12(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{\sigma_0}\epsilon^{-\frac{3}{2}}, T_0(\epsilon)\right\},$$

and hence (31) follows from Theorem 2.3.

The second part of the theorem follows trivially from the first one and (19). $\square$

In the case that only exact functions, gradients and Hessians evaluations are considered, i.e., $\mathcal{G}_t = \mathcal{H}_t = \{1,\ldots,d\}$ for all $t \geq 0$, an iteration-complexity bound of $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ for the cubic regularization method is obtained from Corollary 2.6.

We claim that, by following the arguments in [13, Corollary 2 and Remark 4], an iteration-complexity bound of $\mathcal{O}\left(\epsilon^{-\frac{3}{2}}\right)$ similar to the one in (30) for Algorithm 1 can be proven under the assumption that $\nu_t \leq \epsilon/t$, for all $t \geq 0$, instead of summability of $\{\nu_t\}$.

We next establish, as a by-product from the previous complexity estimates, the global convergence of Algorithm 1.

Theorem 2.7 (Global convergence of Algorithm 1)**.** *Let $\{x_t\}_{t=1}^T$ be generated by Algorithm 1. The following statements hold:*

(a) if $\nu_t \to 0$ as $k \to \infty$, then either there exists $t_* \leq T$ such that $\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*+1})\| = 0$ or $\liminf_{t\to\infty} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| = 0$;

(b) if $\sum_{t=0}^{\infty} \nu_t < +\infty$, then $\lim_{t\to\infty} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| = 0$.

Proof. (a) Let $\delta > 0$. Since $\nu_t \to 0$ as $k \to \infty$, it follows that there exist constants C and $t(\delta) > 0$ such that $\nu_t \leq C$ for all t, and $\nu_t \leq \delta$ for all $t \geq t(\delta)$. Hence, for all $T > \max\{2Ct(\delta/2)/\delta, t(\delta/2)+1\}$,

$$\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T} \nu_t &= \frac{1}{T}\sum_{t=0}^{t(\delta/2)-1} \nu_t + \frac{1}{T}\sum_{t=t(\delta/2)}^{T} \nu_t \\
&\leq \frac{1}{T}\sum_{t=0}^{t(\delta/2)-1} C + \frac{1}{T}\sum_{t=t(\delta/2)}^{T} \frac{\delta}{2} \\
&\leq \frac{Ct(\delta/2)}{T} + \frac{\delta}{2} \\
&\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta,
\end{aligned}$$

which implies that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} \nu_t = 0$ and (27) holds for:

$$T_0(\epsilon) := \max\{2Ct(\delta/2)/\delta, t(\delta/2)+1\},$$

with $\delta := \sigma_0 \epsilon^{\frac{3}{2}}/(12\,(3\sigma_{max}/2 + \sigma_0 + \theta)^{\frac{3}{2}})$. Thus, it follows from Theorem 2.3 that, if

$$T \geq \max\left\{\frac{2(f(x_0) - f^*)}{\delta}, \frac{2}{\delta}Ct\left(\frac{\delta}{2}\right), t\left(\frac{\delta}{2}\right) + 1\right\},$$

then

$$\min_{t=0,\ldots,T-1} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| \leq \epsilon.$$

As $\epsilon > 0$ is arbitrary, this proves that:

$$\lim_{T\to\infty}\left(\min_{t=0,\ldots,T-1} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\|\right) = 0.$$

Therefore, either there exists $t_* \leq T$ such that $\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*+1})\| = 0$ or $\liminf_{t\to\infty} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| = 0$.

(b) By taking the limit in (24) as T goes to infinity, we obtain

$$\sum_{t=0}^{\infty} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\|^{\frac{3}{2}} < \infty$$

which in turn implies $\lim_{t\to\infty} \|\nabla f_{\mathcal{G}_t}(x_{t+1})\| = 0$.

□

Remark 2.8. *Using the fact that*

$$0 \leq \|\nabla f(x_{t+1})\| \leq \|\nabla f(x_{t+1}) - \nabla f_{\mathcal{G}_t}(x_{t+1})\| + \|\nabla f_{\mathcal{G}_t}(x_{t+1})\|,$$

and under the additional assumption that $\lim_{t\to+\infty} \|\nabla f_{\mathcal{G}_t}(x_{t+1}) - \nabla f(x_{t+1})\| = 0$, *it follows from Theorem 2.3(b) that* $\lim_{t\to\infty} \|\nabla f(x_t)\| = 0$.

