NEWTON METHOD WITH FEASIBLE INEXACT PROJECTIONS FOR CONSTRAINED EQUATIONS AND NONSMOOTH NEWTON METHOD IN RIEMANNIAN MANIFOLDS

Doctoral Thesis by

FABIANA RODRIGUES DE OLIVEIRA

Supervised by

Prof. Dr. **Orizon Pereira Ferreira**

Funded by
CAPES

IME - INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE FEDERAL DE GOIÁS
Goiânia, Goiás, Brazil
Março 2019

sistema de bibliotecas ufg

PRPG
PRÓ-REITORIA DE PÓS-GRADUAÇÃO

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO 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 nº 9610/98, 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.

1. Identificação do material bibliográfico: **[ ] Dissertação** **[x] Tese**

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Fabiana Rodrigues de Oliveira

Título do trabalho: Newton method with feasible inexact projections for constrained equations and nonsmooth Newton method in Riemannian manifolds

3. Informações de acesso ao documento:

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

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Fabiana Rodrigues de Oliveira
Assinatura do(a) autor(a)[2]

Ciente e de acordo:

Assinatura do(a) orientador(a)[2]

Data: 12 / 04 / 2019

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados 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.

[2] A assinatura deve ser escaneada.

Fabiana Rodrigues de Oliveira

NEWTON METHOD WITH FEASIBLE INEXACT PROJECTIONS FOR CONSTRAINED EQUATIONS AND NONSMOOTH NEWTON METHOD IN RIEMANNIAN MANIFOLDS

Tese apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Doutora em Matemática.
Área de concentração: Otimização
Orientador: Prof. Dr. Orizon Pereira Ferreira

Goiânia
2019

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

Oliveira, Fabiana Rodrigues de
Newton method with feasible inexact projections for constrained equations and nonsmooth Newton method in Riemannian Manifolds [manuscrito] / Fabiana Rodrigues de Oliveira. - 2019.
x, 71 f.: il.

Orientador: Prof. Dr. Orizon Pereira Ferreira.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2019.
Bibliografia.
Inclui gráfico, tabelas, algoritmos.

1. Newton method. 2. Feasible inexact projection. 3. Riemannian manifolds. 4. Vector fields. 5. Convergence analysis. I. Ferreira, Orizon Pereira, orient. II. Título.

CDU 517

UFG
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA
Campus Samambaia – Caixa Postal 131 – CEP: 74.001-970 – Goiânia-GO.
Fones: (62) 3521-1208 e 3521-1137 www.ime.ufg.br

ATA DA REUNIÃO DA BANCA EXAMINADORA DA DEFESA DE TESE DE FABIANA RODRIGUES DE OLIVEIRA – Ao vigésimo sétimo dia do mês de março do ano de dois mil e dezenove (27/03/2019), às 10:00 horas, reuniram-se os componentes da Banca Examinadora: Prof. Orizon Pereira Ferreira - Orientador, Prof. Leandro da Fonseca Prudente, Prof. Roberto Andreani, Prof. Gabriel Haeser e Prof. Douglas Soares Gonçalves, sob a presidência do primeiro, e em sessão pública realizada no auditório do Instituto de Matemática e Estatística, procederem a avaliação da defesa de tese intitulada: **"Newton method with feasible inexact projections for constrained equations and nonsmooth Newton method in Riemannian Manifolds"** em nível de Doutorado, área de concentração em Otimização, de autoria de Fabiana Rodrigues de Oliveira, discente do Programa de Pós-Graduação em Matemática da Universidade Federal de Goiás. A sessão foi aberta pelo Presidente da Banca, Prof. Orizon Pereira Ferreira que fez a apresentação formal dos membros da Banca. A seguir, a palavra foi concedida a autora da tese que, em 45 minutos procedeu a apresentação de seu trabalho. Terminada a apresentação, cada membro da Banca arguiu o examinando, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1513 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Matemática e procedidas às correções recomendadas, a tese foi **APROVADA** por unanimidade, considerando-se integralmente cumprido este requisito para fins de obtenção do título de **DOUTOR(A) EM MATEMÁTICA**, na área de concentração em Otimização pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega na secretaria do PPGM da versão definitiva da tese, com as devidas correções supervisionadas e aprovadas pelo orientador. Cumpridas as formalidades de pauta, às 12:00 horas a presidência da mesa encerrou esta sessão de defesa de tese e para constar eu, Ana Maria Pereira Pinto, secretária do PPGM, lavrei a presente Ata que, depois de lida e aprovada, será assinada pelos membros da Banca Examinadora em quatro vias de igual teor.

