Universidade Federal de Goiás
Instituto de Matemática e Estatística
Departamento de Matemática

Vilmar Gehlen Filho

Cubic Regularization Methods with Lazy Hessian Approximations - Métodos de Regularização Cúbica com Aproximações Preguiçosas da Hessiana

Goiânia-GO
2025

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[X] Dissertação [ ] Tese [ ] Outro*:_____________

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Vilmar Gehlen Filho

3. Título do trabalho

"Cubic Regularization Methods with Lazy Hessian Approximations" - "Métodos de Regularização Cúbica com Aproximações Preguiçosas da Hessiana"

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

sei! assinatura eletrônica

Documento assinado eletronicamente por **Max Leandro Nobre Goncalves**, **Professor do Magistério Superior**, em 14/03/2025, às 08:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **Vilmar Gehlen Filho**, **Discente**, em 14/03/2025, às 08:57, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5230425** e o código CRC **4F9DBE01**.

Referência: Processo nº 23070.004354/2025-61 SEI nº 5230425

Vilmar Gehlen Filho

Cubic Regularization Methods with Lazy Hessian Approximations - Métodos de Regularização Cúbica com Aproximações Preguiçosas da Hessiana

Dissertação apresentada ao Programa de Pós–Graduação em Matemática, do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de concentração: Otimização.

Orientador: Prof. Max Leandro Nobre Gonçalves

Goiânia-GO
2025

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Filho, Vilmar Gehlen
Cubic Regularization Methods with Lazy Hessian Approximations [manuscrito] / Vilmar Gehlen Filho. - 2025.
LXVII, 67 f.: il.

Orientador: Prof. Max Leandro Nobre Gonçalves.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2025.
Bibliografia.

1. Cubic Regularization Method. 2. Complexity Analysis. 3. Lazy Hessian. 4. Non-Convex Optimization. I. Gonçalves, Max Leandro Nobre, orient. II. Título.

CDU 51

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **11** da sessão de Defesa de Dissertação de **Vilmar Gehlen Filho**, que confere o título de Mestre em **Matemática**, na área de concentração em **Otimização**.

Ao **25/02/2025 (vigésimo quinto dia do mês de fevereiro do ano de dois mil e vinte e cinco)**, a partir das **14:00**, via Web videoconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada **"Cubic Regularization Methods with Lazy Hessian Approximations" - "Métodos de Regularização Cúbica com Aproximações Preguiçosas da Hessiana"**. Os trabalhos foram instalados pelo Orientador, Professor Doutor **Max Leandro Nobre Gonçalves - IME/UFG** com a participação dos demais membros da Banca Examinadora: Professor Doutor **Jefferson Divino Gonçalves de Melo - IME/UFG**, membro titular interno; Professor Doutor **Geovani Nunes Grapiglia - ICTEAM/UCLouvain**, membro titular externo e Professor Doutor **Luiz Rafael dos Santos - MAT/UFSC** membro titular externo. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Max Leandro Nobre Gonçalves**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao **vigésimo quinto dia do mês de fevereiro do ano de dois mil e vinte e cinco**.

TÍTULO SUGERIDO PELA BANCA

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **LUIZ RAFAEL DOS SANTOS**, **Usuário Externo**, em 14/03/2025, às 01:07, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **Jefferson Divino Goncalves De Melo**, **Professor do Magistério Superior**, em 14/03/2025, às 08:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **Max Leandro Nobre Goncalves**, **Professor do Magistério Superior**, em 14/03/2025, às 08:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

---

sei! assinatura eletrônica

Documento assinado eletronicamente por **Rodrigo Donizete Euzebio**, **Vice-Coordenador**, em 14/03/2025, às 08:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Referência: Processo nº 23070.004354/2025-61 SEI nº 5230417

Vilmar Gehlen Filho

Cubic Regularization Methods with Lazy Hessian Approximations - Métodos de Regularização Cúbica com Aproximações Preguiçosas da Hessiana

Dissertação defendida no Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Matemática, aprovada em 25 de Fevereiro de 2025, pela Banca Examinadora constituída pelos professores:

Prof. Max Leandro Nobre Gonçalves
Instituto de Matemática e Estatística – UFG
Presidente da Banca

Prof. Jefferson Divino Gonçalves de Melo
Institudo de Matemática e Estatística – UFG

Prof. Luiz Rafael dos Santos
Departamento de Matemática – UFSC

Prof. Geovani Nunes Grapiglia
Departamento de Matemática – Université Catholique de Louvain – UCLouvain

Vilmar Gehlen Filho

Graduou-se em Licenciatura em Matemática (2023) pela Universidade Federal de Jataí (UFJ). Durante a graduação, desenvolveu trabalhos de iniciação científica no departamento de Matemática na área de Análise.

Dedico este trabalho aos meus pais, Luzia Laine Assis da Silva e Vilmar Gehlen, pois é graças a eles que consegui estar onde estou.

Agradecimentos

Agradeço aos meus pais, Luzia Laine Assis da Silva e Vilmar Gehlen, pelos incentivos e apoio durante toda a minha jornada.

Ao meu orientador de graduação, Professor Claudiney Goulart, pois sem sua enorme ajuda, eu não estaria aqui.

Ao meu orientador Professor Max Leandro Nobre Gonçalves, pela paciência, pelos inúmeros ensinamentos e pelas várias merecidas broncas.

Agradeço à CAPES pelo apoio financeiro.

Resumo

Filho, Vilmar Gehlen. **Cubic Regularization Methods with Lazy Hessian Approximations - Métodos de Regularização Cúbica com Aproximações Preguiçosas da Hessiana**. Goiânia-GO, 2025. 67p. Dissertação de Mestrado. Departamento de Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Neste trabalho, apresentamos implementações de uma variante da Regularização Cúbica de Newton, incorporando aproximações preguiçosas da Hessiana para resolver problemas gerais de otimização não convexa (0-3). Nós propomos dois métodos: o primeiro (Algoritmo 1) utiliza de gradiente exato enquanto reutiliza a mesma aproximação da Hessiana para um bloco de m iterações, por outro lado, o segundo (Algoritmo 2) estende essa ideia permitindo o uso de gradientes inexatos. Implementações de métodos, em que a informação sobre as derivadas são computadas por diferenças finitas são apresentadas. Um recurso interessante empregado pelos algoritmos é que ambos os parâmetros de regularização e a precisão das aproximações das derivadas (quando são atualizadas) são ajustadas usando um critério de busca linear não monótona. Estabelecemos complexidades de primeira ordem para ambos os métodos. Especificamente, dado uma precisão $\epsilon > 0$, é mostrado que o Algoritmo 1 e o Algoritmo 2 requerem no máximo $\mathcal{O}\left(m^{1/2}\epsilon^{-3/2}\right)$ iterações externas para gerar um ponto crítico ϵ-aproximado do problema em questão. Quando as derivadas são computadas com aproximações por diferenças finitas, mostramos que o Algoritmo 1 (resp. Algoritmo 2) precisam no máximo $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2} + (n+m)\right)$ (resp. $\mathcal{O}\left((n^2+mn)m^{-1/2}\epsilon^{-3/2} + (n^2+mn)\right)$) avaliações do gradiente e função (resp. função) para gerar um ponto crítico ϵ-aproximado, onde n é a dimensão do domínio da função objetivo.

Palavras–chave

Método de Regularização Cúbica, Análise de Complexidade, Hessiana Preguiçosa, Otimização Não-Convexa.

Abstract

Filho, Vilmar Gehlen. **Cubic Regularization Methods with Lazy Hessian Approximations**. Goiânia-GO, 2025. 67p. MSc. Dissertation. Departamento de Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

In this work, we present variants of the Cubic Regularization Newton's (CRN) method incorporating lazy Hessian approximations for solving general non-convex optimization problems (0-3). We propose two approaches: the first (Algorithm 1) employs the exact gradient while reusing the same Hessian approximation for a block of m iterations, whereas the second (Algorithm 2) extends this idea by additionally allowing the use of inexact gradients. Implementations of methods, where information about derivatives are computed through finite difference strategies, are presented. One interesting feature of our algorithms is that the regularization parameter and the accuracy of the derivative approximations (when they are updated) are jointly adjusted using a nonmonotone line search criterion. We establish first-order complexity results for both methods. Specifically, for a given precision ϵ, it is shown that the Algorithm 1 and Algorithm 2 require at most $\mathcal{O}\left(m^{1/2}\epsilon^{-3/2}\right)$ outer iterations to generate an $\epsilon-$approximate critical point for aforementioned problem. When the derivatives are computed by finite difference approaches, we show that Algorithm 1 (resp. Algorithm 2) needs at most $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2}+(n+m)\right)$ (resp. $\mathcal{O}\left((n^2+mn)m^{-1/2}\epsilon^{-3/2}+(n^2+mn)\right)$) gradient and function (resp. function) evaluations to generate an ϵ-approximate critical point, where n is the dimension of the domain of the objective function.

Keywords

Cubic Regularization Method, Complexity Analysis, Lazy Hessian, Non-Convex Optimization.

Contents

Introduction

Consider the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{0-1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function, potentially nonconvex. Among the various methods for solving (0-1) (see, e.g., [16]), the Cubic Regularization of Newton's (CRN) method [9, 14] is widely regarded as one of the most effective in terms of iteration-complexity bounds. This method generates a sequence $\{x_t\}$ by solving the subproblem

$$x_{t+1} \in \operatorname*{argmin}_{y \in \mathbb{R}^n} M_{x_t,\sigma_t}(y),$$

where $\sigma_t > 0$ is an appropriately chosen regularization parameter, and

$$M_{x,\sigma}(y) = \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle \nabla^2 f(x)(y - x), y - x\rangle + \frac{\sigma}{6}\|y - x\|^3. \tag{0-2}$$

Assuming that $\nabla^2 f$ is L-Lipschitz continuous in $\mathbb{R}^n$ and choosing σ_t proportional to L, it has been shown that the CRN method requires at most $\mathcal{O}(\epsilon^{-3/2})$ iterations to generate an ϵ-approximate stationary point of (0-1), i.e., a point x_t satisfying $\|\nabla f(x_t)\| \leq \epsilon$, where $\epsilon > 0$ is a given accuracy tolerance. Notably, [3] established that, for the same class of problems, the standard Newton method (without regularization) can require a number of iterations arbitrarily close to $\mathcal{O}(\epsilon^{-2})$ to achieve an ϵ-approximate stationary point.

The iterations of the CRN method are computed by minimizing a cubic model that incorporates both the gradient and the Hessian of the objective function f. However, in some applications, evaluating these terms can be computationally expensive. To mitigate this issue, several studies have proposed variants of the CRN method in which the gradient and/or Hessian are computed inexactly; see, for instance, [4, 8, 12, 17].

The recent work [8] introduced a CRN variant where the Hessian in the model (0-2) is approximated using finite differences of gradient evaluations. This approach employs a nonmonotone line search to estimate the Lipschitz constant of the Hessian and requires only an approximate solution of the CRN subproblem. Under the assumption that $\nabla^2 f$ is L-Lipschitz continuous in $\mathbb{R}^n$, it was shown in [8] that the proposed method achieves an iteration complexity of at most $\mathcal{O}(\epsilon^{-3/2})$ iterations to produce an ϵ-approximate stationary point of (0-1). In terms of function and gradient evaluations, the complexity bound is $\mathcal{O}(n\epsilon^{-3/2})$. Numerical experiments were also conducted, demonstrating the computational advantages of this new approach.

In the setting of CRN methods with lazy approximations, some variants have been recently proposed in [5, 6]. As far as we know, paper [5] was the first one to propose a CRN method with lazy Hessian approximations (i.e., the same Hessian approximation is reused during m consecutive iterations). The proposed method with lazy Hessians for solving (0-1) has the iteration complexity bound of $\mathcal{O}(\sqrt{m}\epsilon^{-3/2})$ for nonconvex problems. Moreover, when $m = n$, it requires in the worst-case a number of Hessian evaluations smaller by a factor of CRN. Preliminary numerical experiments were also discussed. On the other hand, work [6] proposes zeroth- and first-order implementations of CRN method with lazy updates for the Hessians in the more general problem

$$\min_{x \in Q} F(x) := f(x) + \psi(x), \tag{0-3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is as in Problem (0-1), $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is *simple*, proper, closed and convex, but possibly nondifferentiable, and $Q = \text{dom}(\psi)$. The term "simple" above means that the corresponding CRN subproblem that involve ψ can be efficiently solved.

It is known that (0-3) contains a wide class of problems arising in applications from science and engineering, including machine learning, compressed sensing, and image processing. There are important examples of this problem such as using ℓ_1-regularization to obtain sparse solutions with applications in signal recovery and signal processing problems [1], the nearest correlation matrix problem [2], and regularized inverse problems with atomic norms [18]. Paper [6] also use a special adaptive search procedure in the algorithms, which simultaneously fits both the regularization constant and the parameters of the finite difference approximations, that makes the schemes free from knowing the actual Lipschitz constants. It is shown that the first-order (Hessian-free) and zeroth-order (gradient-free) methods of [6] need, respectively, at most $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2} + (n+m)\right)$ function and gradient evaluations and $\mathcal{O}\left((mn+n^2)m^{-1/2}\epsilon^{-3/2} + (mn+n^2)\right)$ function evaluations to find an ϵ-approximate stationary point of (0-3). Some preliminary numerical experiments that illustrate the practical efficiency of the proposed methods were presented.

In this work, we also explore variants of the CRN method incorporating lazy Hessian approximations for solving (0-3). We propose two new approaches: the first (Algorithm 1) employs the exact gradient while reusing the same Hessian approximation for a block of m iterations. The second (Algorithm 2) extends this idea by additionally allowing the use of inexact gradients. It is worth pointing out that the inexact conditions for the derivative approximations in our methods are implementable because they depend only on the current variables; for further discussion on this issue, see, for example, [8, 10, 19]. Implementations of methods, where information about derivatives are computed through finite difference strategies, are presented. One interesting feature of our algorithms is that the regularization parameter σ and the accuracy of the derivative approximations (when they are updated) are jointly adjusted using a nonmonotone line search criterion.

We establish first-order complexity results for both methods. Specifically, it is shown that the Algorithm 1 and Algorithm 2 require at most $\mathcal{O}\left(m^{1/2}\epsilon^{-3/2}\right)$ outer iterations to generate an $\epsilon-$approximate critical point for problem (0-3). When the derivatives are computed by finite difference approaches, we show that Algorithm 1 (resp. Algorithm 2) needs $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2}+(n+m)\right)$ (resp. $\mathcal{O}\left((n^2+mn)m^{-1/2}\epsilon^{-3/2}+(n^2+mn)\right)$) gradient and function (resp. function) evaluations to generate an ϵ-approximate critical point of (0-3). At the same time, when we take $m=n$, it gives us the better complexity in terms of functions and gradient (resp. function) evaluations of order $\mathcal{O}\left(n^{1/2}\epsilon^{-3/2}+n\right)$ (resp. $\mathcal{O}\left(n^{3/2}\epsilon^{-3/2}+n^2\right)$) to achieve an ϵ-approximate stationary point of the main problem. In Remarks 2.5(vii) and 3.5(v), we discuss some differences among our methods and the methods in [6].

This work is organized as follows. Chapter 1 presents some definitions, notation and preliminary results. Chapter 2 presents our first method with lazy Hessian approximations and its first-order complexity results, whereas Chapter 3 discuss our second method with inexact gradient and lazy Hessian approximations and its first-order complexity results. Some final remarks are given in 4.

CHAPTER 1

Preliminary material

In this chapter, we present some definitions, notations, and key results essential for this work.

The symbol $\|.\|$ denotes the 2-norm for vectors or matrices (depending on the context), while $\|.\|_F$ denotes the Frobenius norm for matrices. The Euclidean inner product of $x, y \in \mathbb{R}^n$ is denoted by $\langle x, y \rangle$. For $j = 1, \ldots, n$, $e_j \in \mathbb{R}^n$ is the j-th orthonormal vector of the canonical basis for $\mathbb{R}^n$.

Throughout this work, the following conditions, besides those on f and ψ stated after Problem (0-3), are assumed:

(A1) The Hessian of f is L-Lipschitz continuous, i.e.

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|, \forall \quad x, y \in \mathbb{R}^n, \tag{1-1}$$

where $L \geq 0$ is the Lipschitz constant.

(A2) Exists a constant $F^\star$ such that $F(x) \geq F^\star$ for all $x \in \mathbb{R}^n$.

Next, we establish some function related definitions.

Definition 1.1. *A set $C \subset \mathbb{R}^n$ is said to be convex if the set*

$$[x, y] = \{(1-t)x + ty : x, y \in \mathbb{R}^n, t \in [0, 1]\},$$

is all contained in C. Let $C \subset \mathbb{R}^n$ be a convex set, f is said to be convex if

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y),$$

for all $x, y \in C$ and $t \in [0, 1]$.

Definition 1.2. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called proper if $dom(f) \neq \emptyset$, and for every $x \in dom(f)$,*

$$f(x) > -\infty.$$

Definition 1.3. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is closed if its epigraph, defined as*

$$epi(f) := \{(x, y) \in \mathbb{R}^{n+1} : f(x) \leq y\},$$

is a closed set.

Definition 1.4. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function and let $\bar{x} \in \mathrm{dom}(f)$. An element $r \in \mathbb{R}^n$ is called a subgradient of f at $\bar{x}$ if*

$$f(x) \geq f(\bar{x}) + \langle r, x - \bar{x}\rangle, \text{ for all } x \in \mathbb{R}^n.$$

The collection of all the subgradients of f at $\bar{x}$ is called the subdifferential of the function at this point and is denoted by $\partial f(\bar{x})$, equivalently:

$$\partial f(\bar{x}) := \{r \in \mathbb{R}^n \mid f(x) \geq f(\bar{x}) + \langle r, x - \bar{x}\rangle, \forall x \in \mathbb{R}^n\}.$$

In the following, we present our concept of approximate first-order stationary points.

Definition 1.5. *Given $\epsilon > 0$, we say that $\bar{x} \in \mathbb{R}^n$ is an ϵ-approximate first-order stationary point of (0-3) if $\|\nabla f(\bar{x}) + \psi'(\bar{x})\| \leq \epsilon$, for some $\psi'(\bar{x}) \in \partial\psi(\bar{x})$.*

We begin by presenting well-known properties for the function f when it satisfies **A1**. The proofs are shown for the sake of completeness.

Proposition 1.6. *([14, Lemma 1]) Suppose* **A1** *holds, then, for all $x, y \in \mathbb{R}^n$, we have*

$$\|\nabla f(y) - \nabla f(x) - \nabla^2 f(x)(y - x)\| \leq \frac{L}{2}\|y - x\|^2, \tag{1-2}$$

and

$$\left| f(y) - f(x) - \langle\nabla f(x), y - x\rangle - \frac{1}{2}\langle\nabla^2 f(x)(y - x), y - x\rangle \right| \leq \frac{L}{6}\|y - x\|^3. \tag{1-3}$$

Proof. Since

$$\frac{d}{d\tau}\nabla f(x + \tau(y - x)) = \nabla^2 f(x + \tau(y - x))(y - x),$$

we obtain

$$\int_0^1 \nabla^2 f(x + \tau(y - x))(y - x)d\tau = \nabla f(y) - \nabla f(x).$$

Hence, using **A1**, we find

$$\begin{aligned}
\|\nabla f(y)-\nabla f(x)-\nabla^2 f(x)(y-x)\| &= \left\|\int_0^1 [\nabla^2 f(x+\tau(y-x))-\nabla^2 f(x)](y-x)d\tau\right\| \\
&\leq \int_0^1 \left\|\nabla^2 f(x+\tau(y-x))-\nabla^2 f(x)\right\| \|y-x\| d\tau \\
&\leq L\|y-x\|^2 \int_0^1 \tau d\tau = \frac{1}{2}L\|y-x\|^2,
\end{aligned}$$

which concludes the proof of (1-2). Now, since

$$f(y)-f(x)=\int_0^1 \langle \nabla f(x+\tau(y-x)), y-x\rangle d\tau,$$

we obtain

$$\begin{aligned}
|f(y)-&f(x)-\langle \nabla f(x), y-x\rangle - \frac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle| \\
&= \left|\int_0^1 \langle \nabla f(x+\tau(y-x))-\nabla f(x)-\tau\nabla^2 f(x)(y-x), y-x\rangle d\tau\right| \\
&\leq \int_0^1 \|\nabla f(x+\tau(y-x))-\nabla f(x)-\tau\nabla^2 f(x)(y-x)\|\|y-x\| d\tau.
\end{aligned}$$

Take $\bar{x}=x+\tau(y-x)$, then $\tau(y-x)=\bar{x}-x$, so

$$\begin{aligned}
|f(y)-&f(x)-\langle \nabla f(x), y-x\rangle - \frac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle| \\
&\leq \int_0^1 \|\nabla f(\bar{x})-\nabla f(x)-\nabla^2 f(x)(\bar{x}-x)\|\|y-x\| d\tau,
\end{aligned}$$

combining the last inequality with (1-2), yields

$$\begin{aligned}
|f(y)-&f(x)-\langle \nabla f(x), y-x\rangle - \frac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle| \\
&\leq \frac{1}{2}L\|y-x\|^3 \int_0^1 \tau^2 d\tau = \frac{L}{6}\|y-x\|^3,
\end{aligned}$$

which proves (1-3). □

The next result, provides a useful inequality that is fundamental in the analysis of first-order methods for optimization.