2.1 Second-order iteration-complexity and global convergence analyses

In this section, we present second-order iteration-complexity bounds and global convergence results for Algorithm 1. For this, it is required that $x_{t,i}^+$ satisfies, besides (5), the following condition

$$\nabla^2 f_{\mathcal{H}_{t,i}}(x_t) + \frac{2^i \sigma_t}{2}\|x_{t,i}^+ - x_t\| I \succeq -\theta \|x_{t,i}^+ - x_t\| I, \tag{32}$$

and, regarding the acceptance criteria in Step 2.2, it is required (besides (6) and (7)) the following inequality

$$-\lambda_{min}\left(\nabla^2 f_{\mathcal{G}_t}(x_{t,i}^+)\right) \leq \frac{\sigma + 2(\theta + \sigma_0 + L)}{2}\|x_{t,i}^+ - x_t\|. \tag{33}$$

Note that (32), together with (5), means that $x_{t,i}^+$ approximately satisfies the first- and second-order optimality conditions for a local minimizer of the cubic model $M_{x_t, 2^i\sigma_t}^{\mathcal{G}_t, \mathcal{H}_{t,i}}(\cdot)$. The well-definedness of this new variant of Algorithm 1 can be proven using the second part of Lemma 2.1 and similar arguments as those in the proof of Lemma 2.2.

Theorem 2.9. *Let* $\{x_t\}_{t=1}^T$ *be generated by Algorithm 1 with (5)–*(7), (32) *and* (33) *being satisfied for all* t *and* $i \leq i_t$. *Assume that*

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} \nu_t = 0, \tag{34}$$

where ν_t *is as in* (22). *Given* $\epsilon_g, \epsilon_H \in (0,1)$, *define* $\hat{T}_0(\epsilon_g, \epsilon_H)$ *any non-negative integer such that:*

$$T \geq \hat{T}_0(\epsilon_g, \epsilon_H) \Longrightarrow \frac{1}{T}\sum_{t=0}^{T-1} \nu_t \leq \min\left\{ \frac{\sigma_0 \epsilon_g^{\frac{3}{2}}}{24\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}, \frac{\sigma_0 \epsilon_H^3}{3[\sigma + 2(\theta + \sigma_0 + L)]^3} \right\}. \tag{35}$$

If

$$T \geq \max\left\{ \frac{12(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{\sigma_0 \epsilon_g^{\frac{3}{2}}}, \hat{T}_0(\epsilon_g, \epsilon_H), \frac{3(f(x_0) - f^*)(\sigma + 2(\theta + \sigma_0 + L))^3}{2\sigma_0 \epsilon_H^3} \right\}, \tag{36}$$

then there exists $t_* \leq T$ *such that*

$$\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*})\| < \epsilon_g \quad or \quad \lambda_{min}\left(\nabla^2 f_{\mathcal{H}_{t_*}}(x_{t_*})\right) > -\epsilon_H. \tag{37}$$

Proof. First of all, it follows from the assumption in (34) that $T_0(\epsilon_g, \epsilon_H)$ is well-defined for any given $\epsilon_g, \epsilon_H > 0$. As in the proof of Lemma 2.3 (see (25)), we have

$$T\|s_{t_*}\|^3 \le \sum_{t=0}^{T-1} \|x_{t+1} - x_t\|^3 \le \frac{6(f(x_0) - f^* + 2\sum_{t=0}^{T-1} \nu_t)}{\sigma_0}.$$

where $s_t = x_{t+1} - x_t$ and $t_* = \operatorname{argmin}_{j\in\{0,\dots,T-1\}} \|s_j\|^3$. Hence,

$$\|s_{t_*}\|^3 \le \frac{6(f(x_0) - f^*)}{T\sigma_0} + \frac{12\sum_{t=0}^{T-1} \nu_t}{T\sigma_0}. \tag{38}$$

On the other hand, it follows from (7), the fact that $\sigma_{t+1} = 2^{i_t-1}\sigma_t$ and (18) that

$$\|\nabla f_{\mathcal{G}_t}(x_{t_*+1})\| \le \left(\frac{3\sigma_{t_*+1}}{2} + \sigma_0 + \theta\right) \|x_{t_*+1} - x_{t_*}\|^2 \le \left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right) \|s_{t_*}\|^2,$$

which, combined with (38), yields

$$\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*+1})\|^{\frac{3}{2}} \le \frac{6(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0} + \frac{12\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{T\sigma_0} \sum_{t=0}^{T-1} \nu_t. \tag{39}$$

Now, it follows from (33) and (38) that

$$\begin{aligned}
-\lambda_{min}\left(\nabla^2 f_{\mathcal{H}_{t_*}}(x_{t_*+1})\right) &\le \frac{\sigma + 2(\theta + \sigma_0 + L)}{2}\|s_{t_*}\| \\
&\le \frac{\sigma + 2(\theta + \sigma_0 + L)}{2}\left[\frac{6(f(x_0) - f^*)}{T\sigma_0} + \frac{12\sum_{t=0}^{T-1} \nu_t}{T\sigma_0}\right]^{1/3} \\
&\le \left[\frac{3(\sigma + 2(\theta + \sigma_0 + L))^3(f(x_0) - f^*)}{4T\sigma_0} + \frac{3(\sigma + 2(\theta + \sigma_0 + L))^3\sum_{t=0}^{T-1} \nu_t}{2T\sigma_0}\right]^{1/3}.
\end{aligned}$$

Therefore, (37) now follows from the last inequality, (39) and (36). □

Under the assumption that $\{\nu_t\}$ is summable, Theorem 2.9 implies, in particular, an iteration-complexity bound of $\mathcal{O}\left(\max\{\epsilon_g^{-\frac{3}{2}}, \epsilon_g^{-3}\}\right)$ for Algorithm 1 to generate an approximate second-order stationary points of problem (1).

Corollary 2.10. *Let $\{x_t\}_{t=1}^T$ be generated by Algorithm 1 with (5)–(7), (32) and (33) being satisfied for all t and $i \le i_t$. Assume that $\sum_{t=0}^{\infty} \nu_t < \infty$, where ν_t is as in (22). Given $\epsilon_g, \epsilon_H \in (0, 1)$, if*

$$T \ge \frac{3}{\sigma_0} \max\left\{\tau_1(f(x_0) - f^*), 2\max\{\tau_1, \tau_2\}\sum_{t=0}^{\infty} \nu_t, \tau_2(f(x_0) - f^*)\right\} \max\{\epsilon_g^{-\frac{3}{2}}, \epsilon_g^{-3}\}, \tag{40}$$

where $\tau_1 := 4\left(3\sigma_{max}/2 + \sigma_0 + \theta\right)^{\frac{3}{2}}$ and $\tau_2 := [\sigma + 2(\theta + \sigma_0 + L)]^3/2$ then there exists $t_* \le T$ such that

$$\|\nabla f_{\mathcal{G}_{t_*}}(x_{t_*})\| < \epsilon_g \quad \text{and} \quad \lambda_{min}\left(\nabla^2 f_{\mathcal{H}_{t_*}}(x_{t_*})\right) > -\epsilon_H. \tag{41}$$

As a consequence, Algorithm 1 needs at most $\mathcal{O}\left(\max\{\epsilon_g^{-\frac{3}{2}}, \epsilon_g^{-3}\}\right)$ *calls of the oracle to generate an iterate* x_{t_*} *satisfying* (41)*. Moreover, either there exists* $\hat{t}$ *such that* $\nabla_{\mathcal{G}_{\hat{t}}} f(x_{\hat{t}}) = 0$ *and* $\nabla^2 f_{\mathcal{H}_{\hat{t}}}(x_{\hat{t}}) \succeq 0$ *or*

$$\lim_{t\to+\infty} \max\left\{\|\nabla f_{\mathcal{G}_t}(x_t)\|, -\lambda_{min}\left(\nabla^2 f_{\mathcal{H}_t}(x_t)\right)\right\} = 0. \tag{42}$$