Proposition 1.7. *([15, Lemma 1.2.3]) Suppose that ∇f is L_1-Lipschitz continuous. Then, for all $x, y \in \mathbb{R}^n$, we have*

$$|f(y)-f(x)-\langle \nabla f(x), y-x\rangle| \leq \frac{L_1}{2}\|y-x\|^2. \tag{1-4}$$

Proof. Note that

$$\int_0^1 \langle \nabla f(x+\tau(y-x)), (y-x)\rangle d\tau = f(y)-f(x),$$

then

$$\int_0^1 \langle \nabla f(x+\tau(y-x)) - \nabla f(x), (y-x)\rangle d\tau = f(y)-f(x)-\langle \nabla f(x), y-x\rangle.$$

Consequently

$$\begin{aligned} |f(y)-f(x)-\langle \nabla f(x), y-x\rangle| &\leq \|y-x\| \int_0^1 ||\nabla f(x+\tau(y-x)) - \nabla f(x)|| d\tau \\ &\leq \|y-x\| \int_0^1 \tau \|y-x\| d\tau \\ &\leq \frac{L_1}{2}\|y-x\|^2, \end{aligned}$$

which gives the desired result. □

The next lemma states a property for convex function known as the Jensen's inequality; see, for example, [13].

Lemma 1.8. *Let $D \subset \mathbb{R}^n$ a convex set and $f : D \to \mathbb{R}$ a convex function in D. So, for any $m \in \mathbb{N}$, $x_i \in \mathbb{R}^n$ and $p_i \in \mathbb{R}_+$, $i = 1, 2, \ldots, m$ such that $\sum_{i=1}^m p_i = 1$, it follows*

$$f\left(\sum_{i=1}^m p_i x_i\right) \leq \sum_{i=1}^m p_i f(x_i). \tag{1-5}$$

Proof. We will prove by induction on m. First, let $m = 2$. As f is a convex function, we have $f((1-t)x_1+tx_2) \leq (1-t)f(x_1)+tf(x_2)$, for all $t \in [0, 1]$. Hence, by defining p_1 and p_2 such that $p_1 = (1-t)$ and $p_2 = t$, we obtain that the inequality (1-5) holds for $m = 2$. Now, suppose that (1-5) is true for some $m > 2$, that is,

$$f\left(\sum_{i=1}^m p_i x_i\right) \leq \sum_{i=1}^m p_i f(x_i),$$

with $\sum_{i=1}^m p_i = 1$. Then,

$$\begin{aligned} f\left(\sum_{i=1}^{m+1} p_i x_i\right) &= f\left(\sum_{i=1}^{m} p_i x_i + p_{m+1}x_{m+1}\right) \\ &= f\left((1-p_{m+1})\sum_{i=1}^{m} \frac{p_i}{1-p_{m+1}} x_i + p_{m+1}x_{m+1}\right), \end{aligned}$$

which, combined with the fact that f is convex, yields

$$f\left(\sum_{i=1}^{m+1} p_i x_i\right) \leq (1-p_{m+1}) f\left(\sum_{i=1}^{m} \frac{p_i}{1-p_{m+1}} x_i\right) + p_{m+1} f(x_{m+1}).$$

Since $p_1 + p_2 + \cdots + p_m = 1 - p_{m+1}$, we have $\sum_{i=1}^{m} p_i/(1-p_{m+1}) = 1$. So, by (1-5), we conclude that

$$f\left(\sum_{i=1}^{m+1} p_i x_i\right) \leq \sum_{i=1}^{m} p_i f(x_i) + p_{m+1} f(x_{m+1}) = \sum_{i=1}^{m+1} p_i f(x_i),$$

which proves the lemma. □

The next inequality is essential to establish the complexity of our algorithms.

Proposition 1.9. *([7, Lemma B1]) If $\{r_i\}_{i=1}^{m-1}$ is a sequence of non-negative numbers, then*

$$\sum_{k=1}^{m-1}\left(\sum_{i=1}^{k} r_i\right)^3 \leq \frac{m^3}{3}\sum_{k=1}^{m-1} r_k^3. \tag{1-6}$$

Proof. We will prove the inequality by induction on m. First, consider $m = 2$. Hence,

$$\sum_{k=1}^{m-1}\left(\sum_{i=1}^{k} r_i\right)^3 = \sum_{k=1}^{1}\left(\sum_{i=1}^{k} r_i\right)^3 = r_1^3 \leq \frac{2^3}{3} r_1^3 = \frac{m^3}{3}\sum_{k=1}^{m-1} r_k^3,$$

Then (1-6) holds for $m = 2$. Assume now that the inequality in (1-6) is true for $m > 2$, let us prove that it holds for $m + 1$. Note that

$$\begin{aligned}\sum_{k=1}^{m}\left(\sum_{i=1}^{k} r_i\right)^3 &= \sum_{k=1}^{m-1}\left(\sum_{i=1}^{k} r_i\right)^3 + \left(\sum_{i=1}^{m} r_i\right)^3 \\ &\leq \frac{m^3}{3}\sum_{k=1}^{m} r_k^3 + \left(\sum_{i=1}^{m} r_i\right)^3.\end{aligned}$$

Since $t \mapsto t^3$, $t \geq 0$, is convex, by applying Lemma (1.8) with $f(t) = t^3$, $p_i = 1/m$ and $x_i = r_i$, we have

$$\left(\sum_{i=1}^{m} r_i\right)^3 \leq m^2 \sum_{k=1}^{m} r_k^3.$$

Therefore, combining the last two inequalities, we get

$$\sum_{k=1}^{m}\left(\sum_{i=1}^{k} r_i\right)^3 \leq \frac{m^3}{3}\sum_{k=1}^{m} r_k^3 + m^2\sum_{k=1}^{m} r_k^3 = \frac{m^3+3m^2}{3}\sum_{k=1}^{m+1} r_k^3 \leq \frac{(m+1)^3}{3}\sum_{k=1}^{m} r_k^3,$$

which implies that the inequality in (1-6) is true for $m+1$. □

The following inequality, see [11, Example 1.2-3], is known as Young's inequality, and it is essential for the construction of our algorithm acceptance conditions.

Proposition 1.10. *If a, b are positive numbers and $p, q > 1$ satisfies*

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Proof. Let $p, q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

From the equation above, we can deduce that

$$1 = \frac{p+q}{pq}.$$

So $pq = p + q$. Additionally, note that

$$(p-1)(q-1) = 1.$$

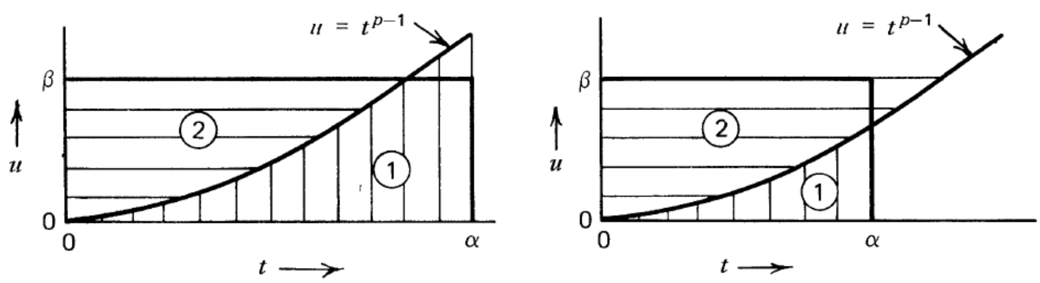
From this, we observe that

$$\frac{1}{p-1} = q - 1.$$

Now, consider $u = t^{p-1}$, then $t = u^{q-1}$. These follow directly from the relationship between p and q. Let α and β be any positive numbers. Since $\alpha\beta$ is the area of a rectangle, see the figure from below, we obtain by integration the inequality

$$\alpha\beta \leq \int_0^\alpha t^{p-1} dt + \int_0^\beta u^{q-1} du = \frac{\alpha^p}{p} + \frac{\beta^q}{q},$$

and it is trivially true if $\alpha = 0$ or $\beta = 0$. □



1.1 Approximations of the gradient and Hessian

In this section, we outline methods for approximating the gradient and Hessian of f using finite difference techniques. Gradient approximations are discussed in detail in Section 1.1.1, while Hessian approximations are presented in Section 1.1.2.

1.1.1 Approximations of the gradient

Here, we discuss how to compute approximations for the gradient of f using finite differences approaches. The first lemma provides an approximation without requiring any additional assumptions beyond **A1**.

Lemma 1.11. *Suppose **A1** holds. Given $x \in \mathbb{R}^n$ and $h > 0$, define $g \in \mathbb{R}^n$ by*

$$g(x) = \left(\frac{f(x+he_1) - f(x-he_1)}{2h}, \dots, \frac{f(x+he_n) - f(x-he_n)}{2h}\right). \tag{1-7}$$

Then,

$$\|g(x) - \nabla f(x)\| \leq \frac{\sqrt{n}L}{6}h^2. \tag{1-8}$$

Proof. Let $i \in \{1, 2, \dots, n\}$. Applying inequality (1-3) at the points $y = x + he_i$ and $y = x - he_i$, we obtain, respectively,

$$\left|f(x+he_i) - f(x) - h\langle \nabla f(x), e_i\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle\right| \leq \frac{L}{6}h^3,$$

and

$$\left|f(x-he_i) - f(x) + h\langle \nabla f(x), e_i\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle\right| \leq \frac{L}{6}h^3.$$

Combining the last two inequalities, we obtain

$$|f(x+he_i) - f(x-he_i) - 2h\langle \nabla f(x), e_i\rangle| \leq \frac{L}{3}h^3,$$

dividing the last inequality by $2h$, and from the definition of g, yields

$$|(g(x) - \nabla f(x))e_i| \leq \frac{L}{6}h^2.$$

Therefore

$$\|g(x) - \nabla f(x)\|^2 \leq \sum_{t=1}^{n} \|(g(x) - \nabla f(x))e_i\|^2 \leq \frac{nL^2}{6^2}h^4,$$

which implies the desired inequality. □

Next, we present an alternative method for approximating the gradient of f. However, this approach requires the additional assumption that the gradient of f is L_1-Lipschitz continuous. Under this condition, we obtain the following lemma.

Lemma 1.12. *Suppose that ∇f is L_1-Lipschitz continuous. Given $x \in \mathbb{R}^n$ and $h > 0$, define $g \in \mathbb{R}^n$ by*

$$g(x) = \left(\frac{f(x+he_1) - f(x)}{h}, \ldots, \frac{f(x+he_n) - f(x)}{h}\right), \tag{1-9}$$

or

$$g(x) = \left(\frac{f(x+he_1) - f(x-he_1)}{2h}, \ldots, \frac{f(x+he_n) - f(x-he_n)}{2h}\right), \tag{1-10}$$

or

$$g(x) = \left(\frac{f(x) - f(x-he_1)}{h}, \ldots, \frac{f(x) - f(x-he_n)}{h}\right). \tag{1-11}$$

Then,

$$\|g(x) - \nabla f(x)\| \leq \frac{\sqrt{n}L_1}{2}h. \tag{1-12}$$

Proof. Let $i \in \{1, 2, \ldots, n\}$. Using the inequality in (1-4) with $y = x + he_i$, we have

$$|f(x+he_i) - f(x) - h\langle \nabla f(x), e_i\rangle| \leq \frac{L_1}{2}h^2, \tag{1-13}$$

or, equivalently,

$$\left|\frac{f(x+he_i) - f(x)}{h} - \langle \nabla f(x), e_i\rangle\right| \leq \frac{L_1}{2}h.$$

Consider g as defined in (1-9). Hence,

$$|(g(x) - \nabla f(x))e_i| \leq \frac{L_1}{2}h. \tag{1-14}$$

With similar arguments, it can be proven that (1-12) also holds for g as in (1-11). On the other hand, taking $y = x - he_i$ in (1-4), we have

$$|f(x) - f(x-he_i) - h\langle \nabla f(x), e_i\rangle| \leq \frac{L_1}{2}h^2.$$

Combing the last inequality with (1-13), we get

$$|f(x+he_i) - f(x-he_i) - 2h\langle \nabla f(x), e_i\rangle| \leq L_1 h^2,$$

which, combined with g as defined in (1-10), implies that (1-14) also holds for the second choice of g. Thus, it follows from (1-14) that

$$\|g(x) - \nabla f(x)\|^2 \leq \sum_{t=1}^{n} \|(g(x) - \nabla f(x))e_i\|^2 \leq \frac{nL_1^2}{2^2}h^2,$$

then

$$\|g(x) - \nabla f(x)\| \leq \frac{\sqrt{n}L_1}{2}h,$$

which proves the inequality (1-12). □

1.1.2 Approximation of the Hessian

In this section, we discuss ways to approximate the Hessian of f using function or gradient evaluations.

We begin by presenting approximations based on gradient evaluations; the result for the first choice of A can be found in [8].

Lemma 1.13. *Suppose* ***A1*** *holds. Given $x \in \mathbb{R}^n$ and $h > 0$, consider $A \in \mathbb{R}^{n\times n}$ defined by*

$$A(x) = \left[\frac{\nabla f(x + he_1) - \nabla f(x)}{h}, \ldots, \frac{\nabla f(x + he_n) - \nabla f(x)}{h}\right], \tag{1-15}$$

or

$$A(x) = \left[\frac{\nabla f(x + he_1) - \nabla f(x - he_1)}{2h}, \ldots, \frac{\nabla f(x + he_n) - \nabla f(x - he_n)}{2h}\right], \tag{1-16}$$

or

$$A(x) = \left[\frac{\nabla f(x) - \nabla f(x - he_1)}{h}, \ldots, \frac{\nabla f(x) - \nabla f(x - he_n)}{h}\right]. \tag{1-17}$$

Then, the matrix

$$B(x) = \frac{1}{2}\left(A(x) + A^\top(x)\right), \tag{1-18}$$

satisfies

$$\|B(x) - \nabla^2 f(x)\| \leq \frac{\sqrt{n}L}{2}h. \tag{1-19}$$

Proof. Let $i \in \{1, 2, \ldots, n\}$. From inequality (1-2) with $y = x + he_i$, we have

$$\|\nabla f(x + he_i) - \nabla f(x) - h\nabla^2 f(x)e_i\| \leq \frac{L}{2}h^2,$$

equivalently,

$$\left\|\frac{\nabla f(x + he_i) - \nabla f(x)}{h} - \nabla^2 f(x)e_i\right\| \leq \frac{L}{2}h. \tag{1-20}$$

For A as in (1-15), it follows from the last inequality that

$$\|(A(x) - \nabla^2 f(x))e_i\| \leq \frac{L}{2}h. \tag{1-21}$$

With similar arguments, it can be proven that (1-21) also holds for A as in (1-17). On the other hand, take A as in (1-16), applying inequality (1-2) at the point $y = x - he_i$, we get

$$\left\|\frac{-\nabla f(x - he_i) + \nabla f(x)}{h} - \nabla^2 f(x)e_i\right\| \leq \frac{L}{2}h.$$

Combining the last inequality and (1-20), we have

$$\left\|\frac{\nabla f(x + he_i) - \nabla f(x + he_i)}{2h} - \nabla^2 f(x)e_i\right\| \leq \frac{L}{2}h,$$

which implies that the inequality (1-21) also holds for the second choice of A. Consequently, it follows from (1-21) that

$$\|A(x) - \nabla^2 f(x)\|^2 \leq \|A(x) - \nabla^2 f(x)\|_F^2 = \sum_{i=1}^{n} \|(A(x) - \nabla^2 f(x))e_i\|^2 \leq n\left(\frac{L}{2}h\right)^2,$$

which gives

$$\|A(x) - \nabla^2 f(x)\| \leq \frac{\sqrt{n}L}{2}h. \tag{1-22}$$

Finally, combining (1-18) and (1-22), we obtain

$$\|B(x) - \nabla^2 f(x)\| \leq \|A(x) - \nabla^2 f(x)\| \leq \frac{\sqrt{n}L}{2}h,$$

which proves the derived result in (1-19). □

In the next two lemmas, we explore methods to approximate the Hessian of f using only function evaluations.

Lemma 1.14. *Suppose **A1** holds. Given $x \in \mathbb{R}^n$ and $h > 0$, consider $A \in \mathbb{R}^{n\times n}$ defined by*

$$A_{ij}(x) = \frac{f(x + he_i + he_j) - f(x + he_i) - f(x + he_j) + f(x)}{h^2}, \tag{1-23}$$

for $i, j = 1, \ldots, n$. Then, the matrix

$$B(x) = \frac{1}{2}\left(A(x) + A^\top(x)\right), \tag{1-24}$$

satisfies

$$\|B(x) - \nabla^2 f(x)\| \leq \frac{5nL}{3}h. \tag{1-25}$$

Proof. Let $i, j \in \{1, 2, \ldots, n\}$. It follows from inequality (1-3) with $y = x + he_i + he_j$ that

$$\left| f(x + he_i + he_j) - f(x) - \langle \nabla f(x), he_i + he_j\rangle - \frac{1}{2}\langle \nabla^2 f(x)(he_i + he_j), he_i + he_j\rangle \right|$$
$$\leq \frac{L}{6}\|he_i + he_j\|^3. \tag{1-26}$$

If $i \neq j$, we obtain

$$\frac{L}{6}\|he_i + he_j\|^3 = \frac{Lh^3}{6}.\sqrt{2}^3 = \frac{Lh^3\sqrt{2}}{3}. \tag{1-27}$$

Now, if $i = j$, we have

$$\frac{L}{6}\|he_i + he_j\|^3 = \frac{Lh^3}{6}.\sqrt{4}^3 = \frac{8Lh^3}{6} = \frac{4Lh^3}{3}. \tag{1-28}$$

From (1-26), (1-27) and (1-28), it follows that

$$\Big| f(x + he_i + he_j) - f(x) - h\langle \nabla f(x), e_i\rangle - h\langle \nabla f(x), e_j\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle$$
$$- \frac{h^2}{2}\langle \nabla^2 f(x)e_j, e_j\rangle - h^2\langle \nabla^2 f(x)e_i, e_j\rangle \Big| \leq \frac{4Lh^3}{3}. \tag{1-29}$$

Now, using the inequality in (1-3) at the points $y = x + he_i$ and $y = x + he_j$, we obtain, respectively

$$\left| f(x) - f(x + he_i) + h\langle \nabla f(x), e_i\rangle + \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle \right| \leq \frac{L}{6}\|he_i\|^3 = \frac{Lh^3}{6}, \tag{1-30}$$

and

$$\left| f(x) - f(x + he_j) + h\langle \nabla f(x), e_j\rangle + \frac{h^2}{2}\langle \nabla^2 f(x)e_j, e_j\rangle \right| \leq \frac{L}{6}\|he_j\|^3 = \frac{Lh^3}{6}. \tag{1-31}$$

Combining (1-29), (1-30) and (1-31), we get

$$\left| f(x + he_i + he_j) - f(x + he_i) - f(x + he_j) + f(x) - h^2\langle \nabla^2 f(x)e_i, e_j\rangle \right| \leq \frac{5Lh^3}{3}.$$