Proof. As discussed in Remark 2.5, if $\{\nu_t\}$ is summable, then (34) holds. Moreover, by defining

$$\hat{T}_0(\epsilon_g, \epsilon_H) := \max\left\{\frac{24\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{\sigma_0 \epsilon_g^{\frac{3}{2}}}, \frac{3[\sigma + 2(\theta + \sigma_0 + L)]^3}{\sigma_0 \epsilon_H^3}\right\} \sum_{t=0}^{\infty} \nu_t < \infty,$$

we find, using the fact that $\min\{a, b\}^{-1} = \max\{1/a, 1/b\}$ for all $a, b > 0$, that

$$T \geq \hat{T}_0(\epsilon_g, \epsilon_H) \Longrightarrow \frac{1}{T}\sum_{t=0}^{T-1} \nu_t \leq \frac{1}{T}\sum_{t=0}^{\infty} \nu_t \leq \min\left\{\frac{\sigma_0 \epsilon_g^{\frac{3}{2}}}{24\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}, \frac{\sigma_0 \epsilon_H^3}{3[\sigma + 2(\theta + \sigma_0 + L)]^3}\right\},$$

that is, (35) holds. Thus, (40) implies that

$$T \geq \max\left\{\frac{12(f(x_0) - f^*)\left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{\frac{3}{2}}}{\sigma_0 \epsilon_g^{\frac{3}{2}}}, \hat{T}_0(\epsilon_g, \epsilon_H), \frac{3(f(x_0) - f^*)(\sigma + 2(\theta + \sigma_0 + L))^3}{2\sigma_0 \epsilon_H^3}\right\},$$

and hence (41) follows from Theorem 2.9.

The second part of the theorem follows trivially from the first one and (19).

Now, by (33) and (25), we have

$$\begin{aligned}
-\sum_{t=0}^{T}[\lambda_{min}\left(\nabla^2 f_{\mathcal{G}_t}(x_{t+1})\right)]^3 &\leq \frac{[\sigma + 2(\theta + \sigma_0 + L)]^3}{8}\sum_{t=0}^{T}\|x_{t+1} - x_t\|^3 \\
&\leq \frac{3[\sigma + 2(\theta + \sigma_0 + L)]^3(f(x_0) - f^* + 2\sum_{t=0}^{T-1}\nu_t)}{4\sigma_0}.
\end{aligned}$$

Therefore, (42) now follows by taking the limit in the last inequality and (24) as T goes to infinity. □

Remark 2.11. *It follows from (42) that the addition of requirement (32) in Step 2.1 of Algorithm 1 allows the iterates to escape from nondegenerate saddle points.*

2.2 Local convergence analysis

This section is dedicated to establish some local converge properties of Algorithm 1. Forward this goal, we assume that full precision for the function and gradient has been reach at some iteration $\bar{t}$ and that the algorithm continues with $|\mathcal{G}_t| = d$ for all $t \geq \bar{t}$. Moreover, the following additional hypothesis is assumed:

(A3) There exists $\mu > 0$ such that $\nabla^2 f(x) \succeq \mu I$ whenever

$$f(x) \leq f(x_{\bar{t}}).$$

By (23) and the fact that $\nu_t = 0$ for all $t \geq \bar{t}$, we have

$$f(x_t) \leq f(x_{t-1}) \leq \ldots \leq f(x_{\bar{t}+1}) \leq f(x_{\bar{t}}), \quad \forall t \geq \bar{t},$$

which, combined with **A3**, yields

$$\nabla^2 f(x_t) \succeq \mu I, \quad \forall t \geq \bar{t}.$$

From this remark, we are able to show the local quadratic convergence rate for Algorithm 1.

Theorem 2.12. *Let $\{x_t\}_{t\geq 0}$ be generated by Algorithm 1. Given $\alpha \in [0,1)$, if*

$$\|\nabla^2 f(x_t) - \nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t)\|\| \leq \alpha\mu, \tag{43}$$

and

$$\|\nabla f(x_0)\| \leq \left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{-1} \frac{(1-\alpha)^3\mu^2}{(1+\theta)^2}, \tag{44}$$

then

$$\|\nabla f(x_t)\| \leq \left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{-1} \frac{\mu^2}{(\theta+1)^2}(1-\alpha)^{2^{t+2}}, \quad \forall t \geq 1. \tag{45}$$

Proof. First, we will show that

$$\|\nabla f(x_{t+1})\| \leq \left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right) \frac{(1+\theta)^2}{(1-\alpha)^2\mu^2}\|\nabla f(x_t)\|^2 \tag{46}$$

for all $t \geq 0$. Assume that

$$\|\nabla f(x_t)\| \leq \left(\frac{3\sigma_{max}}{2} + \sigma_0 + \theta\right)^{-1} \frac{(1-\alpha)^3\mu^2}{(1+\theta)^2}. \tag{47}$$

for some $t \geq 0$. From (43), we obtain, for any $v \neq 0$, that

$$v^T(\nabla^2 f(x_t) - \nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t))v \leq \|\nabla^2 f(x_t) - \nabla^2 f_{\mathcal{H}_{t,i_t}}\|\|v\|^2 \leq \alpha\mu\|v\|^2 = v^T(\alpha\mu I)v,$$

which implies

$$\nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t)) \succeq \nabla^2 f(x_t) - \alpha\mu I.$$

Hence, by Weyl's inequality, we find

$$\lambda_{min}\left(\nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t))\right) \geq \lambda_{min}\left(\nabla^2 f(x_t)\right) - \alpha\mu \geq (1-\alpha)\mu > 0. \tag{48}$$

On the other hand, it follows from the second inequality in (5) that

$$\|M^{\mathcal{H}_{t,i_t}}_{x_t,2^{i_t}\sigma_t}(x_{t+1})\| \leq \theta\|\nabla f(x_t)\|, \tag{49}$$

where

$$\nabla M^{\mathcal{H}_{t,i_t}}_{x_t,2^{i_t}\sigma_t}(x_{t+1}) = \nabla f(x_t) + \nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t)\left(x_{t+1} - x_t\right) + \frac{2^{i_t}\sigma_t}{2}\|x_{t+1} - x_t\|\left(x_{t+1} - x_t\right).$$

From the last equality, we get

$$\left(\nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t) + \frac{2^{i_t}\sigma_t}{2}\|x_{t+1} - x_t\| I\right)(x_{t+1} - x_t) = \nabla M^{\mathcal{H}_{t,i_t}}_{x_t,2^{i_t}\sigma_t}(x_{t+1}) - \nabla f(x_t)$$

$$\Longrightarrow x_{t+1}-x_t=-\left(\nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t)+\frac{2^{i_t}\sigma_t}{2}\|x_{t+1}-x_t\|I\right)^{-1}\left(\nabla M^{\mathcal{H}_{t,i_t}}_{x_t,2^{i_t}\sigma_t}(x_{t+1})-\nabla f(x_t)\right).$$

Then, by (48) and (49), we have

$$\begin{aligned}
\|x_{t+1}-x_t\| &= \left\|\left(\nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t)+\frac{2^{i_t}\sigma_t}{2}\|x_{t+1}-x_t\|I\right)^{-1}\left(\nabla M^{\mathcal{H}_{t,i_t}}_{x_t,2^{i_t}\sigma_t}(x_{t+1})-\nabla f(x_t)\right)\right\| \\
&\leq \frac{\|\nabla M^{\mathcal{H}_{t,i_t}}_{x_t,2^{i_t}\sigma_t}(x_{t+1})-\nabla f(x_t)\|}{\lambda_{min}\left(\nabla^2 f_{\mathcal{H}_{t,i_t}}(x_t)\right)} \\
&\leq \frac{(1+\theta)\|\nabla f(x_t)\|}{(1-\alpha)\mu},
\end{aligned}$$

which, combined with (7) and (18), yields

$$\begin{aligned}
\|\nabla f(x_{t+1})\| &\leq \left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)\|x_{t+1}-x_t\|^2 \\
&\leq \left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)\frac{(1+\theta)^2}{(1-\alpha)^2\mu^2}\|\nabla f(x_t)\|^2.
\end{aligned}$$

Consequently, by (47) and the fact that $\alpha\in[0,1)$, we also have

$$\begin{aligned}
\|\nabla f(x_{t+1})\| &\leq \left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)\frac{(1+\theta)^2}{(1-\alpha)^2\mu^2}\left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)^{-1}\frac{(1-\alpha)^3\mu^2}{(1+\theta)^2}\|\nabla f(x_t)\| \\
&= (1-\alpha)\|\nabla f(x_t)\|<\left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)^{-1}\frac{(1-\alpha)^3\mu^2}{(1+\theta)^2}
\end{aligned}$$

Thus, by induction, (46) holds for all $t\geq 0$.