Dividing the last inequality by h^2 and from the definition of A, it follows that

$$|A_{ij}(x) - \nabla^2 f(x)_{ij}| \leq \frac{5Lh}{3},$$

which, combined with the definition of B in (1-24), yields

$$\|B(x) - \nabla^2 f(x)\| \leq \|A(x) - \nabla^2 f(x)\| \leq n \max_{i=1,\ldots,n} |A_{ij}(x) - \nabla^2 f(x)_{ij}| \leq \frac{5nLh}{3},$$

proving the inequality in (1-25). $\square$

Lemma 1.15. *Suppose* ***A1*** *holds. Given $x \in \mathbb{R}^n$ and $h > 0$, consider $A \in \mathbb{R}^{n\times n}$ defined by*

$$A_{ij}(x) = \frac{f(x + h(e_i + e_j)) - f(x + h(e_i - e_j)) - f(x + h(e_j - e_i)) + f(x - h(e_i + e_j))}{4h^2}, \quad (1\text{-}32)$$

for $i, j = 1, \ldots, n$. Then, the matrix

$$B(x) = \frac{1}{2}\left(A(x) + A^\top(x)\right),$$

satisfies

$$\|B(x) - \nabla^2 f(x)\| \leq \frac{2nLh}{3}.$$

Proof. Let $i, j \in \{1, 2, \ldots, n\}$. Using the inequality (1-3) at four different points, namely $y = x + he_i + he_j$, $y = x + he_i - he_j$, $y = x + he_j - he_i$, and $y = x - he_i - he_j$, we obtain, respectively:

$$\Big| f(x + he_i + he_j) - f(x) - h\langle \nabla f(x), e_i\rangle - h\langle \nabla f(x), e_j\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle$$
$$- \frac{h^2}{2}\langle \nabla^2 f(x)e_j, e_j\rangle - h^2\langle \nabla^2 f(x)e_i, e_j\rangle \Big| \leq \frac{L}{6}\|he_i + he_j\|^3.$$

$$\Big| f(x + he_i - he_j) - f(x) - h\langle \nabla f(x), e_i\rangle + h\langle \nabla f(x), e_j\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle$$
$$- \frac{h^2}{2}\langle \nabla^2 f(x)e_j, e_j\rangle + h^2\langle \nabla^2 f(x)e_i, e_j\rangle \Big| \leq \frac{L}{6}\|he_i - he_j\|^3$$

$$\Big| f(x + he_j - he_i) - f(x) - h\langle \nabla f(x), e_j\rangle - h\langle \nabla f(x), e_i\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_j, e_j\rangle$$
$$- \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle + h^2\langle \nabla^2 f(x)e_i, e_j\rangle \Big| \leq \frac{L}{6}\|he_j - he_i\|^3.$$

$$\Big| f(x - he_j - he_i) - f(x) + h\langle \nabla f(x), e_i\rangle + h\langle \nabla f(x), e_j\rangle - \frac{h^2}{2}\langle \nabla^2 f(x)e_i, e_i\rangle$$
$$- \frac{h^2}{2}\langle \nabla^2 f(x)e_j, e_j\rangle - h^2\langle \nabla^2 f(x)e_i, e_j\rangle \Big| \leq \frac{L}{6}\|he_i + he_j\|^3.$$

Combining these four inequalities, we get

$$\begin{aligned}&\big|f(x+he_i+he_j)-f(x+he_i-he_j)-f(x+he_j-he_i)+f(x-he_j-he_i)\\&-4h^2\langle\nabla^2 f(x)e_i,e_j\rangle\big|\leq\frac{Lh^3}{6}(2\|e_i+e_j\|^3+2\|e_i-e_j\|^3)\leq\frac{16Lh^3}{6}=\frac{8Lh^3}{3}.\end{aligned}$$

Dividing the last inequality by $4h^2$, and from the definition of A, it follows

$$|A_{ij}(x)-\nabla^2 f(x)_{ij}|\leq\frac{2Lh}{3}.$$

From the definition of B and the last inequality, we have

$$\|B(x)-\nabla^2 f(x)\|\leq\|A(x)-\nabla^2 f(x)\|\leq n\max_{i=1,\ldots,n}|A_{ij}(x)-\nabla^2 f(x)_{ij}|\leq\frac{2nLh}{3},$$

which concludes the proof. □

The following chapter presents some auxiliary results, our first algorithm for solving the general problem (0-3) and last the global complexity results related to the algorithm.

CHAPTER 2

Cubic Regularization Method with Lazy Hessian Approximations

In this chapter, we present our first algorithm for solving (0-3). It consists of a variant of the cubic regularization method in which the gradient of the function f is assumed to be available, and the Hessian is approximated in such a way that a suitable condition is satisfied. It is worth mentioning that the approximation of the Hessian is not updated at every iteration.

The Chapter is organized as follows. Section 2.1 introduces some essential auxiliary results. Section 2.2 presents the first algorithm. And finally, Section 2.3 shows iteration complexity results for the algorithm.

2.1 Auxiliary results

The subproblem of the proposed algorithm will be associated to the following cubic model

$$M_{x,z,\sigma}(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\langle B(z)(y-x), y-x \rangle + \frac{\sigma}{6}\|y-x\|^3, \quad (2\text{-}1)$$

where $x, z \in \mathbb{R}^n$, $\sigma > 0$ and $B(z)$ is an approximation to the Hessian of f on a point z. In the context of the implementation of the algorithm, the point $x^+ \in \mathbb{R}^n$ is taken to be an approximate solution to the subproblem

$$\min_{y \in Q} M_{x,z,\sigma}(y) + \psi(y).$$

Recall that $F(x) = f(x) + \psi(x)$, we will first derive an estimate for the expression $F(x) - F(x^+)$, which is related to our algorithm acceptance condition. Subsequently, we acquire a bound on $\|\nabla f(x^+) + \psi'(x^+)\|$, which is later used in the complexity results.

This follow proposition provides an estimate for $F(x) - F(x^+)$ if $M_{x,z,\sigma}(\cdot)$, $B(\cdot)$, $x^+ \in \mathbb{R}^n$ and $\sigma > 0$ satisfy certain conditions.

Proposition 2.1. *Suppose* ***A1*** *holds and assume that* $x^+ \in \mathbb{R}^n$ *satisfies*

$$M_{x,x,\sigma}(x^+) + \psi(x^+) \leq F(x), \tag{2-2}$$

for some $x \in \mathbb{R}^n$ *and* $\sigma > 0$. *Moreover, suppose that for some* $\kappa_B \geq 0$ *and* $\hat{x} \in \mathbb{R}^n$, *it holds that*

$$\|B(x) - \nabla^2 f(x)\| \leq \kappa_B \|x - \hat{x}\|. \tag{2-3}$$

If

$$\sigma \geq 2(L + \sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}}), \tag{2-4}$$

for some $\bar{\rho} > 0$, *then*

$$F(x) - F(x^+) \geq \frac{\sigma}{12}\|x^+ - x\|^3 - \frac{1}{3\bar{\rho}}\|x - \hat{x}\|^3. \tag{2-5}$$

Proof. From (1-3) and the definition of $M_{x,z,\sigma}(\cdot)$ in (2-1), we have

$$\begin{aligned} f(x^+) &\leq f(x) + \langle \nabla f(x), x^+ - x \rangle + \frac{1}{2}\langle \nabla^2 f(x)(x^+ - x), x^+ - x \rangle + \frac{L}{6}\|x^+ - x\|^3 \\ &\leq M_{x,x,\sigma}(x^+) + \frac{1}{2}\langle (\nabla^2 f(x) - B(x))(x^+ - x), x^+ - x \rangle + \frac{L - \sigma}{6}\|x^+ - x\|^3. \end{aligned}$$

Therefore, by the last inequality, along with (2-2), (2-3), $F = f + \psi$ and Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} F(x^+) &\leq F(x) + \frac{1}{2}\|B(x) - \nabla^2 f(x)\|\|x^+ - x\|^2 + \frac{L - \sigma}{6}\|x^+ - x\|^3 \\ &\leq F(x) + \frac{\kappa_B}{2}\|x - \hat{x}\|\|x^+ - x\|^2 + \frac{L - \sigma}{6}\|x^+ - x\|^3 \\ &= F(x) + \frac{\|x - \hat{x}\|}{\bar{\rho}^{\frac{1}{3}}}\frac{\kappa_B\bar{\rho}^{\frac{1}{3}}\|x^+ - x\|^2}{2} + \frac{L - \sigma}{6}\|x^+ - x\|^3. \end{aligned} \tag{2-6}$$

From Proposition 1.10 with

$$p = 3, \quad q = \frac{3}{2} \quad a = \frac{\|x - \hat{x}\|}{\bar{\rho}^{\frac{1}{3}}} \quad \text{and} \quad b = \frac{\kappa_B\bar{\rho}^{\frac{1}{3}}\|x^+ - x\|^2}{2},$$

we have

$$\frac{\|x - \hat{x}\|}{\bar{\rho}^{\frac{1}{3}}}\frac{\kappa_B\bar{\rho}^{\frac{1}{3}}\|x^+ - x\|^2}{2} \leq \frac{\|x - \hat{x}\|^3}{3\bar{\rho}} + \frac{\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}}\|x^+ - x\|^3}{3\sqrt{2}}.$$

Combining the last inequality and (2-6), we obtain

$$F(x^+) \leq F(x) + \left(\frac{L + \sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}} - \sigma}{6}\right)\|x^+ - x\|^3 + \frac{1}{3\bar{\rho}}\|x - \hat{x}\|^3,$$

which, combined with (2-4), implies (2-5). □

The next proposition gives another estimate for $F(x) - F(x^+)$. In the context of the algorithm implementation, the approximation B of the Hessian that we made on $x \in \mathbb{R}^n$ is reused for the next $m \geq 0$ fixed iterations.

Proposition 2.2. *Suppose* ***A1*** *holds and assume that $x^+ \in \mathbb{R}^n$ satisfies*

$$M_{x,z,\sigma}(x^+) + \psi(x^+) \leq F(x), \tag{2-7}$$

for some $x, z \in \mathbb{R}^n$ and $\sigma > 0$. Moreover, suppose that for some $\kappa_B \geq 0$, $\hat{z} \in \mathbb{R}^n$, it holds that

$$\|B(z) - \nabla^2 f(z)\| \leq \kappa_B \|z - \hat{z}\|. \tag{2-8}$$

If

$$\sigma \geq 2(L + \sqrt{2}L^{\frac{3}{2}}\hat{\rho}^{\frac{1}{2}} + \sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}}), \tag{2-9}$$

for some $\hat{\rho}, \bar{\rho} > 0$, then

$$F(x) - F(x^+) \geq \frac{\sigma}{12}\|x^+ - x\|^3 - \frac{1}{3\hat{\rho}}\|x - z\|^3 - \frac{1}{3\bar{\rho}}\|z - \hat{z}\|^3. \tag{2-10}$$

Proof. From (1-3) and definition of $M_{x,z,\sigma}(\cdot)$ in (2-1), we have

$$\begin{aligned} f(x^+) &\leq f(x) + \langle \nabla f(x), x^+ - x\rangle + \frac{1}{2}\langle \nabla^2 f(x)(x^+ - x), x^+ - x\rangle + \frac{L}{6}\|x^+ - x\|^3 \\ &= M_{x,z,\sigma}(x^+) + \frac{1}{2}\langle (\nabla^2 f(x) - B(z))(x^+ - x), x^+ - x\rangle + \frac{L - \sigma}{6}\|x^+ - x\|^3. \end{aligned}$$

The last inequality, combined with (2-7), (2-8), $F = f + \psi$ and Cauchy-Schwarz inequality, yields

$$\begin{aligned} F(x^+) &\leq F(x) + \frac{1}{2}\|\nabla^2 f(x) - \nabla^2 f(z)\|\|x^+ - x\|^2 + \frac{1}{2}\|B(z) - \nabla^2 f(z)\|\|x^+ - x\|^2 \\ &\quad + \frac{L - \sigma}{6}\|x^+ - x\|^3 \\ &\leq F(x) + \frac{L}{2}\|x - z\|\|x^+ - x\|^2 + \frac{\kappa_B}{2}\|z - \hat{z}\|\|x^+ - x\|^2 + \frac{L - \sigma}{6}\|x^+ - x\|^3 \\ &= F(x) + \frac{\|x - z\|}{\hat{\rho}^{\frac{1}{3}}}\frac{L\hat{\rho}^{\frac{1}{3}}}{2}\|x^+ - x\|^2 + \frac{\|z - \hat{z}\|}{\bar{\rho}^{\frac{1}{3}}}\frac{\kappa_B}{2}\bar{\rho}^{\frac{1}{3}}\|x^+ - x\|^2 \\ &\quad + \frac{L - \sigma}{6}\|x^+ - x\|^3, \end{aligned} \tag{2-11}$$

From Proposition 1.10 with

$$p = 3 \quad q = \frac{3}{2} \quad a = \frac{\|x - z\|}{\hat{\rho}^{\frac{1}{3}}} \quad \text{and} \quad b = \frac{L\hat{\rho}^{\frac{1}{3}}}{2}\|x^+ - x\|^2,$$

we obtain

$$\frac{\|z - \hat{z}\|}{\bar{\rho}^{\frac{1}{3}}}\frac{\kappa_B}{2}\bar{\rho}^{\frac{1}{3}}\|x^+ - x\|^2 \leq \frac{1}{3\bar{\rho}}\|z - \hat{z}\|^3 + \frac{\sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}}}{6}\|x^+ - x\|^3. \tag{2-12}$$

Again, from Proposition 1.10 with

$$p = 3 \quad q = \frac{3}{2} \quad a = \frac{\|z - \hat{z}\|}{\bar{\rho}^{\frac{1}{3}}} \quad \text{and} \quad b = \frac{\kappa_B}{2}\bar{\rho}^{\frac{1}{3}}\|x^+ - x\|^2,$$

we get

$$\frac{1}{\hat{\rho}^{\frac{1}{3}}}\|x - z\|\frac{L\hat{\rho}^{\frac{1}{3}}}{2}\|x^+ - x\|^2 \leq \frac{1}{3\hat{\rho}}\|x - z\|^3 + \frac{\sqrt{2}L^{\frac{3}{2}}\hat{\rho}^{\frac{1}{2}}}{6}\|x^+ - x\|^3. \tag{2-13}$$

It follows from inequality (2-11), (2-12) and (2-13) that

$$F(x^+) \leq F(x) + \left(\frac{L + \sqrt{2}L^{\frac{3}{2}}\hat{\rho}^{\frac{1}{2}} + \sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}} - \sigma}{6}\right)\|x^+ - x\|^3 + \frac{\|x - z\|^3}{3\hat{\rho}} + \frac{\|z - \hat{z}\|^3}{3\bar{\rho}},$$

which, combined with (2-9), implies the desired inequality. □

Combining the last proposition with Lemma 1.13, we can establish conditions on h and σ such that $\|B(z) - \nabla^2 f(z)\| \leq \kappa_B\|z - \hat{z}\|$, and the inequality (2-10) holds.

Theorem 2.3. *Suppose* ***A1*** *holds. Let $x, z, \hat{z} \in \mathbb{R}^n$, $\kappa_B \geq 0$ and $\sigma > 0$. Consider $M_{x,z,\sigma}(\cdot)$ as in (2-1) where the matrix $B(z) \in \mathbb{R}^{n\times n}$ is defined by*

$$B(z) = \frac{1}{2}(A(z) + A^\top(z)),$$

where A is as in (1-15), (1-16) or (1-17) with $x = z$ and

$$0 < h \leq \frac{2\kappa_B}{\sqrt{n}\sigma}\|z - \hat{z}\|. \tag{2-14}$$

Assume also that $x^+ \in \mathbb{R}^n$ satisfies (2-7). If

$$\sigma \geq 2(L + \sqrt{2}L^{\frac{3}{2}}\hat{\rho}^{\frac{1}{2}} + \sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}}), \tag{2-15}$$

for some $\hat{\rho}, \bar{\rho} > 0$*, then*

$$F(x) - F(x^+) \geq \frac{\sigma}{12}\|x^+ - x\|^3 - \frac{1}{3\hat{\rho}}\|x - z\|^3 - \frac{1}{3\bar{\rho}}\|z - \hat{z}\|^3.$$

Proof. It follows from Lemma 1.13 with $x = z$ and (2-14) that

$$\|B(z) - \nabla^2 f(z)\| \leq \frac{\sqrt{n}L}{2}h \leq \frac{\sqrt{n}L}{2}\frac{2\kappa_B}{\sqrt{n}\sigma}\|z - \hat{z}\| = \frac{\kappa_B L}{\sigma}\|z - \hat{z}\|.$$

Since (2-15) implies $L/\sigma \leq 1$, we have

$$\|B(z) - \nabla^2 f(z)\| \leq \kappa_B\|z - \hat{z}\|.$$

Therefore, since all the hypothesis of Proposition 2.2 hold, then the desired inequality follows trivially from Proposition 2.2. □

The next proposition establishes a bound on $\|\nabla f(x^+) + \psi'(x^+)\|$ if $B(\cdot)$, $\nabla M_{x,z,\sigma}(\cdot)$ and $x^+ \in \mathbb{R}^n$ satisfy some conditions.

Proposition 2.4. *Suppose* ***A1*** *holds and assume that* (2-8) *holds for some* $\kappa_B \geq 0$ *and* $z, \hat{z} \in \mathbb{R}^n$*. Moreover, assume that* $x^+ \in \mathbb{R}^n$ *satisfies*

$$\|\nabla M_{x,z,\sigma}(x^+) + \psi'(x^+)\| \leq \theta\|x^+ - x\|^2, \tag{2-16}$$

for some $\psi'(x^+) \in \partial\psi(x^+)$*,* $x \in \mathbb{R}^n$*,* $\theta \geq 0$ *and* $\sigma > 0$*. If* $\rho, \rho^\star > 0$*, then*

$$\|\nabla f(x^+) + \psi'(x^+)\|^{\frac{3}{2}} \leq \sqrt{3}\left[\left(\frac{\sigma + L + L^2\rho + \kappa_B^2\rho^\star + 2\theta}{2}\right)^{\frac{3}{2}}\|x^+ - x\|^3 + \frac{\|x - z\|^3}{2^{\frac{3}{2}}\rho^{\frac{3}{2}}} + \frac{\|z - \hat{z}\|^3}{2^{\frac{3}{2}}\rho^{\star\frac{3}{2}}}\right]. \tag{2-17}$$

Proof. From the definition of $M_{x,z,\sigma}(\cdot)$ in (2-1), we have

$$\nabla M_{x,z,\sigma}(y) = \nabla f(x) + B(z)(y - x) + \frac{\sigma}{2}\|y - x\|(y - x).$$

Hence, using (2-16) and the triangle inequality, we obtain

$$
\begin{aligned}
\|\nabla f(x^+) + \psi'(x^+)\| &\leq \|\nabla f(x^+) - \nabla M_{x,z,\sigma}(x^+)\| + \|\nabla M_{x,z,\sigma}(x^+) + \psi'(x^+)\| \\
&\leq \|\nabla f(x^+) - \nabla f(x) - B(z)(x^+ - x)\| + \theta\|x^+ - x\|^2 \\
&+ \frac{\sigma}{2}\|x^+ - x\|^2 \\
&\leq \|\nabla f(x^+) - \nabla f(x) - \nabla^2 f(x)(x^+ - x)\| + \left(\frac{\sigma}{2} + \theta\right)\|x^+ - x\|^2 \\
&+ \|(\nabla^2 f(x) - B(z))(x^+ - x)\| \\
&\leq \|\nabla f(x^+) - \nabla f(x) - \nabla^2 f(x)(x^+ - x)\| + \left(\frac{\sigma}{2} + \theta\right)\|x^+ - x\|^2 \\
&+ \|(\nabla^2 f(x) - \nabla^2 f(z))(x^+ - x)\| + \|(\nabla^2 f(z) - B(z))(x^+ - x)\|.
\end{aligned}
$$

From last inequality, (1-1), (1-2) and (2-8), we get

$$
\begin{aligned}
\|\nabla f(x^+) + \psi'(x^+)\| \leq \left(\frac{L+\sigma}{2} + \theta\right)\|x^+ - x\|^2 + L\|x - z\|\|x^+ - x\| \\
+ \kappa_B\|z - \hat{z}\|\|x^+ - x\|.
\end{aligned}
\tag{2-18}
$$

On the other hand, it follows from Proposition 1.10 with $p = q = 2$ that

$$
ab \leq \frac{a^2}{2} + \frac{b^2}{2}, \forall\ a, b \geq 0.
$$

Hence, using the last property with $a = \|x - z\|/\rho^{1/2}$ and $b = L\rho^{1/2}\|x^+ - x\|$, we obtain

$$
L\|x - z\|\|x^+ - x\| \leq \frac{\|x - z\|^2}{2\rho} + \frac{L^2\rho\|x^+ - x\|^2}{2}. \tag{2-19}
$$

Again, by the same property with $a = \|z - \hat{z}\|/\rho^{\star\frac{1}{2}}$ and $b = \kappa_B\rho^{\star\frac{1}{2}}\|x^+ - x\|$, we have

$$
\kappa_B\|z - \hat{z}\|\|x^+ - x\| \leq \frac{\|z - \hat{z}\|^2}{2\rho^\star} + \frac{\kappa_B^2\rho^\star\|x^+ - x\|^2}{2}. \tag{2-20}
$$

It follows from (2-18), (2-19) and (2-20) that

$$
\begin{aligned}
\|\nabla f(x^+) + \psi'(x^+)\| \leq \left(\frac{\sigma + L + L^2\rho + \kappa_B^2\rho^\star + 2\theta}{2}\right)\|x^+ - x\|^2 + \frac{1}{2\rho}\|x - z\|^2 \\
+ \frac{1}{2\rho^\star}\|z - \hat{z}\|^2.
\end{aligned}
$$

Raising the last inequality to the power of 3/2 and using the inequality (1-5) for the function $t \mapsto t^{3/2}$, $t \geq 0$ with $p_1 = p_2 = p_3 = 1/3$, we get

$$\|\nabla f(x^+) + \psi'(x^+)\|^{\frac{3}{2}} \leq \sqrt{3}\left[\left(\frac{\sigma + L + L^2\rho + \kappa_B^2\rho^\star + 2\theta}{2}\right)^{\frac{3}{2}}\|x^+ - x\|^3 + \frac{\|x - z\|^3}{2^{\frac{3}{2}}\rho^{\frac{3}{2}}} + \frac{\|z - \hat{z}\|^3}{2^{\frac{3}{2}}\rho^{\star\frac{3}{2}}}\right],$$

which proves (2-17). $\square$

2.2 Cubic Regularization Method with Lazy Hessian Approximations

In this section, we propose and analyze an algorithm that finds an ϵ-approximate critical points for the problem (0-3).