Denoting

$$\delta_t=\left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)\frac{(1+\theta)^2}{(1-\alpha)^2\mu^2}\|\nabla f(x_t)\|,$$

it follows from (46) that

$$\delta_{t+1}\leq\delta_t^2\quad\forall t\geq 0.$$

Moreover, by (44), we also have

$$\delta_0=\left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)\frac{(1+\theta)^2}{(1-\alpha)^2\mu^2}\|\nabla f(x_0)\|\leq 1-\alpha.$$

Therefore, for all $t\geq 1$,

$$\begin{aligned}
\|\nabla f(x_t)\|=\left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)^{-1}\frac{(1-\alpha)^2\mu^2}{(1+\theta)^2}\delta_t &\leq \left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)^{-1}\frac{(1-\alpha)^2\mu^2}{(1+\theta)^2}\delta_0^{2^t} \\
&\leq \left(\frac{3\sigma_{max}}{2}+\sigma_0+\theta\right)^{-1}\frac{(1-\alpha)^2\mu^2}{(1+\theta)^2}(1-\alpha)^{2^t},
\end{aligned}$$

which implies (45). □

Table 1: Datasets: problem dimension n and the number of training samples d

Name	n	d
Cina0	132	16033
Gisette	5000	6000
Madelon	500	2000
Secom	590	1567

3 Numerical Experiments

In this section, we explore the numerical behavior of Algorithm 1 (subsampled-CRM) to solve the l_2-logistic problem. The computational results were obtained using MATLAB R2018a on a machine with a 3.5 GHz Dual-Core Intel Core i5 processor and 16 GB 2400 MHz DDR4 memory. The proposed algorithm was compared with two other schemes: the full-CRM, which corresponds to Algorithm 1 with $\mathcal{G}_t = \mathcal{H}_t = \{1, \ldots, d\}$ for all $t \geq 0$, and the subsampled Adaptive Cubic Regularization method (subsampled-ARC) of [3] where, at each iteration, the Hessian is approximated by $\nabla^2 f_{\mathcal{H}_k}(x) := 1/|\mathcal{H}_k| \sum_{i \in \mathcal{H}_k} \nabla^2 f_i(x)$ with

$$|\mathcal{H}_k| = \max\left\{0.005d, \min\left\{0.1d, \left\lceil \frac{4\rho}{C_k}\left(\frac{2\rho}{C_k} + \frac{1}{3}\right) \log\left(\frac{2n}{0.2}\right)\right\rceil\right\}\right\},$$

and ρ and C_k are updated as described in [3]. The initialization parameters in the subsampled-ARC method were fixed as suggested in [3, Section 8.1], while, in the subsampled-CRM, they were set as $\theta = 5$, $\sigma_0 = 0.1$, $\mathcal{G}_0 = 0.1d$ and $\mathcal{H}_0 = \mathcal{H}_t = \mathcal{G}_0$. Moreover, we update the sample $\mathcal{G}_t$ such that $|\mathcal{G}_t| = \min\{d, \lceil 1.25^t |\mathcal{G}_0| \rceil\}$ for all $t \geq 0$. For all methods, each cubic subproblem was approximately solved by the Barzilai–Borwein gradient method [1] combined with the nonmonotone linesearch of [16]. The major per iteration cost of the last method is one Hessian-vector product, needed to compute the gradient of the cubic model.

We use as test problem the l_2-logistic problem of the form

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{d} \sum_{i=1}^{d} f_i(x) = \left[\log(1 + e^{-b_i a_i^T x}) + \frac{\mu}{2}\|x\|_2^2\right], \tag{50}$$

where $\{(a_i, b_i)\}_{i=1}^d \subset \mathbb{R}^n \times \{-1, 1\}$ is the dataset and $\mu > 0$ is the regularization parameter. We set $\mu = 0.001$. The used data sets are from [12, 21] and each one of them is described in Table 1. For all instances, we used $x_0 = (0, \ldots, 0)^T$ as starting point.

Following [23], we consider, as the performance measurement, the total number of propagations, which, in this instance, corresponds to the number of oracle calls of function and Hessian-vector product. Note that, due to the particular structure of f_i in (50) and its gradient and Hessian, once that f_i, in particular $e^{-b_i a_i^T x}$, has been computed the evaluation of ∇f_i and $\nabla^2 f_i$ comes for free. Moreover, the evaluation of Hessian-vector product, which is required for each iteration of Barzilai–Borwein gradient method, corresponds to two gradient evaluations. Table 2 describes, taking into account the possibility of partial evaluations of function and Hessian, the total number of propagations per iterations for each algorithm considered in this section.