Algorithm 1. Cubic Regularization Method with Lazy Hessian Approximations

Step 0. Choose $x_{-1}, x_0 \in \mathbb{R}^n$, $m \in \mathbb{N}$, $\sigma_0 > 0$, $\theta \geq 0$, $\bar{\kappa}_B \geq 0$, and set $t := 0$.
Step 1. Find the smallest integer $i \geq 0$ such that $2^{i-1}\sigma_t \geq \sigma_0(m+1)$.
Step 1.1. Construct $B_{t,i}$ such that

$$\|B_{t,i}(x_t) - \nabla^2 f(x_t)\| \leq \frac{\bar{\kappa}_B}{2^{i-1}}\|x_t - x_{t-1}\|. \tag{2-21}$$

Step 1.2. Compute $x_{t,i}^+$ such that

$$M_{x_t,x_t,2^i\sigma_t}(x_{t,i}^+) + \psi(x_{t,i}^+) \leq F(x_t), \quad \left\|\nabla M_{x_t,x_t,2^i\sigma_t}(x_{t,i}^+) + \psi'(x_{t,i}^+)\right\| \leq \theta\|x_{t,i}^+ - x_t\|^2, \tag{2-22}$$

for some $\psi'(x_{t,i}^+) \in \partial\psi(x_{t,i}^+)$, where $M_{x,z,\sigma}(\cdot)$ is as in (2-1).
Step 1.3. Take $\gamma_1 := m$ if $t = 0$ and $\gamma_1 := 0$ otherwise. If

$$F(x_t) - F(x_{t,i}^+) \geq \frac{2^i\sigma_t}{12}\|x_{t,i}^+ - x_t\|^3 - \frac{\sigma_0}{24(\gamma_1 + 1)}\|x_t - x_{t-1}\|^3, \tag{2-23}$$

set $i_t := i$ and go to Step 2. Otherwise, set $i := i + 1$ and go to Step 1.1.
Step 2. Set $x_{t+1} := x_{t,i_t}^+$, $\sigma_{t+1} := 2^{i_t-1}\sigma_t$, $B_t := B_{t,i_t}$, $t := t + 1$.
Step 3. If $m > 0$, let $\hat{t} := t$, $B := B_{\hat{t}-1}$ and go to Step 3.1. Otherwise, return to Step 1.
Step 3.1. Find the smallest integer $j \geq 0$ such that $2^{j-1}\sigma_t \geq \sigma_0(m+1)$.
Step 3.2. Compute $x_{t,j}^+$ such that

$$M_{x_t,x_{\hat{t}-1},2^j\sigma_t}(x_{t,j}^+) + \psi(x_{t,j}^+) \leq F(x_t), \quad \left\|\nabla M_{x_t,x_{\hat{t}-1},2^j\sigma_t}(x_{t,j}^+) + \psi'(x_{t,j}^+)\right\| \leq \theta\|x_{t,j}^+ - x_t\|^2,$$

for some $\psi'(x_{t,j}^+) \in \partial\psi(x_{t,j}^+)$, where $M_{x,z,\sigma}(\cdot)$ is as in (2-1).
Step 3.3. Take $\gamma_2 := m$ if $\hat{t} = 1$ and $\gamma_2 := 0$ otherwise. If

$$\begin{aligned} F(x_t) - F(x_{t,j}^+) \geq \frac{2^j\sigma_t}{12}\|x_{t,j}^+ - x_t\|^3 - \frac{\sigma_0}{4(m+1)^2}\|x_t - x_{\hat{t}-1}\|^3 \\ - \frac{\sigma_0}{24(\gamma_2+1)}\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3, \end{aligned} \tag{2-24}$$

set $j_t := j$ and go to Step 3.4. Otherwise, set $j := j+1$ and go to Step 3.2.
Step 3.4. Set $x_{t+1} := x_{t,j_t}^+$, $\sigma_{t+1} := 2^{j_t-1}\sigma_t$ and $t := t+1$.
Step 3.5. If $t \leq \hat{t} + m$, go to Step 3.1; otherwise, go to Step 1.

Remark 2.5. *(i) We mention that a Hessian approximation $B_{t,i}$ satisfying the condition (2-21) can be obtained, for example, by means of finite difference approaches. Indeed, it follows from Lemma 1.13 with*

$$x := x_t \quad \text{and} \quad h = \frac{2\|x_t - x_{t-1}\|}{2^{i-1}},$$

that the approximation $B_{t,i}$ as defined in (1-18) satisfies (2-21) with $\bar{\kappa}_B = \sqrt{n}L$. (ii) Note that $x_{t,i}^+$ as in Step 1.2 is an inexact solution of the problem

$$\min_{y\in Q} M_{x_t,x_t,2^i\sigma_t}(y) + \psi(y).$$

Conditions in (2-22) only require a decrease of the cubic regularized model summed with the function ψ and an approximate first-order stationary point of the above problem. (iii) Note that the sequence of parameters $\{\sigma_t\}$ can be nonmonotone. Indeed, if $i_t = 0$, we have $\sigma_{t+1} = 2^{i_t-1}\sigma_t = \sigma_t/2 \leq \sigma_t$. (iv) Note that both conditions (2-23) and (2-24) allow acceptance of a trial point $x_{t,i}^+$ such that

$$F(x_{t,i}^+) > F(x_t).$$

Consequently, the sequence $\{F(x_t)\}_{t\geq 0}$ may be nonmonotone. (v) If $m = 0$, we obtain an algorithm similar to CNM (Cubic Newton Method) with finite-difference Hessian approximations proposed in [8]. (vi) Whenever we refer to a block, it mean a group of iterations which the same Hessian approximation B is used. For instance, if $x_{-1} \neq x_0 \in \mathbb{R}^n$ are the two inicial points of Algorithm 1, then in the first block we have that B approximates the Hessian on the point x_0. Hence, the same approximation B is used for the subsequent m iterations. In this case, the sequence $\{x_t\}_{t=1}^{m+1}$ corresponds to the first block. Similarly, the sequence $\{x_t\}_{t=m+2}^{2m+2}$ corresponds to the second block, and the Hessian approximation B in the second block is computed at the point x_{m+1}. This process repeats block by block, with the Hessian approximation B updated at the beginning of each new block. (vii) Considering $B_{t,i}$ as described in (i), the main differences between our algorithm and the first-order CNM of [6, Algorithm 2] lie in the choice of the parameter h used for the finite-difference approximations and the timing of Hessian updates. In their method, the parameter h

depends on the precision ϵ and may be smaller than our choice. On the other hand, we obtain a suitable Hessian approximation—possibly updating it multiple times—at the first iteration of each block and retain it for the remaining iterations of that block. In contrast, the method in [6, Algorithm 2] constructs a Hessian approximation at the first iteration of each block but terminates the block's iterations (starting a new block) if a certain acceptance condition is not met.

2.3 Iteration-complexity for Algorithm 1

In the following, we proceed to the complexity analysis of Algorithm 1. We begin by proving that the sequence of parameters $\{\sigma_t\}$ is bounded from above. In particular, we show that the inner procedures in Steps 1.3 and 3.3 end in a finite number of iterations.

Lemma 2.6. *Suppose **A1** holds. Then, the regularization parameters σ_t in Algorithm 1 satisfies*

$$\sigma_0(m+1) \leq \sigma_t \leq \sigma_{max} := \sigma_0(m+1) + 2\left(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{4(m+1)^2}{3\sigma_0}\right)^{\frac{1}{2}} + 4\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{8(m+1)}{\sigma_0}\right)^{\frac{1}{2}}\right), \tag{2-25}$$

for all $t \geq 1$. As a consequence, the inner procedures in Steps 1.3 and 3.3 end in a finite number of iterations.

Proof. Let us prove by induction on t that (2-25) holds. For $t=1$, by Step 1, we obtain $\sigma_0(m+1) \leq 2^{i_0-1}\sigma_0 = \sigma_1$. Now, assume by contradiction that $2^{i_0-1}\sigma_0 = \sigma_1 > \sigma_{max}$. Hence, we have

$$2^{i_0-1}\sigma_0 > 2\left(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{4(m+1)^2}{3\sigma_0}\right)^{\frac{1}{2}} + 4\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{8(m+1)}{\sigma_0}\right)^{\frac{1}{2}}\right)$$

$$> 2\left(L + \sqrt{2}\left(\frac{\bar{\kappa}_B}{2^{i_0-1}}\right)^{\frac{3}{2}}\left(\frac{8(m+1)}{\sigma_0}\right)^{\frac{1}{2}}\right) \tag{2-26}$$

$$> 2\left(L + \sqrt{2}\left(\frac{\bar{\kappa}_B}{2^{i_0-1}}\right)^{\frac{3}{2}}\left(\frac{8}{\sigma_0}\right)^{\frac{1}{2}}\right), \tag{2-27}$$

where we used the fact that $4 > \sqrt{2}/(2^{i_0-1})^{3/2}$ and $i_0 > 0$ in the second inequality. Then, by inequality (2-26) and Proposition 2.1 with $\sigma = 2^{i_0-1}\sigma_0$, $\kappa_B = \bar{\kappa}_B/2^{i_0-1}$, $x^+ = x^+_{0,i}$, $x = x_0$, $\hat{x} = x_{-1}$ and $\bar{\rho} = 8(m+1)/\sigma_0$ it follows that

$$F(x_0) - F(x^+_{0,i}) \geq \frac{2^{i_0-1}\sigma_0}{12}\|x^+_{0,i} - x_0\|^3 - \frac{\sigma_0}{24(m+1)}\|x_0 - x_{-1}\|^3.$$

Therefore, (2-23) is satisfied for $i = i_0 - 1$ and $\gamma_1 = m$, contradicting the minimality of i_0, which proves the inequality in (2-25) for $t = 1$. Now, suppose that (2-25) holds for some natural number $t > 1$, that is, $\sigma_0(m+1) \leq \sigma_t \leq \sigma_{max}$. Let us consider the case that σ_{t+1} is given in Step 3.4 (the proof for the case where σ_{t+1} is given in Step 2 follows with similar arguments). We divide the proof into two cases:
Case ($j_t = 0$): From Step 3.1, we obtain

$$\sigma_0(m+1) \leq \sigma_{t+1} = 2^{j_t - 1} = 2^{0-1}\sigma_t = \frac{1}{2}\sigma_t \leq \sigma_t \leq \sigma_{\max}.$$

Case ($j_t > 0$): From Step 3.1, we have $\sigma_{t+1} = 2^{j_t-1}\sigma_t \geq \sigma_0(m+1)$. Now, assume by contradiction that $\sigma_{t+1} = 2^{j_t-1}\sigma_t > \sigma_{max}$. Hence, we have

$$\begin{aligned} 2^{j_t-1}\sigma_t &> 2\left(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{4(m+1)^2}{3\sigma_0}\right)^{\frac{1}{2}} + 4\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{8(m+1)}{\sigma_0}\right)^{\frac{1}{2}}\right) \\ &> 2\left(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{4(m+1)^2}{3\sigma_0}\right)^{\frac{1}{2}} + \sqrt{2}\left(\frac{\bar{\kappa}_B}{2^{j_t-1}}\right)^{\frac{3}{2}}\left(\frac{8(m+1)}{\sigma_0}\right)^{\frac{1}{2}}\right) \qquad (2\text{-}28) \\ &> 2\left(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{4(m+1)^2}{3\sigma_0}\right)^{\frac{1}{2}} + \sqrt{2}\left(\frac{\bar{\kappa}_B}{2^{j_t-1}}\right)^{\frac{3}{2}}\left(\frac{8}{\sigma_0}\right)^{\frac{1}{2}}\right), \qquad (2\text{-}29) \end{aligned}$$

where we used the fact that $4 > \sqrt{2}/(2^{j_t-1})^{3/2}$ in the last inequality. It follows from inequality (2-29) and Proposition 2.2 with $\sigma = 2^{j_t-1}\sigma_t$, $\kappa_B = \bar{\kappa}_B/2^{j_t-1}$, $x^+ = x^+_{t,j}$, $x = x_t$, $z = x_{\hat{t}-1}$, $\hat{z} = x_{\hat{t}-2}$, $\hat{\rho} = 4(m+1)^2/(3\sigma_0)$ and $\bar{\rho} = 8/\sigma_0$ that

$$F(x_t) - F(x^+_{t,j}) \geq \frac{2^{j_t-1}\sigma_t}{12}\|x^+_{t,j} - x_t\|^3 - \frac{\sigma_0}{4(m+1)^2}\|x_t - x_{\hat{t}-1}\|^3 - \frac{\sigma_0}{24}\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3.$$

Therefore, (2-24) is satisfied for $j = j_t - 1$ and $\gamma_2 = 0$, contradicting the minimality of j_t. On the other hand, from inequality (2-28) and Proposition 2.2 with $\sigma = 2^{j_t-1}\sigma_t$, $\kappa_B = \bar{\kappa}_B/2^{j_t-1}$, $x^+ = x^+_{t,j}$, $x = x_t$, $z = x_{\hat{t}-1}$, $\hat{z} = x_{\hat{t}-2}$, $\hat{\rho} = 4(m+1)^2/(3\sigma_0)$ and $\bar{\rho} = 8(m+1)/\sigma_0$, we get

$$F(x_t) - F(x^+_{t,j}) \geq \frac{2^{j_t-1}\sigma_t}{12}\|x^+_{t,j} - x_t\|^3 - \frac{\sigma_0}{4(m+1)^2}\|x_t - x_{\hat{t}-1}\|^3 - \frac{\sigma_0}{24(m+1)}\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3.$$

Consequently, inequality (2-24) is satisfied for $j = j_t - 1$ and $\gamma_2 = m$, contradicting again the minimality of j_t. So, $\sigma_{t+1} \leq \sigma_{max}$, which concludes the proof. □

Next, we present some key results to establish a global iteration-complexity bound for Algorithm 1.

Lemma 2.7. *Suppose **A1** holds. Let $\{x_t\}_{t=1}^T$ be the sequence generated by Algorithm 1. Let $\tau \in \mathbb{N} - \{0\}$ be the block number associated with the T-th iteration of Algorithm 1, that is, $T = (\tau - 1) * (m+1) + \ell$ with $\ell \in \mathbb{N}$ and $1 \leq \ell \leq m+1$. Then,*

(i) $\sum_{t=(\tau-1)(m+1)+1}^{T-1} \|x_t - x_{(\tau-1)(m+1)}\|^3 \leq \frac{(m+1)^3}{3} \sum_{t=(\tau-1)(m+1)}^{T-1} \|x_{t+1} - x_t\|^3,$

(ii) $F(x_T) \leq F(x_{(\tau-1)(m+1)}) - \frac{\sigma_0(m+1)}{12}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1} - x_t\|^3 + \frac{\sigma_0(m+1)}{24}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3$, *if* $\tau > 1$.

(iii) $F(x_T) \leq F(x_0) - \frac{\sigma_0(m+1)}{12}\sum_{t=0}^{T-1}\|x_{t+1} - x_t\|^3 + \frac{\sigma_0}{24}\|x_0 - x_{-1}\|^3$, *if* $\tau = 1$.

Proof. (i) Using the triangle inequality, we have

$$\sum_{t=(\tau-1)(m+1)+1}^{T-1}\|x_t - x_{(\tau-1)(m+1)}\|^3 \leq \sum_{t=(\tau-1)(m+1)+1}^{T-1}\left(\sum_{i=(\tau-1)(m+1)+1}^{t}\|x_i - x_{i-1}\|\right)^3$$
$$\leq \sum_{\bar{t}=1}^{\ell-1}\left(\sum_{\bar{i}=1}^{\bar{t}}\|x_{\bar{i}+(\tau-1)(m+1)} - x_{\bar{i}+(\tau-1)(m+1)-1}\|\right)^3,$$

where we used the variable changes $\bar{t} := t - (\tau-1)(m+1)$ and $\bar{i} := i - (\tau-1)(m+1)$ in the last inequality. Then, from last inequality, (1-6), $\ell \leq m+1$ and $\bar{t} := t-(\tau-1)(m+1)$, we obtain

$$\sum_{t=(\tau-1)(m+1)+1}^{T-1}\|x_t - x_{(\tau-1)(m+1)}\|^3 \leq \frac{\ell^3}{3}\sum_{\bar{t}=1}^{\ell-1}\|x_{\bar{t}+(\tau-1)(m+1)} - x_{\bar{t}+(\tau-1)(m+1)-1}\|^3$$
$$\leq \frac{(m+1)^3}{3}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1} - x_t\|^3,$$

which proves the inequality in (i).

(ii) For $1 \leq \ell \leq m+1$, consider the sequence $\{x_t\}_{(\tau-1)(m+1)+1}^{(\tau-1)(m+1)+l}$ associated to the block $\tau > 1$. It follows from Algorithm 1, that the $((\tau-1)(m+1)+1)$-th iteration satisfies (2-23), whereas the $(\ell-1)$-th consecutive ones satisfy (2-24). Hence,

$$F(x_{(\tau-1)(m+1)+1}) \leq F(x_{(\tau-1)(m+1)}) - \frac{\sigma_{(\tau-1)(m+1)+1}}{6}\|x_{(\tau-1)(m+1)+1} - x_{(\tau-1)(m+1)}\|^3$$
$$+ \frac{\sigma_0}{24}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3,$$

$$F(x_{(\tau-1)(m+1)+2}) \leq F(x_{(\tau-1)(m+1)+1}) - \frac{\sigma_{(\tau-1)(m+1)+2}}{6}\|x_{(\tau-1)(m+1)+2} - x_{(\tau-1)(m+1)+1}\|^3$$
$$+ \frac{\sigma_0}{4(m+1)^2}\|x_{(\tau-1)(m+1)+1} - x_{(\tau-1)(m+1)}\|^3 + \frac{\sigma_0}{24}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3,$$

$$F(x_{(\tau-1)(m+1)+3}) \leq F(x_{(\tau-1)(m+1)+2}) - \frac{\sigma_{(\tau-1)(m+1)+3}}{6}\|x_{(\tau-1)(m+1)+3} - x_{(\tau-1)(m+1)+2}\|^3$$
$$+ \frac{\sigma_0}{4(m+1)^2}\|x_{(\tau-1)(m+1)+2} - x_{(\tau-1)(m+1)}\|^3 + \frac{\sigma_0}{24}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3,$$

$$\vdots$$

$$F(x_{(\tau-1)(m+1)+l}) \leq F(x_{(\tau-1)(m+1)+(l-1)}) - \frac{\sigma_{(\tau-1)(m+1)+l}}{6}\|x_{(\tau-1)(m+1)+l} - x_{(\tau-1)(m+1)+(l-1)}\|^3$$
$$+ \frac{\sigma_0}{4(m+1)^2}\|x_{(\tau-1)(m+1)+(l-1)} - x_{(\tau-1)(m+1)}\|^3 + \frac{\sigma_0}{24}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3.$$

Combining the above inequalities with (2-25), we get

$$F(x_T) \leq F(x_{(\tau-1)(m+1)}) - \frac{\sigma_0(m+1)}{6}\sum_{t=(\tau-1)(m+1)}^{(\tau-1)(m+1)+l-1}\|x_{t+1} - x_t\|^3$$
$$+ \frac{\sigma_0}{4(m+1)^2}\sum_{t=(\tau-1)(m+1)+1}^{(\tau-1)(m+1)+l-1}\|x_t - x_{(\tau-1)(m+1)}\|^3 + \frac{\ell\sigma_0\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3}{24},$$

which, combined with $T = (\tau-1)(m+1) + \ell$ and $\ell \leq m+1$, yields

$$F(x_T) \leq F(x_{(\tau-1)(m+1)}) + \frac{\sigma_0}{4(m+1)^2}\sum_{t=(\tau-1)(m+1)+1}^{T-1}\|x_t - x_{(\tau-1)(m+1)}\|^3$$
$$- \frac{\sigma_0(m+1)}{6}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1} - x_t\|^3 + \frac{\sigma_0(m+1)\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3}{24}.$$

Combining the last inequality with (i), we obtain the desired result. Analogously, we obtain the result in (iii). □

As a consequence of the last lemma, we obtain the following bound to the sum of the sequence $\{\|x_{t+1} - x_t\|^3\}$.

Lemma 2.8. *Suppose **A1** and **A2** hold. Let $\{x_t\}_{t=1}^T$ be the sequence generated by Algorithm 1. Let $\tau \in \mathbb{N} - \{0\}$ be the block number associated with the T-th iteration as defined in Lemma 2.7. Then,*

$$\sum_{t=0}^{T-1}\|x_{t+1} - x_t\|^3 \leq \frac{1}{m+1}\left(\frac{24(F(x_0) - F^\star)}{\sigma_0} + \|x_0 - x_{-1}\|^3\right).$$

Proof. It follows from Lemma 2.7(ii) for multiples values of τ and Lemma 2.7(iii) that

$$F(x_{m+1}) \leq F(x_0) - \frac{\sigma_0(m+1)}{12}\sum_{t=0}^{m}\|x_{t+1} - x_t\|^3 + \frac{\sigma_0}{24}\|x_0 - x_{-1}\|^3,$$

$$F(x_{2(m+1)}) \leq F(x_{m+1}) - \frac{\sigma_0(m+1)}{12}\sum_{t=m+1}^{2m+1}\|x_{t+1} - x_t\|^3 + \frac{\sigma_0(m+1)}{24}\|x_{m+1} - x_m\|^3,$$

$$F(x_{3(m+1)}) \leq F(x_{2(m+1)}) - \frac{\sigma_0(m+1)}{12} \sum_{t=2m+2}^{3m+2} \|x_{t+1} - x_t\|^3 + \frac{\sigma_0(m+1)}{24} \|x_{2(m+1)} - x_{2(m+1)-1}\|^3.$$

$$\vdots$$

$$F(x_T) \leq F(x_{(\tau-1)(m+1)}) - \frac{\sigma_0(m+1)}{12} \sum_{t=(\tau-1)(m+1)}^{T-1} \|x_{t+1} - x_t\|^3 + \frac{\sigma_0(m+1)}{24} \|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3.$$

Combining the above inequalities, we obtain

$$F(x_T) \leq F(x_0) - \frac{\sigma_0(m+1)}{12} \sum_{t=0}^{T-1} \|x_{t+1} - x_t\|^3 + \frac{\sigma_0(m+1)}{24} \sum_{k=1}^{\tau-1} \|x_{k(m+1)} - x_{k(m+1)-1}\|^3 + \frac{\sigma_0 \|x_0 - x_{-1}\|^3}{24},$$

where $T = (\tau-1)(m+1) + \ell$ with $\ell \in \mathbb{N}$ and $1 \leq \ell \leq m+1$. Hence, using that $F(x_T) > F^\star$ and some algebraic manipulations, we have

$$F^\star \leq F(x_0) - \frac{\sigma_0(m+1)}{24} \sum_{t=0}^{T-1} \|x_{t+1} - x_t\|^3 + \frac{\sigma_0}{24} \|x_0 - x_{-1}\|^3,$$

which implies the desired inequality. □

The following lemma gives us a key relation in the analysis of the algorithm complexity.

Lemma 2.9. *Suppose **A1** holds. Let $\{x_t\}_{t=1}^{T}$ be the sequence generated by Algorithm 1. Let $\tau \in \mathbb{N} - \{0\}$ be the block number associated with the T-th iteration as defined in Lemma 2.7. Then,*

$$(\sqrt{3})^{-1} \sum_{t=(\tau-1)(m+1)}^{T-1} \|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\tilde{\lambda}^{\frac{3}{2}} + \frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}} \right) \times \sum_{t=(\tau-1)(m+1)}^{T-1} \|x_{t+1} - x_t\|^3 + \|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3, \tag{2-30}$$

where $\tilde{\lambda} := \sigma_{max} + (L + L^2(m+1))/2 + \theta + 2(m+1)^{2/3}\bar{\kappa}_B^2$.

Proof. For $1 \leq \ell \leq m+1$, let $\{x_t\}_{(\tau-1)(m+1)+1}^{(\tau-1)(m+1)+\ell}$ be the sequence associated with the block τ. Consider first $t = (\tau-1)(m+1)$. Since $x_{(\tau-1)(m+1)+1}$ is the first iteration of this block, it follows from Algorithm 1 and Proposition 2.4 with $\sigma = 2^{i_t}\sigma_t$, $x^+ = x_{t+1}$, $x = x_t$, $z = x_t$,

$\hat{z} = x_{t-1}$, $\kappa_B = \bar{\kappa}_B/2^{i_t-1}$, $\rho = m+1$ and $\rho^\star = (m+1)^{2/3}$ that

$$(\sqrt{3})^{-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\frac{2^{i_t}\sigma_t + L + L^2(m+1)}{2} + \theta + \frac{(m+1)^{\frac{2}{3}}}{2}\left(\frac{\bar{\kappa}_B}{2^{i_t-1}}\right)^2\right)^{\frac{3}{2}}$$
$$\times \|x_{t+1} - x_t\|^3 + \frac{\|x_t - x_{t-1}\|^3}{2^{\frac{3}{2}}(m+1)},$$

which, combined with $2^{i_t-1}\sigma_t = \sigma_{t+1} \leq \sigma_{max}$ (see Step 2 of Algorithm 1 and (2-25)) and $i_t \geq 0$, yields

$$(\sqrt{3})^{-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \tilde{\lambda}^{\frac{3}{2}}\|x_{t+1} - x_t\|^3 + \frac{\|x_t - x_{t-1}\|^3}{2^{\frac{3}{2}}(m+1)}. \tag{2-31}$$

Let us now consider the case where x_{t+1} corresponds to the remaining iterations of the sequence, i.e., t satisfies $\hat{t} := (\tau-1)(m+1)+1 \leq t \leq (\tau-1)(m+1)+\ell-1$. Hence, from Proposition 2.4 with $\sigma = 2^{j_t}\sigma_t$, $x^+ = x_{t+1}$, $x = x_t$, $z = x_{\hat{t}-1}$, $\hat{z} = x_{\hat{t}-2}$, $\kappa_B = \bar{\kappa}_B/2^{j_t-1}$, $\rho = m+1$ and $\rho^\star = (m+1)^{2/3}$ it holds that

$$(\sqrt{3})^{-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\frac{2^{j_t}\sigma_t + L + L^2(m+1)}{2} + \frac{(m+1)^{\frac{2}{3}}}{2}\left(\frac{\bar{\kappa}_B}{2^{j_t-1}}\right)^2 + \theta\right)^{\frac{3}{2}}$$
$$\times \|x_{t+1} - x_t\|^3 + \frac{\|x_t - x_{\hat{t}-1}\|^3}{2^{\frac{3}{2}}(m+1)^{\frac{3}{2}}} + \frac{\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3}{2^{\frac{3}{2}}(m+1)}.$$

Since $2^{j_t-1}\sigma_t = \sigma_{t+1} \leq \sigma_{max}$ (see Step 3.4 of Algorithm 1 and (2-25)) and $j_t \geq 0$, we get

$$(\sqrt{3})^{-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \tilde{\lambda}^{\frac{3}{2}}\|x_{t+1} - x_t\|^3 + \frac{\|x_t - x_{\hat{t}-1}\|^3}{2^{\frac{3}{2}}(m+1)^{\frac{3}{2}}} + \frac{\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3}{2^{\frac{3}{2}}(m+1)},$$

for all $\hat{t} \leq t \leq (\tau-1)(m+1)+\ell-1$. Combining the last inequalities and (2-31), we get

$$(\sqrt{3})^{-1}\sum_{t=(\tau-1)(m+1)}^{(\tau-1)(m+1)+\ell-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \tilde{\lambda}^{\frac{3}{2}}\sum_{t=(\tau-1)(m+1)}^{(\tau-1)(m+1)+\ell-1}\|x_{t+1} - x_t\|^3$$
$$+ \frac{1}{2^{\frac{3}{2}}(m+1)^{\frac{3}{2}}}\sum_{t=(\tau-1)(m+1)+1}^{(\tau-1)(m+1)+\ell-1}\|x_t - x_{(\tau-1)(m+1)}\|^3 + \frac{\ell}{2^{\frac{3}{2}}(m+1)}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3.$$

Since $T = (\tau-1)(m+1)+\ell$ and $\ell \leq m+1$, we obtain

$$(\sqrt{3})^{-1}\sum_{t=(\tau-1)(m+1)}^{T-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \tilde{\lambda}^{\frac{3}{2}}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1} - x_t\|^3$$
$$+ \frac{1}{2^{\frac{3}{2}}(m+1)^{\frac{3}{2}}}\sum_{t=(\tau-1)(m+1)+1}^{T-1}\|x_t - x_{(\tau-1)(m+1)}\|^3 + \frac{\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3}{2^{\frac{3}{2}}},$$

which, combined with Lemma 2.7(i), yields

$$(\sqrt{3})^{-1}\sum_{t=(\tau-1)(m+1)}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \left(\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)$$
$$\times \sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3+\frac{\|x_{(\tau-1)(m+1)}-x_{(\tau-1)(m+1)-1}\|^3}{2^{\frac{3}{2}}},$$

which implies the desired result. □

We are now ready to present an iteration complexity bound for the Algorithm 1 in terms of the outer iteration number.

Theorem 2.10. *Suppose* ***A1*** *and* ***A2*** *hold. Let $\{x_t\}_{t=1}^{T}$ be the sequence generated by Algorithm 1. Let $\tau\in\mathbb{N}-\{0\}$ be the block number associated with the T-th iteration as defined in Lemma 2.7. Then,*

$$\sum_{t=0}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \sqrt{3}\left(1+\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\lambda+\sqrt{3}\|x_0-x_{-1}\|^3, \tag{2-32}$$

where $\tilde{\lambda}$ is as in Lemma 2.9, and $\lambda := (24(F(x_0)-F^\star)/\sigma_0+\|x_0-x_{-1}\|^3)/(m+1)$. As a consequence, given $\epsilon>0$, Algorithm 1 needs at most $\mathcal{O}\left(m^{1/2}\epsilon^{-3/2}\right)$ iterations to generate an $\epsilon-$approximate critical point for problem (0-3).

Proof. It follows from inequality (2-30) for multiples values of τ that

$$(\sqrt{3})^{-1}\sum_{t=0}^{m}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \left(\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\sum_{t=0}^{m}\|x_{t+1}-x_t\|^3$$
$$+\|x_0-x_{-1}\|^3,$$

$$(\sqrt{3})^{-1}\sum_{t=m+1}^{2m+1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \left(\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\sum_{t=m+1}^{2m+1}\|x_{t+1}-x_t\|^3$$
$$+\|x_{m+1}-x_m\|^3,$$

$$(\sqrt{3})^{-1}\sum_{t=2(m+1)}^{3m+2}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \left(\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\sum_{t=2(m+1)}^{3m+2}\|x_{t+1}-x_t\|^3$$
$$+\|x_{2(m+1)}-x_{2(m+1)-1}\|^3,$$

$$\vdots$$

$$(\sqrt{3})^{-1}\sum_{t=(\tau-1)(m+1)}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3 + \|x_{(\tau-1)(m+1)}-x_{(\tau-1)(m+1)-1}\|^3.$$

Combining the last inequalities, we get

$$(\sqrt{3})^{-1}\sum_{t=0}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3 + \sum_{k=0}^{\tau-1}\|x_{k(m+1)}-x_{k(m+1)-1}\|^3.$$

where $T=(\tau-1)(m+1)+\ell$ with $\ell\in\mathbb{N}$ and $1\leq\ell\leq m+1$. From last inequality and some algebraic manipulation, we have

$$(\sqrt{3})^{-1}\sum_{t=0}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(1+\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3 + \|x_0-x_{-1}\|^3.$$

Hence, using Lemma 2.8, we obtain

$$\sum_{t=0}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \leq \sqrt{3}\left(1+\tilde{\lambda}^{\frac{3}{2}}+\frac{(m+1)^{\frac{3}{2}}}{3(2)^{\frac{3}{2}}}\right)\lambda+\sqrt{3}\|x_0-x_{-1}\|^3,$$

which concludes the first result. Now, note that the inequality in (2-32), along with the definitions of σ_{max}, λ and $\tilde{\lambda}$, implies the second result of the theorem. □

The next theorem gives us a complexity on the number of functions and gradient evaluations of Algorithm 1.

Theorem 2.11. *Suppose **A1** and **A2** hold. Consider that Algorithm 1 is implemented such that $B_{t,i}$ in Step 1.1 is computed as in (1-15). Then, the number of gradient and function evaluations up to the T-th iteration, $FGE(T)$, is bounded as follows:*

$$FGE(T) \leq \frac{(n+m+2)(T+m)}{m+1}\left(2+\log_2\frac{\sigma_{max}}{\sigma_0}\right)+\log_2\frac{\sigma_{max}}{\sigma_0(m+1)}. \tag{2-33}$$

As a consequence, given $\epsilon>0$, the total number of FGE required to generate an ϵ-approximate critical point is $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2}+(n+m)\right)$.

Proof. Consider an arbitrary block τ. Hence, using the notation in Lemma 2.7, that is, $T=(\tau-1)(m+1)+\ell$ with $\ell\in\mathbb{N}$, $1\leq\ell\leq m+1$, it follows from Algorithm 1 that the number of functions and gradient evaluations for the first and for the other iterations ($t\in\{(\tau-1)(m+1)+1,(\tau-1)(m+1)+2,\ldots,(\tau-1)(m+1)+l-1\}$) of block τ are bounded, respectively, by

$$(i_{(\tau-1)(m+1)}+1)(n+1), \quad (j_t+1), \tag{2-34}$$

if τ is not the first block and

$$2+(i_{(\tau-1)(m+1)}+1)(n+1), \quad (j_t+1),$$

otherwise. From Steps 2 and 3.4 of Algorithm 1, we have $2^{i_t-1}\sigma_t=\sigma_{t+1}$ and $2^{j_t-1}\sigma_t=\sigma_{t+1}$, which implies that

$$i_{(\tau-1)(m+1)}+1=\log_2\sigma_{(\tau-1)(m+1)+1}-\log_2\sigma_{(\tau-1)(m+1)}+2,$$

and $j_t+1=\log_2\sigma_{t+1}-\log_2\sigma_t+2$ for all $(\tau-1)(m+1)+1\le t\le T-1$. Combining the last equalities with (2-34), we have

$$\begin{aligned}
&(i_{(\tau-1)(m+1)}+1)(n+1)+\sum_{t=(\tau-1)(m+1)+1}^{T-1}(j_t+1)\\
&=(n+1)(\log_2\sigma_{(\tau-1)(m+1)+1}-\log_2\sigma_{(\tau-1)(m+1)}+2)+\sum_{t=(\tau-1)(m+1)+1}^{T-1}(\log_2\sigma_{t+1}-\log_2\sigma_t+2)\\
&\le 2(n+m+2)+(n+1)(\log_2\sigma_{(\tau-1)(m+1)+1}-\log_2\sigma_{(\tau-1)(m+1)})+\log_2\sigma_T-\log_2\sigma_{(\tau-1)(m+1)+1}.
\end{aligned}$$

Applying the last inequalities for multiples values of τ, we obtain

$$\begin{aligned}
2+(i_0+1)(n+1)+\sum_{t=1}^{m}(j_t+1)&=2(n+m+2)+(n+1)(\log_2\sigma_1-\log_2\sigma_0)\\
&\quad+\log_2\sigma_{m+1}-\log_2\sigma_1,
\end{aligned}$$

$$\begin{aligned}
(i_{m+1}+1)(n+1)+\sum_{t=m+2}^{2m+1}(j_t+1)&\le 2(n+m+2)+(n+1)(\log_2\sigma_{m+2}-\log_2\sigma_{m+1})\\
&\quad+\log_2\sigma_{2(m+1)}-\log_2\sigma_{m+2},
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
&(i_{(\tau-1)(m+1)}+1)(n+1)+\sum_{t=(\tau-1)(m+1)+1}^{T-1}(j_t+1)\le 2(n+m+2)\\
&+(n+1)(\log_2\sigma_{(\tau-1)(m+1)+1}-\log_2\sigma_{(\tau-1)(m+1)})+\log_2\sigma_T-\log_2\sigma_{(\tau-1)(m+1)+1}.
\end{aligned}$$

Combining the last inequalities and (2-25), we obtain

$$FGE(T) \leq 2\tau(n+m+2) + (n+2)\sum_{k=0}^{\tau-1} \log_2 \frac{\sigma_{k(m+1)+1}}{\sigma_{k(m+1)}} + \log_2 \frac{\sigma_{max}}{\sigma_0(m+1)}.$$

which, combined with $\sigma_t \geq \sigma_0$, implies that

$$FGE(T) \leq \frac{2(T+m)(n+m+2)}{m+1} + \frac{(T+m)(n+2)}{m+1} \log_2 \frac{\sigma_{max}}{\sigma_0} + \log_2 \frac{\sigma_{max}}{\sigma_0(m+1)},$$

where in the last inequality we used $\tau \leq (T+m)/(m+1)$, which implies the desired inequality. On the other hand, the second result follows from (2-33) and the complexity on the term T. □

Remark 2.12. *(i) If $B_{t,i}$ is updated as in (1-16), it can be shown that the statement of the last theorem still holds, however with a slight modification:*

$$FGE(T) \leq \frac{(2n+m+2)(T+m)}{m+1}\left(2 + \log_2 \frac{\sigma_{\max}}{\sigma_0}\right) + \log_2 \frac{\sigma_{\max}}{\sigma_0(m+1)}.$$

(ii) Comparing the results in Theorem 2.11 with the ones in [6, Algorithm 2], we see that the bound in (2-33) is similar, in terms of n and m, to the one in [6, Lemma 4.4]. At the same time, their bound on the total number of function and gradient evaluations to generate an ϵ-approximate critical point, is $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2} + (n+m)\right)$, so the complexity in Theorem 2.11 is maintained. (iii) Taking $m = n$ in Theorem 2.11 we get a complexity of $\mathcal{O}\left(n^{1/2}\epsilon^{-3/2} + n\right)$ to the number of function and gradient evaluations, which is a reduction by a factor of $n^{1/2}$ in comparison with the CNM proposed in [8].

The next section establishes an algorithm that finds an approximate solution to the problem (0-3) when both the gradient and Hessian are not known.

CHAPTER 3

Cubic Regularization Method with Inexact Gradient and Lazy Hessian approximations

In this chapter, we present our second algorithm for solving (0-3). It consist of a variant of the cubic regularization method in which the gradient and Hessian of the function f is assumed to be unavailable and both the Hessian and gradient are approximated such that some conditions are satisfied.

The Chapters are organized as follows. Section 3.1 introduces some essential auxiliary results. Section 3.2 presents the second algorithm. And finally, Section 3.3 shows iterations complexity results for the second algorithm.

3.1 Auxiliary Results

The subproblem of the proposed algorithm will be associated to the following cubic model

$$M^g_{x,z,\sigma}(y) = f(x) + \langle g(x), y - x\rangle + \frac{1}{2}\langle B(z)(y - x), y - x\rangle + \frac{\sigma}{6}\|y - x\|^3, \tag{3-1}$$

where $x, z \in \mathbb{R}^n$, $\sigma > 0$, $g(x)$ is an approximation to the gradient of f on x, and $B(z)$ is an approximation to the Hessian of f on the point z. Note that the approximation B constructed by (1-23) uses n times more evaluations of f compared to the approximation g in (1-9). Therefore, computing the Hessian and maintaining it fixed for $m \geq 0$ iterations, while we update the gradient approximation g every iteration, might reduce the computational cost in terms of f.

Under some assumptions on $M^g_{x,z,\sigma}(\cdot)$, $g(\cdot)$, $B(\cdot)$ and σ, we can establish a bound on the term $F(x) - F(x^+)$, which follows from this next proposition.