Table 2: Total number of propagations per iterations for each algorithm, where j denotes the number of Barzilai–Borwein iterations for solving the corresponding cubic subproblems.

subsampled-CRM	full-CRM	subsampled-ARC
$(\lvert\mathcal{G}_t\rvert + 2\lvert\mathcal{H}_t\rvert j)/d$	$1 + 2j$	$1 + 2\lvert\mathcal{H}_t\rvert j/d$

Figure 1 shows a comparison among the three algorithms in terms of the function values versus the total number of propagations. As can be seen, the subsampled-CRM clearly outperforms the full-CRM in all datasets, demonstrating the advantages of using subsampling techniques. Moreover, the subsampled-CRM is more efficient than subsampled-ARC mainly due to the fact that it also uses subsampled approach to evaluate the function f in (50) as well as its gradient.

4 Final remarks

We proposed a subsampled cubic regularization method for solving finite-sum optimization problems. Under suitable hypotheses, we proved that the proposed algorithm needs at most $\mathcal{O}\left(\epsilon^{-3/2}\right)$ (resp. $\mathcal{O}(\max\{\epsilon_g^{-\frac{3}{2}}, \epsilon_g^{-3}\})$) calls of the oracle to generate an ϵ-approximate first-order (resp. an (ϵ_g, ϵ_H)-approximate second-order) stationary point of the objective function. Global convergence properties for finding approximate first- and second-order stationary points were also established. We further proved a quadratic convergence result for the method. Finally, numerical experiments were presented, illustrating the advantages of using subsampling techniques.

References

[1] J. Barzilai and J. M. Borwein, *Two-Point Step Size Gradient Methods*, IMA J. Numer. Anal. **8** (1988), no. 1, 141–148.

[2] S. Bellavia and Gurioli. G, *Stochastic analysis of an adaptive cubic regularization method under inexact gradient evaluations and dynamic hessian accuracy*, Optimization **71** (2022), no. 1, 227–261.

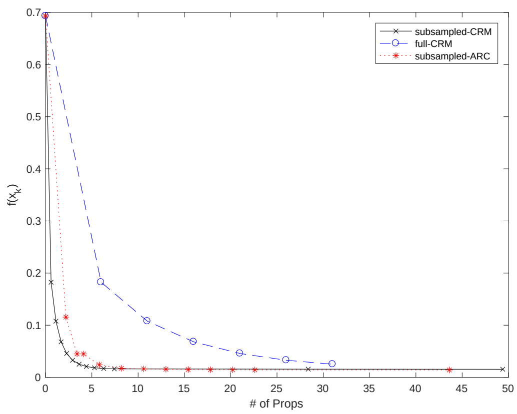
[3] S. Bellavia, G. Gurioli, and B. Morini, *Adaptive cubic regularization methods with dynamic inexact Hessian information and applications to finite-sum minimization*, IMA J. Numer. Anal. **41** (2020), no. 1, 764–799.

[4] S. Bellavia, G. Gurioli, B. Morini, and P. L. Toint, *Quadratic and cubic regularisation methods with inexact function and random derivatives for finite-sum minimisation*, Available on https://doi.org/10.48550/arXiv.2104.00592.

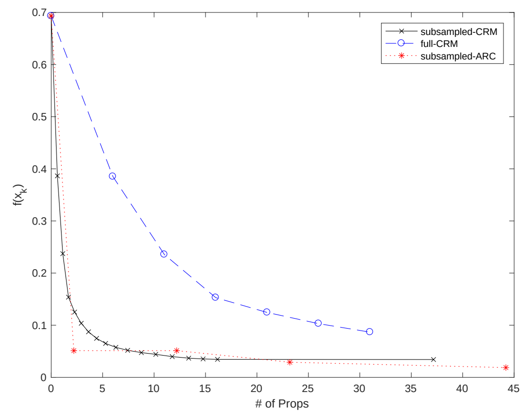
[5] S. Bellavia, N. Krejić, and N. K. Jerinkić, *Subsampled inexact Newton methods for minimizing large sums of convex functions*, IMA J. Numer. Anal. **40** (2019), no. 4, 2309–2341.

[6] S. Bellavia, N. Krejić, and B. Morini, *Inexact restoration with subsampled trust-region methods for finite-sum minimizations*, Comput. Optim. Appl. **76** (2020), no. 2, 701–736.