Proposition 3.1. *Suppose **A1** holds and assume that $x^+ \in \mathbb{R}^n$ satisfies*

$$M^g_{x,x,\sigma}(x^+) + \psi(x^+) \leq F(x), \tag{3-2}$$

for some $x \in \mathbb{R}^n$ and $\sigma > 0$. Moreover, suppose that for some $\kappa_g, \kappa_B \geq 0$ and $\hat{x} \in \mathbb{R}^n$, it *holds that*

$$\|g(x) - \nabla f(x)\| \leq \kappa_g \|x - \hat{x}\|^2 \quad \textit{and} \quad \|B(x) - \nabla^2 f(x)\| \leq \kappa_B \|x - \hat{x}\|. \tag{3-3}$$

If

$$\sigma \geq 2(L + \sqrt{6}\kappa_B^{\frac{3}{2}} \bar{\rho}^{\frac{1}{2}} + 18\bar{\rho}^2 \kappa_g^3), \tag{3-4}$$

for some $\bar{\rho} > 0$, *then*

$$F(x) - F(x^+) \geq \frac{\sigma}{12}\|x^+ - x\|^3 - \frac{1}{3\bar{\rho}}\|x - \hat{x}\|^3.$$

Proof. From inequality (1-3) and definition of $M^g_{x,z,\sigma}(\cdot)$ in (3-1), we have

$$\begin{aligned} f(x^+) &\leq f(x) + \langle \nabla f(x), x^+ - x\rangle + \frac{1}{2}\langle \nabla^2 f(x)(x^+ - x), x^+ - x\rangle + \frac{L}{6}\|x^+ - x\|^3 \\ &\leq M^g_{x,x,\sigma}(x^+) + \langle \nabla f(x) - g(x), x^+ - x\rangle + \frac{1}{2}\langle (\nabla^2 f(x) - B(x))(x^+ - x), x^+ - x\rangle \\ &+ \frac{L - \sigma}{6}\|x^+ - x\|^3. \end{aligned}$$

From (3-2), (3-3), $F = f + \psi$ and Cauchy-Schwarz inequality, we get

$$\begin{aligned} F(x^+) &\leq F(x) + \|g(x) - \nabla f(x)\|\|x^+ - x\| + \frac{1}{2}\|B(x) - \nabla^2 f(x)\|\|x^+ - x\|^2 \\ &+ \frac{L-\sigma}{6}\|x^+ - x\|^3 \\ &\leq F(x) + \kappa_g\|x - \hat{x}\|^2\|x^+ - x\| + \frac{\kappa_B}{2}\|x - \hat{x}\|\|x^+ - x\|^2 + \frac{L-\sigma}{6}\|x^+ - x\|^3 \\ &= F(x) + \frac{\|x - \hat{x}\|^2}{(3\bar{\rho})^{\frac{2}{3}}}\kappa_g(3\bar{\rho})^{\frac{2}{3}}\|x^+ - x\| + \frac{\|x - \hat{x}\|}{(3\bar{\rho})^{\frac{1}{3}}}\frac{\kappa_B(3\bar{\rho})^{\frac{1}{3}}\|x^+ - x\|^2}{2} \\ &+ \frac{L-\sigma}{6}\|x^+ - x\|^3. \end{aligned} \tag{3-5}$$

Using Proposition 1.10 with

$$a = \frac{\|x - \hat{x}\|^2}{(3\bar{\rho})^{\frac{2}{3}}} \quad b = \kappa_g(3\bar{\rho})^{\frac{2}{3}}\|x^+ - x\| \quad p = \frac{3}{2} \quad \text{and} \quad q = 3,$$

we get

$$\frac{\|x - \hat{x}\|^2}{(3\bar{\rho})^{\frac{2}{3}}}\kappa_g(3\bar{\rho})^{\frac{2}{3}}\|x^+ - x\| \leq \frac{2\|x - \hat{x}\|^3}{9\bar{\rho}} + \frac{9\bar{\rho}^2\|x^+ - x\|^3\kappa_g^3}{3}. \tag{3-6}$$

Again, by Proposition 1.10 with

$$a = \frac{\|x - \hat{x}\|}{(3\bar{\rho})^{\frac{1}{3}}} \quad b = \frac{\kappa_B(3\bar{\rho})^{\frac{1}{3}}\|x^+ - x\|^2}{2} \quad p = 3 \quad \text{and} \quad q = \frac{3}{2},$$

we have

$$\frac{\|x-\hat{x}\|}{(3\bar{\rho})^{\frac{1}{3}}}\frac{\kappa_B(3\bar{\rho})^{1/3}\|x^+-x\|^2}{2} \leq \frac{\sqrt{2}\kappa_B^{\frac{3}{2}}(3\bar{\rho})^{\frac{1}{2}}\|x^+-x\|^3}{6} + \frac{\|x-\hat{x}\|^3}{9\bar{\rho}}. \tag{3-7}$$

Combining (3-5), (3-6) and (3-7), it follows that

$$F(x^+) \leq F(x) + \left(\frac{\sqrt{6}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}} + 18\bar{\rho}^2\kappa_g^3 + L - \sigma}{6}\right)\|x^+-x\|^3 + \frac{\|x-\hat{x}\|^3}{3\bar{\rho}},$$

From last inequality and (3-4), we have

$$F(x^+) \leq F(x) - \frac{\sigma}{12}\|x^+-x\|^3 + \frac{1}{3\bar{\rho}}\|x-\hat{x}\|^3,$$

which implies the desired result. □

The next proposition gives another estimate for $F(x) - F(x^+)$. In the context of the algorithm implementation, the approximation B of the Hessian that we made on $x \in \mathbb{R}^n$ is reused for the next $m \geq 0$ fixed iterations.

Proposition 3.2. *Suppose **A1** holds and assume that $x^+ \in \mathbb{R}^n$ satisfies*

$$M^g_{x,z,\sigma}(x^+) + \psi(x^+) \leq F(x), \tag{3-8}$$

for some $x, z \in \mathbb{R}^n$ and $\sigma > 0$. Moreover, suppose that for some $\kappa_g, \kappa_B \geq 0$ and $\hat{x}, \hat{z} \in \mathbb{R}^n$, it holds that

$$\|g(x) - \nabla f(x)\| \leq \kappa_g\|x-\hat{x}\|^2 \quad \textit{and} \quad \|B(z) - \nabla^2 f(z)\| \leq \kappa_B\|z-\hat{z}\|. \tag{3-9}$$

If

$$\sigma > 2(L + \sqrt{2}L^{\frac{3}{2}}\hat{\rho}^{\frac{1}{2}} + \sqrt{2}\kappa_B^{\frac{3}{2}}\bar{\rho}^{\frac{1}{2}} + +2\kappa_g^3\tilde{\rho}^2), \tag{3-10}$$

for some $\hat{\rho}, \bar{\rho}, \tilde{\rho} > 0$, then

$$F(x) - F(x^+) \geq \frac{\sigma}{12}\|x^+-x\|^3 - \frac{2}{3\tilde{\rho}}\|x-\hat{x}\|^3 - \frac{1}{3\hat{\rho}}\|x-z\|^3 - \frac{1}{3\bar{\rho}}\|z-\hat{z}\|^3. \tag{3-11}$$

Proof. From inequality (1-3) and definition of $M^g_{x,z,\sigma}(\cdot)$ in (3-1), we get

$$\begin{aligned} f(x^+) &\leq f(x) + \langle\nabla f(x), x^+-x\rangle + \frac{1}{2}\langle\nabla^2 f(x)(x^+-x), x^+-x\rangle + \frac{L}{6}\|x^+-x\|^3 \\ &\leq M^g_{x,z,\sigma}(x^+) + \|g(x)-\nabla f(x)\|\|x^+-x\| + \frac{1}{2}\|\nabla^2 f(x) - B(z)\|\|x^+-x\|^2 \\ &+ \frac{L-\sigma}{6}\|x^+-x\|^3. \end{aligned}$$

Combining the last inequality with (3-8), (3-9), $F = f + \psi$ and Cauchy-Schwarz inequality, we obtain

$$F(x^+) \leq F(x) + \kappa_g \|x - \hat{x}\|^2 \|x^+ - x\| + \frac{L}{2}\|x - z\|\|x^+ - x\|^2 + \frac{\kappa_B}{2}\|z - \hat{z}\|\|x^+ - x\|^2 + \frac{L - \sigma}{6}\|x^+ - x\|^3.$$

From Proposition 1.10 with

$$a = \frac{\|x - \hat{x}\|^2}{\tilde{\rho}^{\frac{2}{3}}} \quad b = \kappa_g \tilde{\rho}^{\frac{2}{3}}\|x^+ - x\| \quad p = \frac{3}{2} \quad \text{and} \quad q = 3,$$

we have

$$\kappa_g \|x - \hat{x}\|^2 \|x^+ - x\| = \frac{\|x - \hat{x}\|^2}{\tilde{\rho}^{\frac{2}{3}}} \kappa_g \tilde{\rho}^{\frac{2}{3}} \|x^+ - x\| \leq \frac{2\|x - \hat{x}\|^3}{3\tilde{\rho}} + \frac{\kappa_g^3 \tilde{\rho}^2 \|x^+ - x\|^3}{3}. \tag{3-12}$$

Hence, by inequalities (2-12), (2-13) and (3-12), we have

$$F(x^+) \leq F(x) + \left(\frac{\sqrt{2}\kappa_B^{\frac{3}{2}} \bar{\rho}^{\frac{1}{2}} + \sqrt{2} L^{\frac{3}{2}} \hat{\rho}^{\frac{1}{2}} + 2\kappa_g^3 \tilde{\rho}^2 + L - \sigma}{6} \right) \|x^+ - x\|^3 + \frac{2}{3\tilde{\rho}}\|x - \hat{x}\|^3 + \frac{1}{3\hat{\rho}}\|x - z\|^3 + \frac{1}{3\bar{\rho}}\|z - \hat{z}\|^3.$$

From (3-10), it follows the desired result in (3-11). □

Combining the Lemmas 1.11 and 1.14, with the Proposition 3.2, we can find h_g, h_B associated with the approximations g and B respectively and σ such that the inequality (3-11) holds.

Theorem 3.3. *Suppose* ***A1*** *holds. Let $x, \hat{x}, z, \hat{z} \in \mathbb{R}^n$, $\kappa_B, \kappa_g \geq 0$ and $\sigma > 0$. Consider $M^g_{x,z,\sigma}(\cdot)$ as in (3-1) where the matrix $B(z) \in \mathbb{R}^{n\times n}$ is defined by*

$$B(z) = \frac{1}{2}(A(z) + A^\top(z)),$$

and A is as in (1-23) with $x = z$, and g as defined in (1-7), with

$$h_g \leq \left(\frac{6\kappa_g}{\sqrt{n}\sigma}\|x - \hat{x}\|^2 \right)^{\frac{1}{2}} \quad \text{and} \quad h_B \leq \frac{3\kappa_B}{5n\sigma}\|z - \hat{z}\|,$$

where h_g and h_B are associated with the approximations g and B respectively. Assume also that $x^+ \in \mathbb{R}^n$ satisfies (3-8). If

$$\sigma > 2(L + \sqrt{2}L^{\frac{3}{2}} \hat{\rho}^{\frac{1}{2}} + \sqrt{2}\kappa_B^{\frac{3}{2}} \bar{\rho}^{\frac{1}{2}} + 2\kappa_g^3 \tilde{\rho}^2), \tag{3-13}$$

for some $\hat{\rho}, \bar{\rho}, \tilde{\rho} > 0$, *then*

$$F(x) - F(x^+) \geq \frac{\sigma}{12}\|x^+ - x\|^3 - \frac{2}{3\tilde{\rho}}\|x - \hat{x}\|^3 - \frac{1}{3\hat{\rho}}\|x - z\|^3 - \frac{1}{3\bar{\rho}}\|z - \hat{z}\|^3.$$

Proof. It follows from Lemma 1.14 with $x = z$ and (1-25) that

$$\|B(z) - \nabla^2 f(z)\| \leq \frac{5nL}{3}h_B \leq \frac{5nL}{3}\frac{3\kappa_B}{5n\sigma}\|z - \hat{z}\| = \frac{\kappa_B L}{\sigma}\|z - \hat{z}\|. \tag{3-14}$$

On the other hand, from Lemma 1.11 and (1-8), we have

$$\|g(x) - \nabla f(x)\| \leq \frac{\sqrt{n}L}{6}h_g^2 \leq \frac{\sqrt{n}L}{6}\frac{6\kappa_g}{\sqrt{n}\sigma}\|x - \hat{x}\|^2 \leq \frac{\kappa_g L}{\sigma}\|x - \hat{x}\|^2. \tag{3-15}$$

The inequalities (3-14) and (3-15), combined with the fact that (3-13) implies $L/\sigma \leq 1$, yields

$$\|B(z) - \nabla^2 f(z)\| \leq \kappa_B\|z - \hat{z}\| \quad \text{and} \quad \|g(x) - \nabla f(x)\| \leq \kappa_g||x - \hat{x}||^2.$$

Therefore, since the all assumptions of Proposition 3.2 hold, then the desired inequality follows trivially from Proposition 3.2. □

The next proposition gives us an upper bound for $\|\nabla f(x^+) + \psi'(x^+)\|$ if $g(\cdot), B(\cdot), \nabla M^g_{x,z,\sigma}(\cdot)$ and x^+ satisfy some conditions.

Proposition 3.4. *Suppose **A1** holds and assume that the matrix* $B(z) \in \mathbb{R}^{n\times n}$, $g(x) \in \mathbb{R}^n$ *in* (3-1) *satisfy* (3-9) *for some* $\kappa_B, \kappa_g \geq 0$ *and* $x, \hat{x}, z, \hat{z} \in \mathbb{R}^n$. *Moreover, assume that* $x^+ \in \mathbb{R}^n$ *satisfies*

$$\|\nabla M^g_{x,z,\sigma}(x^+) + \psi'(x^+)\| \leq \theta\|x^+ - x\|^2, \tag{3-16}$$

for some $\psi'(x^+) \in \partial\psi(x^+)$, $\theta \geq 0$ *and* $\sigma > 0$. *If* $\rho, \rho^\star > 0$, *then*

$$\|\nabla f(x^+) + \psi'(x^+)\|^{\frac{3}{2}} \leq 2\Bigg[\left(\frac{\sigma + L + L^2\rho + \kappa_B^2\rho^\star + 2\theta}{2}\right)^{\frac{3}{2}}\|x^+ - x\|^3 + \kappa_g^{\frac{3}{2}}\|x - \hat{x}\|^3 + \frac{\|x - z\|^3}{2^{\frac{3}{2}}\rho^{\frac{3}{2}}} + \frac{\|z - \hat{z}\|^3}{2^{\frac{3}{2}}\rho^{\star\frac{3}{2}}}\Bigg].$$

Proof. From (1-1), (1-2), (3-9), (3-16), triangle inequality and definition of $M^g_{x,z,\sigma}(\cdot)$ in

(3-1), we get

$$
\begin{aligned}
\|\nabla f(x^+) + \psi'(x^+)\| &\le \|\nabla f(x^+) - \nabla M^g_{x,z,\sigma}(x^+)\| + \|\nabla M^g_{x,z,\sigma}(x^+) + \psi'(x^+)\| \\
&\le \|\nabla f(x^+) - g(x) - B(z)(x^+ - x) - \frac{\sigma}{2}\|x^+ - x\|(x^+ - x)\| \\
&+ \theta\|x^+ - x\|^2 \\
&\le \|\nabla f(x) - g(x) + \nabla f(x^+) - \nabla f(x) - B(z)(x^+ - x)\| \\
&+ \left(\frac{\sigma}{2} + \theta\right)\|x^+ - x\|^2 \\
&\le \|\nabla f(x^+) - \nabla f(x) - \nabla^2 f(x)(x^+ - x)\| + \|\nabla^2 f(x) - B(z)\|\|x^+ - x\| \\
&+ \|g(x) - \nabla f(x)\| + \left(\frac{\sigma}{2} + \theta\right)\|x^+ - x\|^2 \\
&\le \left(\frac{L+\sigma}{2} + \theta\right)\|x^+ - x\|^2 + \kappa_g\|x - \hat{x}\|^2 + \|B(z) - \nabla^2 f(z)\|\|x^+ - x\| \\
&+ \|\nabla^2 f(x) - \nabla^2 f(z)\|\|x^+ - x\| \\
&\le \left(\frac{L+\sigma}{2} + \theta\right)\|x^+ - x\|^2 + \kappa_g\|x - \hat{x}\|^2 + \kappa_B\|z - \hat{z}\|\|x^+ - x\| \\
&+ L\|x - z\|\|x^+ - x\|. \qquad (3\text{-}17)
\end{aligned}
$$

Hence, using Proposition 1.10 with $a = \|x - z\|/\rho^{1/2}$, $b = L\rho^{1/2}\|x^+ - x\|$ and $p = q = 2$, we obtain

$$
L\|x - z\|\|x^+ - x\| \le \frac{\|x - z\|^2}{2\rho} + \frac{L^2\rho\|x^+ - x\|^2}{2}. \qquad (3\text{-}18)
$$

Again, by Proposition 1.10 with $a = \|z - \hat{z}\|/\rho^{\star\frac{1}{2}}$, $b = \kappa_B\rho^{\star\frac{1}{2}}\|x^+ - x\|$ and $p = q = 2$, we have

$$
\kappa_B\|z - \hat{z}\|\|x^+ - x\| \le \frac{\|z - \hat{z}\|^2}{2\rho^\star} + \frac{\kappa_B^2\rho^\star\|x^+ - x\|^2}{2}. \qquad (3\text{-}19)
$$

It follows from (3-17), (3-18) and (3-19) that

$$
\begin{aligned}
\|\nabla f(x^+) + \psi'(x^+)\| \le \left(\frac{\sigma + L + L^2\rho + \kappa_B^2\rho^\star + 2\theta}{2}\right)\|x^+ - x\|^2 + \kappa_g\|x - \hat{x}\|^2 + \frac{1}{2\rho}\|x - z\|^2 \\
+ \frac{1}{2\rho^\star}\|z - \hat{z}\|^2.
\end{aligned}
$$

Raising the last inequality to the power of 3/2 and using the inequality (1-5) for the function $t \mapsto t^{3/2}$, $t \ge 0$ with $p_1 = p_2 = p_3 = p_4 = 1/4$, we get

$$
\begin{aligned}
\|\nabla f(x^+) + \psi'(x^+)\|^{\frac{3}{2}} \le 2\Bigg[\left(\frac{\sigma + L + L^2\rho + \kappa_B^2\rho^\star + 2\theta}{2}\right)^{\frac{3}{2}}\|x^+ - x\|^3 \\
+ \kappa_g^{\frac{3}{2}}\|x - \hat{x}\|^3 + \frac{\|x - z\|^3}{2^{\frac{3}{2}}\rho^{\frac{3}{2}}} + \frac{\|z - \hat{z}\|^3}{2^{\frac{3}{2}}\rho^{\star\frac{3}{2}}}\Bigg].
\end{aligned}
$$

which gives the desired inequality. □

3.2 Cubic Regularization Method with Inexact Gradient and Lazy Hessian approximations

In this section we discuss another algorithm that finds an ϵ-approximate critical point for the problem (0-3).