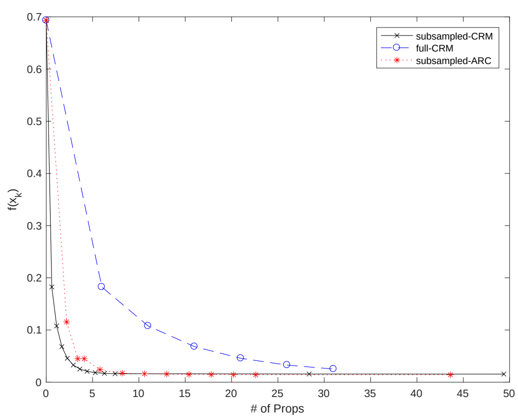
Cina0



Gisette



Madelon



Secom

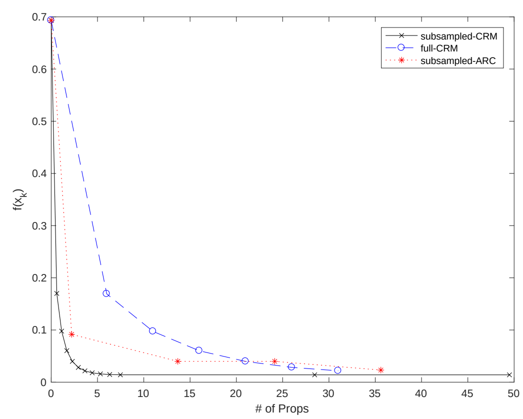


Figure 1: Comparison of subsampled-CRM, full-CRM and subsampled-ARC in terms of the function value versus the total number of propagations.

[7] E. G. Birgin, J. L. Gardenghi, J. M. Martnez, S. A. Santos, and Ph. L. Toint, *Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models*, Math. Program. **163** (2017), no. 1, 359–368.

[8] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *On the complexity of steepest descent, Newton's and regularized newton's methods for nonconvex unconstrained optimization problems*, SIAM J. Optim. **20** (2010), no. 6, 2833–2852.

[9] X. Chen, B. Jiang, T. Lin, and S. Zhang, *Accelerating adaptive cubic regularization of Newton's method via random sampling*, J. Mach. Learn. Res. **23** (2022), no. 90, 1–38.

[10] J. W. Demmel, *Applied numerical linear algebra*, SIAM, Philadelphia, 1997.

[11] D. di Serafino, N. Krejić, N. K. Jerinkić, and M. Viola, *LSOS: Line-search second-order stochastic optimization methods for nonconvex finite sums*, Available on https://doi.org/10.48550/arXiv.2007.15966.

[12] D. Dua and C. Graff, *UCI machine learning repository. http://archive.ics.uci.edu/ml*, 2017.

[13] G. N. Grapiglia and E.W. Sachs, *A generalized worst-case complexity analysis for non-monotone line searches*, Numer. Algorithms **87** (2021), no. 2, 779–796.

[14] G.N. Grapiglia, M.L.N. Gonçalves, and G.N. Silva, *A cubic regularization of Newton's method with finite difference hessian approximations*, Numer. Algorithms **90** (2022), 607–630.

[15] A Griewank, *The modification of Newton method for unconstrained optimization by bounding cubic terms.*, Technical Report NA/12, Department of Applied Mathematics and Theoretical Physics. University of Cambridge, Cambridge (1981).

[16] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal. **23** (1986), no. 4, 707–716.

[17] J. Kohler and A. Lucchi, *Sub-sampled cubic regularization for non-convex optimization*, In Proceedings of the 34th International Conference on Machine Learning (ICML 2017), vol. 70, 2017, pp. 1895 – 1904.

[18] Y. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program. **184** (2006), no. 1, 177–205.

[19] S. Park, S. H. Jung, and P. M. Pardalos, *Combining stochastic adaptive cubic regularization with negative curvature for nonconvex optimization*, J. Optim. Theory Appl. **184** (2020), no. 3, 953–971.

[20] B. Stefania, N. K. Jerinkić, and G. Malaspina, *Subsampled nonmonotone spectral gradient methods*, Commun. Appl. Ind. Math. **11** (2020), no. 1, 19–34.

[21] Causality Workbench Team, *A marketing dataset. Available at http://www.causality.inf.ethz.ch/data/CINA.html*, 2008.

[22] Z. Wang, Y. Zhou, Y. Liang, and G. Lan, *Cubic regularization with momentum for nonconvex optimization*, Available on https://doi.org/10.48550/arXiv.1810.03763.

[23] P. Xu, F. Roosta, and M. W. Mahoney, *Second-order optimization for non-convex machine learning: an empirical study*, pp. 199–207, in Proceedings of the 2020 SIAM International Conference on Data Mining (SDM), SIAM, 2020.