Algorithm 2. Cubic Regularization Method with Inexact Gradient and Lazy Hessian approximations

Step 0. Choose $x_{-1}, x_0 \in \mathbb{R}^n$ such that $\|x_0 - x_{-1}\|^3 = c/(m+1)$ for some $c > 0$. Take $\sigma_0 > 0$, $\theta \geq 0$, $m > 0$, $\bar{\kappa}_g, \bar{\kappa}_B \geq 0$, and set $t := 0$.
Step 1. Find the smallest integer $i \geq 0$ such that $2^{i-1}\sigma_t \geq \sigma_0(m+1)$.
Step 1.1. Construct $g_{t,i}$ and $B_{t,i}$ such that

$$\|B_{t,i}(x_t) - \nabla^2 f(x_t)\| \leq \frac{\bar{\kappa}_B}{2^{i-1}}\|x_t - x_{t-1}\| \quad \text{and} \quad \|g_{t,i}(x_t) - \nabla f(x_t)\| \leq \frac{\bar{\kappa}_g}{2^{i-1}}\|x_t - x_{t-1}\|^2. \tag{3-20}$$

Step 1.2. Compute $x_{t,i}^+$ such that

$$M^g_{x_t,x_t,2^i\sigma_t}(x_{t,i}^+) + \psi(x_{t,i}^+) \leq F(x_t), \quad \left\|\nabla M^g_{x_t,x_t,2^i\sigma_t}(x_{t,i}^+) + \psi'(x_{t,i}^+)\right\| \leq \theta\|x_{t,i}^+ - x_t\|^2, \tag{3-21}$$

for some $\psi'(x_{t,i}^+) \in \partial\psi(x_{t,i}^+)$, where $M^g_{x,z,\sigma}(\cdot)$ is as in (3-1).
Step 1.3. If

$$F(x_t) - F(x_{t,i}^+) \geq \frac{2^i\sigma_t}{12}\|x_{t,i}^+ - x_t\|^3 - \frac{\sigma_0}{48}\|x_t - x_{t-1}\|^3, \tag{3-22}$$

set $i_t := i$ and go to Step 2. Otherwise, set $i := i+1$ and go to Step 1.1.
Step 2. Set $x_{t+1} := x_{t,i_t}^+$, $\sigma_{t+1} := 2^{i_t-1}\sigma_t$, $B_t := B_{t,i_t}$, $t := t+1$.
Step 3. If $m > 0$, let $\hat{t} := t$, $B := B_{\hat{t}-1}$ and go to Step 3.1. Otherwise, return to Step 1.
Step 3.1. Find the smallest integer $j \geq 0$ such that $2^{j-1}\sigma_t \geq \sigma_0(m+1)$.
Step 3.2 Construct $g_{t,j}$ such that

$$\|g_{t,j}(x_t) - \nabla f(x_t)\| \leq \frac{\bar{\kappa}_g}{2^{j-1}}\|x_t - x_{t-1}\|.$$

Step 3.3 Compute $x_{t,j}^+$ such that

$$M^g_{x_t,x_{\hat{t}-1},2^j\sigma_t}(x_{t,j}^+) + \psi(x_{t,j}^+) \leq F(x_t), \quad \left\|\nabla M^g_{x_t,x_{\hat{t}-1},2^j\sigma_t}(x_{t,j}^+) + \psi'(x_{t,j}^+)\right\| \leq \theta\|x_{t,j}^+ - x_t\|^2,$$

for some $\psi'(x_{t,j}^+) \in \partial\psi(x_{t,j}^+)$, where $M^g_{x,z,\sigma}(\cdot)$ is as in (3-1).
Step 3.4 If

$$\begin{aligned} F(x_t) - F(x_{t,j}^+) \geq{}& \frac{2^j\sigma_t}{12}\|x_{t,j}^+ - x_t\|^3 - \frac{\sigma_0(m+1)}{12}\|x_t - x_{t-1}\|^3 - \frac{\sigma_0}{8(m+1)^2}\|x_t - x_{\hat{t}-1}\|^3 \\ &- \frac{\sigma_0}{48}\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3, \end{aligned} \tag{3-23}$$

set $j_t := j$ and go to Step 3.5. Otherwise, set $j := j+1$ and go to Step 3.2.
Step 3.5 Set $x_{t+1} := x^+_{t,j_t}$, $\sigma_{t+1} := 2^{j_t-1}\sigma_t$ and $t := t+1$.
Step 3.6 If $t \leq \hat{t} + m$, go to Step 3.1; otherwise, go to Step 1.

Remark 3.5. *(i) We mention that a Hessian approximation $B_{t,i}$ and gradient approximation $g_{t,i}$ satisfying the condition (3-20) can be obtained by evaluations of the function f. Indeed, it follows from Lemma 1.11 with*

$$x := x_t \quad \text{and} \quad h_g = \left(\frac{6\|x_t - x_{t-1}\|^2}{2^{i-1}}\right)^{\frac{1}{2}},$$

that the approximation $g_{t,i}$ as defined in (1-7) satisfies (3-20) with $\bar{\kappa}_g = \sqrt{n}L$. On the other hand, from Lemma 1.14 with

$$x := x_t \quad \text{and} \quad h_B = \frac{3\|x_t - x_{t-1}\|}{5.2^{i-1}},$$

that the approximation $B_{t,i}$ as defined in (1-23) satisfies (3-20) with $\bar{\kappa}_B = nL$. (ii) Note that $x^+_{t,i}$ as in Step 1.2 is an inexact solution of the problem

$$\min_{y \in Q} M^g_{x_t, x_t, 2^i\sigma_t}(y) + \psi(y).$$

Conditions in (3-21) only require a decrease of the cubic regularized model summed with the function ψ and an approximate first-order stationary point of the above problem. (iii) Note that the sequence of parameters $\{\sigma_t\}$ can be nonmonotone. Indeed, if $i_t = 0$, we have $\sigma_{t+1} = 2^{i_t-1}\sigma_t = \sigma_t/2 \leq \sigma_t$. (iv) Note that both conditions (3-22) and (3-23) allow acceptance of a trial point $x^+_{t,i}$ such that

$$F(x^+_{t,i}) > F(x_t).$$

Consequently, the sequence $\{F(x_t)\}_{t\geq 0}$ may be nonmonotone. (v) Considering $B_{t,i}$ as described in (i), the main differences between our algorithm and the first-order CNM of [6, Algorithm 4] lie in the choice of the parameters h_g and h_B used for the finite-difference approximations and the timing of Hessian updates. In their method, both the parameters h_g and h_B depend on the precision ϵ and may be smaller than our choice. On the other hand, we obtain a suitable Hessian approximation—possibly updating it multiple times—at the first iteration of each block and retain it for the remaining iterations of that block. In contrast, the method in [6, Algorithm 4] constructs a Hessian approximation at the first iteration of each block but terminates the block's iterations (starting a new block) if a certain acceptance condition is not met.

3.3 Iteration-complexity for Algorithm 2

In the following, we proceed to the complexity analysis of Algorithm 2. We begin by proving that the sequence of parameters $\{\sigma_t\}$ is bounded from above. In particular, we show that the inner procedures in Steps 1.3 and 3.4 end in a finite number of iterations.

Lemma 3.6. *Suppose* ***A1*** *holds. Then, the regularization parameters σ_t in Algorithm 2 for Step 1.3 satisfy*

$$\sigma_0(m+1) \leq \sigma_t \leq \bar{\sigma}_{max} := \sigma_0(m+1) + 2\left(L + 4\sqrt{3}\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{16}{\sigma_0}\right)^{\frac{1}{2}} + 144\bar{\kappa}_g^3\left(\frac{16}{\sigma_0}\right)^2\right), \quad (3\text{-}24)$$

for all $t \geq 1$. As a consequence, the inner procedure in Step 1.3 end in a finite number of iterations.

Proof. Let us prove by induction on t that (3-24) holds. For $t = 1$ the proof is analogous to the one in Lemma 2.6.

Now, suppose that the inequality (3-24) holds for some natural number $t > 1$, then $\sigma_0(m+1) \leq \sigma_t \leq \bar{\sigma}_{max}$. Let us divide the proof into two cases:
Case ($i_t = 0$): From Step 1, we have $2^{0-1}\sigma_t \geq \sigma_0(m+1)$, consequently

$$\sigma_0(m+1) \leq \sigma_{t+1} = 2^{0-1}\sigma_t = \frac{1}{2}\sigma_t \leq \sigma_t \leq \bar{\sigma}_{\max}.$$

Case ($i_t > 0$): From Step 1, we get $\sigma_{t+1} = 2^{i_t-1}\sigma_t \geq \sigma_0(m+1)$. Now, assume by contradiction that $2^{i_t-1}\sigma_t > \bar{\sigma}_{max}$. Hence, we have

$$\begin{aligned} 2^{i_t-1}\sigma_t &> 2\left(L + 4\sqrt{3}\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{16}{\sigma_0}\right)^{\frac{1}{2}} + 144\bar{\kappa}_g^3\left(\frac{16}{\sigma_0}\right)^2\right) \\ &\geq 2\left(L + \sqrt{6}\left(\frac{\bar{\kappa}_B}{2^{i_t-1}}\right)^{\frac{3}{2}}\left(\frac{16}{\sigma_0}\right)^{\frac{1}{2}} + 18\left(\frac{\bar{\kappa}_g}{2^{i_t-1}}\right)^3\left(\frac{16}{\sigma_0}\right)^2\right), \end{aligned}$$

where we used the fact that $4 > \sqrt{2}/(2^{i_t-1})^{3/2}$ in the last inequality. Then, it follows from Proposition 3.1 with $\sigma = 2^{i_t-1}\sigma_t$, $\kappa_B = \bar{\kappa}_B/2^{i_t-1}$, $\kappa_g = \bar{\kappa}_g/2^{i_t-1}$, $x^+ = x_{t,i}^+$, $x = x_t$, $\hat{x} = x_{t-1}$, $\bar{\rho} = 16/\sigma_0$, that

$$F(x_t) - F(x_{t,i}^+) \geq \frac{2^{i_t-1}\sigma_t}{12}\|x_{t,i}^+ - x_t\|^3 - \frac{\sigma_0}{48}\|x_t - x_{t-1}\|^3,$$

hence, the inequality (3-22) holds for $i = i_t - 1$, contradicting the minimality of i_t. So $\sigma_{t+1} \leq \bar{\sigma}_{max}$, which implies the desired inequality. □

Lemma 3.7. *Suppose* ***A1*** *holds. Then, the regularization parameters σ_t in Algorithm 2* *for Step* 3.4 *satisfy*

$$\sigma_0(m+1) \leq \sigma_t \leq \tilde{\sigma}_{max} := \sigma_0(m+1) + 2\Bigg(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{24(m+1)^2}{9\sigma_0}\right)^{\frac{1}{2}} + 4\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{16}{\sigma_0}\right)^{\frac{1}{2}} + 16\bar{\kappa}_g^3\left(\frac{8}{\sigma_0(m+1)}\right)^2\Bigg), \tag{3-25}$$

for all $t \geq 1$. *As a consequence, the inner procedure in Step* 3.4 *end in a finite number of iterations.*

Proof. Let us prove by induction on t that (3-24) holds. For $t=1$ the proof is analogous to the one in Lemma 2.6.

Now, suppose that the inequality (3-25) holds for some natural number $t > 1$, then $\sigma_0(m+1) \leq \sigma_t \leq \tilde{\sigma}_{max}$. Let us divide the proof into two cases:
Case ($j_t = 0$): From Step 3.1, we have $2^{0-1}\sigma_t \geq \sigma_0(m+1)$, consequently

$$\sigma_0(m+1) \leq \sigma_{t+1} = 2^{0-1}\sigma_t = \frac{1}{2}\sigma_t \leq \sigma_t \leq \tilde{\sigma}_{\max}.$$

Case ($j_t > 0$): From Step 3.1, we get $\sigma_{t+1} = 2^{j_t-1}\sigma_t \geq \sigma_0(m+1)$, so the first inequality holds. Now, assume by contradiction that $2^{j_t-1}\sigma_t > \tilde{\sigma}_{max}$. Hence, we get

$$\begin{aligned}
2^{j_t-1}\sigma_t &> 2\left(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{24(m+1)^2}{9\sigma_0}\right)^{\frac{1}{2}} + 4\bar{\kappa}_B^{\frac{3}{2}}\left(\frac{16}{\sigma_0}\right)^{\frac{1}{2}} + 16\bar{\kappa}_g^3\left(\frac{8}{\sigma_0(m+1)}\right)^2\right)\\
&\geq 2\Bigg(L + \sqrt{2}L^{\frac{3}{2}}\left(\frac{24(m+1)^2}{9\sigma_0}\right)^{\frac{1}{2}} + \sqrt{2}\left(\frac{\bar{\kappa}_B}{2^{j_t-1}}\right)^{\frac{3}{2}}\left(\frac{16}{\sigma_0}\right)^{\frac{1}{2}}\\
&\quad + 2\left(\frac{\bar{\kappa}_g}{2^{j_t-1}}\right)^3\left(\frac{8}{\sigma_0(m+1)}\right)^2\Bigg),
\end{aligned}$$

where we used the fact that $4 > \sqrt{2}/(2^{j_t-1})^{3/2}$ in the last inequality. Then, it follows from Proposition 3.2 with $\sigma = 2^{j_t-1}\sigma_t$, $\kappa_B = \bar{\kappa}_B/2^{j_t-1}$, $\kappa_g = \bar{\kappa}_g/2^{j_t-1}$, $x^+ = x^+_{t,j}$, $x = x_t$, $\hat{x} = x_{t-1}$, $z = x_{\hat{t}-1}$, $\hat{z} = x_{\hat{t}-2}$, $\tilde{\rho} = 8/(\sigma_0(m+1))$, $\hat{\rho} = 24(m+1)^2/(9\sigma_0)$ and $\bar{\rho} = 16/\sigma_0$, that

$$\begin{aligned}
F(x_t) - F(x^+_{t,j}) &\geq \frac{2^{j_t-1}\sigma_t}{12}\|x^+_{t,i} - x_t\|^3 - \frac{\sigma_0(m+1)}{12}\|x_t - x_{t-1}\|^3 - \frac{\sigma_0}{8(m+1)^2}\|x_t - x_{\hat{t}-1}\|^3\\
&\quad - \frac{\sigma_0}{48}\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3.
\end{aligned}$$

Therefore, the inequality (3-23) holds for $j = j_t - 1$, contradicting the minimality of j_t. So $\sigma_{t+1} \leq \tilde{\sigma}_{max}$, which implies the desired inequality. □

Next, we present some key results to establish a global iteration-complexity bound for Algorithm 2.

Lemma 3.8. *Suppose **A1** and **A2** holds. Let $\{x_t\}_{t=1}^{T}$ be a sequence generated by Algorithm 2. Let τ be the block number associated to the T-th iteration, that is, $T = (\tau-1)*(m+1)+\ell$ with $\ell \in \mathbb{N}$ and $1 \le \ell \le m+1$. Then,*

(i) $F(x_T) \le F(x_{(\tau-1)(m+1)}) + \frac{\sigma_0(m+1)}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3 - \frac{\sigma_0(m+1)}{24}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3$, *if* $\tau > 1$.

(ii) $F(x_T) \le F(x_0) + \frac{\sigma_0 c}{48} - \frac{\sigma_0(m+1)}{24}\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3$, *if* $\tau = 1$.

Proof. For $1 \le \ell \le m+1$, consider the sequence $\{x_t\}_{(\tau-1)(m+1)+1}^{(\tau-1)(m+1)+\ell}$ associated to the block τ. It follows from Algorithm 2, that the $((\tau-1)(m+1)+1)$-th iteration satisfies (3-22), whereas the $(\ell-1)$-th consecutive ones satisfy (3-23). Hence,

$$F(x_{(\tau-1)(m+1)+1}) \le F(x_{(\tau-1)(m+1)}) - \frac{\sigma_{(\tau-1)(m+1)+1}}{6}\|x_{(\tau-1)(m+1)+1} - x_{(\tau-1)(m+1)}\|^3 + \frac{\sigma_0}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3,$$

$$F(x_{(\tau-1)(m+1)+2}) \le F(x_{(\tau-1)(m+1)+1}) - \frac{\sigma_{(\tau-1)(m+1)+2}}{6}\|x_{(\tau-1)(m+1)+2} - x_{(\tau-1)(m+1)+1}\|^3 + \frac{\sigma_0(m+1)}{12}\|x_{(\tau-1)(m+1)+1} - x_{(\tau-1)(m+1)}\|^3 + \frac{\sigma_0}{8(m+1)^2}\|x_{(\tau-1)(m+1)+1} - x_{(\tau-1)(m+1)}\|^3 + \frac{\sigma_0}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3,$$

$$F(x_{(\tau-1)(m+1)+3}) \le F(x_{(\tau-1)(m+1)+2}) - \frac{\sigma_{(\tau-1)(m+1)+3}}{6}\|x_{(\tau-1)(m+1)+3} - x_{(\tau-1)(m+1)+2}\|^3 + \frac{\sigma_0(m+1)}{12}\|x_{(\tau-1)(m+1)+2} - x_{(\tau-1)(m+1)+1}\|^3 + \frac{\sigma_0}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3 + \frac{\sigma_0}{8(m+1)^2}\|x_{(\tau-1)(m+1)+2} - x_{(\tau-1)(m+1)}\|^3,$$

$$\vdots$$

$$F(x_{(\tau-1)(m+1)+l}) \le F(x_{(\tau-1)(m+1)+(l-1)}) + \frac{\sigma_0(m+1)}{12}\|x_{(\tau-1)(m+1)+(l-1)} - x_{(\tau-1)(m+1)+(l-2)}\|^3 - \frac{\sigma_{(\tau-1)(m+1)+l}}{6}\|x_{(\tau-1)(m+1)+l} - x_{(\tau-1)(m+1)+(l-1)}\|^3 + \frac{\sigma_0}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3 + \frac{\sigma_0}{8(m+1)^2}\|x_{(\tau-1)(m+1)+(l-1)} - x_{(\tau-1)(m+1)}\|^3.$$

Combining all the above inequalities with the fact that $T-(\tau-1)(m+1) = \ell$, (3-24) and

(3-25) hold, we get

$$
\begin{aligned}
F(x_T) \le F(x_{(\tau-1)(m+1)}) &+ \frac{(T-(\tau-1)(m+1))\sigma_0}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3 \\
&- \frac{\sigma_0(m+1)}{6}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3 + \frac{\sigma_0}{8(m+1)^2}\sum_{t=(\tau-1)(m+1)+1}^{T-1}\|x_t - x_{(\tau-1)(m+1)}\|^3 \\
&+ \frac{\sigma_0(m+1)}{12}\sum_{t=(\tau-1)(m+1)}^{T-2}\|x_{t+1}-x_t\|^3.
\end{aligned}
$$

Hence, from $\ell \le m+1$, some algebraic manipulations and Lemma 2.7(i), we get the desired result. The result in Lemma 3.8(ii) follows analogously. □

Finally, from this lemma, we obtain a key bound result for the sequence $\{\|x_{t+1} - x_t\|^3\}$.

Lemma 3.9. *Suppose **A1** and **A2** holds. Let $\{x_t\}_{t=1}^T$ be a sequence generated by Algorithm 2. Let $\tau \in \mathbb{N} - \{0\}$ be the block number associated with the T-th iteration as in Lemma 3.8. Then,*

$$
\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3 \le \frac{1}{m+1}\left(\frac{48(F(x_0)-F^\star)}{\sigma_0} + c\right).
$$

Proof. It follows from Lemma 3.8(i) with multiples values of τ and Lemma 3.8(ii) that

$$
F(x_{m+1}) \le F(x_0) - \frac{\sigma_0(m+1)}{24}\sum_{t=0}^{m}\|x_{t+1}-x_t\|^3 + \frac{\sigma_0 c}{48},
$$

$$
F(x_{2(m+1)}) \le F(x_{m+1}) - \frac{\sigma_0(m+1)}{24}\sum_{t=m+1}^{2m+1}\|x_{t+1}-x_t\|^3 + \frac{\sigma_0(m+1)}{48}\|x_{m+1}-x_m\|^3,
$$

$$
\vdots
$$

$$
\begin{aligned}
F(x_T) \le F(x_{(\tau-1)(m+1)}) &- \frac{\sigma_0(m+1)}{24}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3 \\
&+ \frac{\sigma_0(m+1)}{48}\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3.
\end{aligned}
$$

Combining all the last inequalities, it follows that

$$
\begin{aligned}
F(x_T) \le F(x_0) &- \frac{\sigma_0(m+1)}{24}\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3 + \frac{\sigma_0(m+1)}{48}\sum_{k=1}^{\tau-1}\|x_{k(m+1)} - x_{k(m+1)-1}\|^3 \\
&+ \frac{\sigma_0 c}{48}.
\end{aligned}
$$

From last inequality, some algebraic manipulations and $F(x_T) > F^\star$, we have

$$F^\star \leq F(x_0) - \frac{\sigma_0(m+1)}{48}\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3 + \frac{\sigma_0 c}{48},$$

which implies the desired inequality. □

The following lemma gives us another key bound in the algorithm complexity analysis.

Lemma 3.10. *Suppose* ***A1*** *holds. Let* $\{x_t\}_{t=1}^T$ *be the sequence generated by Algorithm* 2*. Let* $\tau \in \mathbb{N} - \{0\}$ *be the block number associated with the* T*-th iteration as in Lemma 3.8. Then,*

$$\begin{aligned} 2^{-1}\sum_{t=(\tau-1)(m+1)}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} &\leq \left(\tilde{\eta}^{3/2}+\frac{(m+1)^{\frac{3}{2}}}{3}+(2\kappa_g)^{\frac{3}{2}}\right)\\ \times \sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3 &+ \left((2\kappa_g)^{\frac{3}{2}}+1\right)\|x_{(\tau-1)(m+1)}-x_{(\tau-1)(m+1)-1}\|^3, \end{aligned} \tag{3-26}$$

where $\tilde{\eta} := \hat{\sigma}_{max} + (L+L^2(m+1))/2 + 2\bar{\kappa}_B^2(m+1)^{\frac{2}{3}} + \theta$ *and* $\hat{\sigma}_{max} := \max\{\bar{\sigma}_{max}, \tilde{\sigma}_{max}\}$.

Proof. For $1 \leq \ell \leq m+1$, let $\{x_t\}_{(\tau-1)(m+1)+1}^{(\tau-1)(m+1)+\ell}$ be the sequence associated with the block τ. Since $x_{(\tau-1)(m+1)+1}$ is the first iteration, it follows from Algorithm 2 and Proposition 3.4 with $\sigma = 2^{i_t}\sigma_t$, $x^+ = x_{t+1}$, $x = x_t$, $z = x_t$, $\hat{z} = x_{t-1}$, $\kappa_g = \bar{\kappa}_g/2^{i_t-1}$, $\kappa_B = \bar{\kappa}_B/2^{i_t-1}$, $\rho = m+1$, $\rho^\star = (m+1)^{2/3}$ and $t = (\tau-1)(m+1)$ that

$$\begin{aligned} 2^{-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} &\leq \left(\frac{2^{i_t}\sigma_t + L + L^2(m+1)}{2} + \frac{(m+1)^{\frac{2}{3}}}{2}\left(\frac{\bar{\kappa}_B}{2^{i_t-1}}\right)^2 + \theta\right)^{\frac{3}{2}}\\ \times \|x_{t+1}-x_t\|^3 &+ \left(\frac{\bar{\kappa}_g}{2^{i_t-1}}\right)^{\frac{3}{2}}\|x_t-x_{t-1}\|^3 + \frac{\|x_t-x_{t-1}\|^3}{2^{\frac{3}{2}}(m+1)}, \end{aligned}$$

which, combined with $2^{i_t-1}\sigma_t = \sigma_{t+1} \leq \hat{\sigma}_{max}$ (see Step 2 of Algorithm 2, (3-24) and (3-25)) and $i_t \geq 0$, yields

$$\begin{aligned} 2^{-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} &\leq \tilde{\eta}^{\frac{3}{2}}\|x_{t+1}-x_t\|^3 + (2\bar{\kappa}_g)^{\frac{3}{2}}\|x_t-x_{t-1}\|^3\\ &+ \frac{\|x_t-x_{t-1}\|^3}{2^{\frac{3}{2}}(m+1)}. \end{aligned} \tag{3-27}$$

Let us now consider the case where x_{t+1} corresponds to the remaining iterations of the sequence, i.e, t satisfies $\hat{t} := (\tau-1)(m+1)+1 \leq t \leq (\tau-1)(m+1)+\ell-1$. Hence, it follows from Algorithm 2 and Proposition 3.4 with $\sigma = 2^{j_t}\sigma_t$, $x^+ = x_{t+1}$, $x = x_t$, $z = x_{\hat{t}-1}$,

$\hat{z} = x_{\hat{t}-2}$, $\kappa_g = \bar{\kappa}_g/2^{j_t-1}$, $\kappa_B = \bar{\kappa}_B/2^{j_t-1}$, $\rho = m+1$ and $\rho^\star = (m+1)^{2/3}$ that

$$2^{-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\frac{2^{j_t}\sigma_t + L + L^2(m+1)}{2} + \frac{(m+1)^{\frac{2}{3}}}{2}\left(\frac{\bar{\kappa}_B}{2^{j_t-1}}\right)^2 + \theta\right)^{\frac{3}{2}}$$
$$\times \|x_{t+1} - x_t\|^3 + \left(\frac{\bar{\kappa}_g}{2^{j_t-1}}\right)^{\frac{3}{2}} \|x_t - x_{t-1}\|^3 + \frac{\|x_t - x_{\hat{t}-1}\|^3}{2^{\frac{3}{2}}(m+1)^{\frac{3}{2}}} + \frac{\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3}{2^{\frac{3}{2}}(m+1)}.$$

Since $2^{j_t-1}\sigma_t = \sigma_{t+1} \leq \hat{\sigma}_{max}$ (see Step 3.5 of Algorithm 2, (3-24) and (3-25)) and $j_t \geq 0$, we get

$$2^{-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \tilde{\eta}^{\frac{3}{2}}\|x_{t+1} - x_t\|^3 + (2\bar{\kappa}_g)^{\frac{3}{2}}\|x_t - x_{t-1}\|^3 + \frac{\|x_t - x_{\hat{t}-1}\|^3}{2^{\frac{3}{2}}(m+1)^{\frac{3}{2}}}$$
$$+ \frac{\|x_{\hat{t}-1} - x_{\hat{t}-2}\|^3}{2^{\frac{3}{2}}(m+1)} \tag{3-28}$$

for all $\hat{t} \leq t \leq (\tau-1)(m+1)+\ell-1$. From relation (3-28) with $\hat{t} \leq t \leq (\tau-1)(m+1)+\ell-1$ and the inequality (3-27) obtained, we have

$$2^{-1}\sum_{t=(\tau-1)(m+1)}^{(\tau-1)(m+1)+l-1} \|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(\tilde{\eta}^{\frac{3}{2}} + \frac{(m+1)^{\frac{3}{2}}}{3} + (2\bar{\kappa}_g)^{\frac{3}{2}}\right)$$
$$\times \sum_{t=(\tau-1)(m+1)}^{(\tau-1)(m+1)+l-1} \|x_{t+1} - x_t\|^3 + \left((2\bar{\kappa}_g)^{\frac{3}{2}} + \frac{\ell}{(m+1)}\right)\|x_{(\tau-1)(m+1)} - x_{(\tau-1)(m+1)-1}\|^3.$$

Where in the last inequality we used Lemma 2.7(i). From $T = (\tau-1)(m+1)+\ell$, and $\ell \leq m+1$, we obtain the desired result. □

We are now ready to present an iteration complexity bound for the Algorithm 2 in terms of the outer iteration number.

Theorem 3.11. *Suppose* ***A1*** *and* ***A2*** *holds. Let* $\{x_t\}_{t=1}^T$ *be a sequence generated by Algorithm* 2*. Let* $\tau \in \mathbb{N} - \{0\}$ *be the block number associated with the* T*-th iteration as in Lemma 3.8, then it follows that*

$$2^{-1}\sum_{t=0}^{T-1}\|\nabla f(x_{t+1}) + \psi'(x_{t+1})\|^{\frac{3}{2}} \leq \left(1 + \tilde{\eta}^{\frac{3}{2}} + \frac{(m+1)^{\frac{3}{2}}}{3} + 2(2\bar{\kappa}_g)^{\frac{3}{2}}\right)\eta$$
$$+ \left((2\bar{\kappa}_g)^{\frac{3}{2}} + 1\right)\frac{c}{m+1}, \tag{3-29}$$

where $\eta := (48(F(x_0) - F^\star)/\sigma_0 + c)/(m+1)$*,* $\tilde{\eta}$ *and* $\hat{\sigma}_{max}$ *are as in Lemma 3.10. As a consequence, given* $\epsilon > 0$ *Algorithm* 2 *needs at most* $\mathcal{O}\left(m^{1/2}\epsilon^{-3/2}\right)$ *iterations to generate an* $\epsilon-$*approximate critical point for problem* (0-3)*.*

Proof. Let $\bar{\eta} := \tilde{\eta}^{\frac{3}{2}} + (m+1)^{\frac{3}{2}}/3 + (2\bar{\kappa}_g)^{\frac{3}{2}}$. It follows from inequality (3-26) with multiples values of τ that

$$2^{-1}\sum_{t=0}^{m}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \bar{\eta}\sum_{t=0}^{m}\|x_{t+1}-x_t\|^3 + \left((2\bar{\kappa}_g)^{\frac{3}{2}}+1\right)\|x_0-x_{-1}\|^3,$$

$$2^{-1}\sum_{t=m+1}^{2m+1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \bar{\eta}\sum_{t=m+1}^{2m+1}\|x_{t+1}-x_t\|^3 + \left((2\bar{\kappa}_g)^{\frac{3}{2}}+1\right)\|x_{m+1}-x_m\|^3,$$

$$2^{-1}\sum_{t=2(m+1)}^{3m+2}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \bar{\eta}\sum_{t=2(m+1)}^{3m+2}\|x_{t+1}-x_t\|^3 + \left((2\bar{\kappa}_g)^{\frac{3}{2}}+1\right)\|x_{2(m+1)}-x_{2(m+1)-1}\|^3,$$

$$\vdots$$

$$2^{-1}\sum_{t=(\tau-1)(m+1)}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \bar{\eta}\sum_{t=(\tau-1)(m+1)}^{T-1}\|x_{t+1}-x_t\|^3 + \left((2\bar{\kappa}_g)^{\frac{3}{2}}+1\right)\|x_{(\tau-1)(m+1)}-x_{(\tau-1)(m+1)-1}\|^3.$$

Combining the last inequalities, we get

$$2^{-1}\sum_{t=0}^{T-1}\|\nabla f(x_{t+1})+\psi'(x_{t+1})\|^{\frac{3}{2}} \le \left(\bar{\eta}+(2\bar{\kappa}_g)^{\frac{3}{2}}+1\right)\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3 + \left((2\bar{\kappa}_g)^{\frac{3}{2}}+1\right)\|x_0-x_{-1}\|^3.$$

where $T=(\tau-1)(m+1)+\ell$ with $\ell\in\mathbb{N}$ and $1\le\ell\le m+1$. From $\sum_{t=0}^{T-1}\|x_{t+1}-x_t\|^3\le\eta$, we obtain the inequality (3-29). Now, combining the inequality (3-29), definition of $\hat{\sigma}_{max}$, η and $\tilde{\eta}$ we get the second part of the theorem. □

The next theorem gives us a complexity on the number of functions and gradient evaluations of Algorithm 2.

Theorem 3.12. *Suppose **A1** and **A2** hold. Consider that Algorithm* 2 *is implemented such that $B_{t,i}$ in Step 1.1 is computed as in* (1-23) *and $g_{t,i}$ in Steps* 1.1 *and* 3.2 *are computed as in* (1-7)*. Then, the number of function evaluations up to the T-th iteration, $FE(T)$, is* *bounded as follows:*

$$FE(T) \leq \frac{(n^2+mn+2n+m+2)(T+m)}{m+1}\left(2+\log_2\frac{\hat{\sigma}_{max}}{\sigma_0}\right) + \log_2\frac{\hat{\sigma}_{max}}{\sigma_0(m+1)}, \tag{3-30}$$

where $\hat{\sigma}_{max}$ is as in Lemma 3.10. As a consequence, given $\epsilon > 0$, the total number of FE required to generated an ϵ-approximate critical point is $\mathcal{O}\left((n^2+mn)m^{-1/2}\epsilon^{-3/2}+(n^2+mn)\right)$.

Proof. Consider an arbitrary block τ. Hence, using the notation in Lemma 3.8, that is, $T = (\tau-1)(m+1)+\ell$ with $1 \leq \ell \leq m+1$, it follows from Algorithm 2 that the number of functions evaluations for the first and for the other iterations ($t \in \{(\tau-1)(m+1)+1, (\tau-1)(m+1)+2, \ldots, (\tau-1)(m+1)+l-1\}$) of block τ are bounded, respectively, by

$$(i_{(\tau-1)(m+1)}+1)(n^2+2n+1), \quad (n+1)(j_t+1), \tag{3-31}$$

if τ is not the first block and

$$2+(i_{(\tau-1)(m+1)}+1)(n^2+2n+1), \quad (n+1)(j_t+1),$$

otherwise. From Steps 2 and 3.4 of Algorithm 1, we have $2^{i_t-1}\sigma_t = \sigma_{t+1}$ and $2^{j_t-1}\sigma_t = \sigma_{t+1}$, which implies that

$$i_{(\tau-1)(m+1)}+1 = \log_2\sigma_{(\tau-1)(m+1)+1} - \log_2\sigma_{(\tau-1)(m+1)} + 2,$$

and $j_t+1 = \log_2\sigma_{t+1} - \log_2\sigma_t + 2$ for all $(\tau-1)(m+1)+1 \leq t \leq T-1$. Combining the last equalities with (3-31), we have

$$\begin{aligned}
&(i_{(\tau-1)(m+1)}+1)(n^2+2n+1) + (n+1)\sum_{t=(\tau-1)(m+1)+1}^{T-1}(j_t+1)\\
&= (n^2+2n+1)(\log_2\sigma_{(\tau-1)(m+1)+1} - \log_2\sigma_{(\tau-1)(m+1)}+2)\\
&+ (n+1)\sum_{t=(\tau-1)(m+1)+1}^{T-1}(\log_2\sigma_{t+1} - \log_2\sigma_t + 2)\\
&\leq 2(n^2+mn+2n+m+2) + (n+1)^2(\log_2\sigma_{(\tau-1)(m+1)+1} - \log_2\sigma_{(\tau-1)(m+1)})\\
&+ \log_2\sigma_T - \log_2\sigma_{(\tau-1)(m+1)+1}.
\end{aligned}$$

Applying the last inequalities for multiples values of τ, we obtain

$$\begin{aligned}
&2+(i_0+1)(n^2+2n+1)+(n+1)\sum_{t=1}^{m}(j_t+1) = 2(n^2+mn+2n+m+2)\\
&+ (n+1)^2(\log_2\sigma_1 - \log_2\sigma_0) + \log_2\sigma_{m+1} - \log_2\sigma_1,
\end{aligned}$$

$$(i_{m+1}+1)(n^2+2n+1)+(n+1)\sum_{t=m+2}^{2m+1}(j_t+1) \leq 2(n^2+mn+2n+m+2)$$
$$+(n+1)^2(\log_2\sigma_{m+2}-\log_2\sigma_{m+1})+\log_2\sigma_{2(m+1)}-\log_2\sigma_{m+2},$$

$$\vdots$$

$$(i_{(\tau-1)(m+1)}+1)(n^2+2n+1)+(n+1)\sum_{t=(\tau-1)(m+1)+1}^{T-1}(j_t+1) \leq$$
$$2(n^2+mn+2n+m+2)+(n+1)^2(\log_2\sigma_{(\tau-1)(m+1)+1}-\log_2\sigma_{(\tau-1)(m+1)})$$
$$+\log_2\sigma_T-\log_2\sigma_{(\tau-1)(m+1)+1}.$$

Combining the last inequalities with (3-24) and (3-25), we obtain

$$FE(T) \leq 2\tau(n^2+mn+2n+m+2)+((n+1)^2+1)\sum_{k=0}^{\tau-1}\log_2\frac{\sigma_{k(m+1)+1}}{\sigma_{k(m+1)}}$$
$$+\log_2\frac{\hat{\sigma}_{max}}{\sigma_0(m+1)}.$$

which, combined with $\sigma_t \geq \sigma_0$, yields

$$FE(T) \leq \frac{2(T+m)(n^2+mn+2n+m+2)}{m+1}+\frac{(T+m)((n+1)^2+1)}{m+1}\log_2\frac{\hat{\sigma}_{max}}{\sigma_0}$$
$$+\log_2\frac{\hat{\sigma}_{max}}{\sigma_0(m+1)},$$

where in the last inequality we used $\tau \leq (T+m)/(m+1)$. The inequality in (3-30) now follows from the last inequality and some algebraic manipulations. On the other hand, the second result follows from (3-30) and the complexity on T. □

Remark 3.13. *(i) If $B_{t,i}$ is updated as in (1-32), it can be shown that the statement of the last theorem still holds, however with a slight modification:*

$$FE(T) \leq \frac{(n^2+mn+3n+m+2)(T+m)}{m+1}\left(2+\log_2\frac{\hat{\sigma}_{max}}{\sigma_0}\right)+\log_2\frac{\hat{\sigma}_{max}}{\sigma_0(m+1)},$$

(ii) Comparing the results in Theorem 3.12 with the ones for [6, Algorithm 4], we see that the bound in (3-30) is similar, in terms of n and m, to the one in [6, Lemma 5.4]. At the same time, their bound on the total number of function evaluations, to generate an ϵ-approximate critical point, is $\mathcal{O}\left((n^2+mn)m^{-1/2}\epsilon^{-3/2}+(n^2+mn)\right)$, so the complexity in Theorem 3.12 is maintained. (iii) Taking $m=n$ in Theorem 2.11 we get a complexity of $\mathcal{O}\left(n^{3/2}\epsilon^{-3/2}+n^2\right)$ to the number of function evaluations, which is a reduction by a factor of $n^{1/2}$ in comparison with the method proposed in [4].

CHAPTER 4

Conclusion

In this work, we proposed and analyzed some methods for solving non-convex optimization problems. We proposed two versions of the cubic regularization method with lazy Hessian approximations for solving the general problem (0-3). For a given precision ϵ, it was shown that the first and second Algorithms require at most $\mathcal{O}\left(m^{1/2}\epsilon^{-3/2}\right)$ outer iterations to generate an $\epsilon-$approximate critical point for the aforementioned problem. When the derivatives are computed by finite difference approaches, we show that Algorithm 1 (resp. Algorithm 2) needs at most $\mathcal{O}\left((n+m)m^{-1/2}\epsilon^{-3/2}+(n+m)\right)$ (resp. $\mathcal{O}\left((n^2+mn)m^{-1/2}\epsilon^{-3/2}+(n^2+mn)\right))$ gradient and function (resp. function) evaluations to generate an ϵ-approximate critical point. In this regard, the two algorithms proposed maintain the same complexity as in [6] in terms of both function and gradient evaluations, as well as the number of outer iterations required to achieve an ϵ-stationary point of (0-3). At the same time, when we take $m=n$, it gives us the better complexity in terms of functions and gradient (resp. function) evaluation of order $\mathcal{O}\left(n^{1/2}\epsilon^{-3/2}+n\right)$ (resp. $\mathcal{O}\left(n^{3/2}\epsilon^{-3/2}+n^2\right))$ to achieve an ϵ-stationary point of the main problem, where n is the dimension of the domain of the objective function. In a future work, we will analyze the numerical behavior of the proposed algorithms and compare them with other existing methods.

Bibliography

[1] BECK, A.; TEBOULLE, M. **Gradient-based algorithms with applications to signal-recovery problems.**, 2010.

[2] BORSDORF, R.; HIGHAM, N. J. **A preconditioned newton algorithm for the nearest correlation matrix**. *IMA Journal of Numerical Analysis*, 30(1):94–107, 2010.

[3] CARTIS, C.; GOULD, N. I.; TOINT, P. L. **On the complexity of steepest descent, newton's and regularized newton's methods for nonconvex unconstrained optimization problems**. *Siam journal on optimization*, 20(6):2833–2852, 2010.

[4] CARTIS, C.; GOULD, N. I.; TOINT, P. L. **On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization**. *SIAM Journal on Optimization*, 22(1):66–86, 2012.

[5] DOIKOV, N.; CHAYTI, E. M.; JAGGI, M. **Second-order optimization with lazy hessians**. In: Krause, A.; Brunskill, E.; Cho, K.; Engelhardt, B.; Sabato, S.; Scarlett, J., editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 de **Proceedings of Machine Learning Research**, p. 8138–8161. PMLR, 23–29 Jul 2023.

[6] DOIKOV, N.; GRAPIGLIA, G. N. **First and zeroth-order implementations of the regularized newton method with lazy approximated hessians**. *arXiv preprint arXiv:2309.02412*, 2023.

[7] DOIKOV, N.; JAGGI, M.; OTHERS. **Second-order optimization with lazy hessians**. In: *International Conference on Machine Learning*, p. 8138–8161. PMLR, 2023.

[8] GRAPIGLIA, G. N.; GONÇALVES, M. L.; SILVA, G. **A cubic regularization of newton's method with finite difference hessian approximations**. *Numerical Algorithms*, p. 1–24, 2022.

[9] GRIEWANK, A. **The modification of newton's method for unconstrained optimization by bounding cubic terms**. Technical report, Technical report NA/12, 1981.

[10] KOHLER, J.; LUCCHI, A. **Sub-sampled cubic regularization for non-convex optimization**. In: *In Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, volume 70, p. 1895 – 1904, 2017.

[11] Kreyszig, E. **Introductory functional analysis with applications**. John Wiley & Sons, 1991.

[12] Melicher, V.; Haber, T.; Vanroose, W. **Fast derivatives of likelihood functionals for ode based models using adjoint-state method**. *Computational Statistics*, 32:1621–1643, 2017.

[13] Mordukhovich, B. S.; Nam, N. M. **An easy path to convex analysis and applications**. Springer, 2014.

[14] Nesterov, Y.; Polyak, B. T. **Cubic regularization of newton method and its global performance**. *Mathematical programming*, 108(1):177–205, 2006.

[15] Nesterov, Y.; others. **Lectures on convex optimization**, volume 137. Springer, 2018.

[16] Nocedal, J.; Wright, S. J. **Numerical optimization**. Springer, 1999.

[17] Stapor, P.; Fröhlich, F.; Hasenauer, J. **Optimization and profile calculation of ode models using second order adjoint sensitivity analysis**. *Bioinformatics*, 34(13):i151–i159, 2018.

[18] Villa, S.; Salzo, S.; Baldassarre, L.; Verri, A. **Accelerated and inexact forward-backward algorithms**. *SIAM Journal on Optimization*, 23(3):1607–1633, 2013.

[19] Wang, Z.; Zhou, Y.; Liang, Y.; Lan, G. **A note on inexact gradient and hessian conditions for cubic regularized Newton's method**. *Oper. Res. Lett.*, 47(2):146–149, 2019.