Prof. Dr. Orizon Pereira Ferreira
Presidente - IME/UFG

Prof. Dr. Leandro da Fonseca Prudente
Membro – IME/UFG

Prof. Dr. Roberto Andreani
Membro - IMECC/UNICAMP

Prof. Dr. Gabriel Haeser
Membro – IME/USP

Prof. Dr. Douglas Soares Gonçalves
Membro – DM/UFSC

FABIANA RODRIGUES DE OLIVEIRA

NEWTON METHOD WITH FEASIBLE INEXACT PROJECTIONS FOR CONSTRAINED EQUATIONS AND NONSMOOTH NEWTON METHOD IN RIEMANNIAN MANIFOLDS

Tese 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 Doutor(a) em Matemática, aprovada em 27 de março de 2019, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Orizon Pereira Ferreira
Instituto de Matemática e Estatística - UFG
Presidente da Banca

Prof. Dr. Leandro da Fonseca Prudente
Instituto de Matemática e Estatística - UFG

Prof. Dr. Roberto Andreani
Departamento de Matemática Aplicada - UNICAMP

Prof. Dr. Gabriel Haeser
Instituto de Matemática e Estatística - USP

Prof. Dr. Douglas Soares Gonçalves
Departamento de Matemática - UFSC

Dedicado a:

todas as pessoas que contribuíram para a realização deste trabalho, em especial,

Meus pais, Fernando e Dilourdes

Meus irmãos, Fernando e Fabrícia

Meu orientador, Orizon

Agradecimentos

Primeiramente, a Deus por permitir que eu realizasse mais este sonho, por se fazer presente em minha vida, por ter me dado sabedoria para transpor as dificuldades e forças para nunca desistir. Aos meus familiares, em especial, meus pais Fernando e Dilourdes e meus irmãos Fabrícia e Fernando, por serem o meu porto seguro, por me acompanhar em todos os momentos, por todas as palavras de incentivo e por sempre acreditarem em mim. Ao professor Orizon Ferreira, meus sinceros agradecimentos por todo ensinamento, pelo direcionamento e estímulo, pela confiança, pela enorme paciência e disponibilidade em esclarecer minhas dúvidas, pela dedicação no desenvolvimento de toda esta tese e por ter me aceito como orientanda. À todos os funcionários do IME/UFG pela cordialidade. Aos professores do IME/UFG que contribuíram na minha formação, em especial, Armando Corro, Jefferson Melo, Leandro Prudente, Jhone Caldeira e José Valdo. Também estendo meus agradecimentos à todos os professores do grupo de otimização desta instituição pelas valiosas contribuições nos seminários. Aos professores Leandro Prudente, Roberto Andreani, Gabriel Haeser e Douglas Gonçalves por terem aceito o convite para participarem da comissão examinadora e pelas valiosas observações e sugestões. Da mesma forma, agradeço ao professor suplente Max Gonçalves. À todos os amigos que conviveram comigo durante esses quatro anos de curso, pela companhia, pelos momentos de estudo e descontração. À CAPES pelo recurso financeiro, indispensável na realização deste trabalho. Aqueles que não foram lembrados aqui, e que contribuíram direta ou indiretamente para o desenvolvimento desta tese, meus sinceros agradecimentos.

Abstract

In this thesis, we will study three versions of the Newton method for solving problems in two contexts, namely Euclidean and Riemannian. In the Euclidean context, we will present the Newton method with feasible inexact projections for solving generalized equations subject to a set of constraints. Under local assumptions, the linear or superlinear convergence of a sequence generated by the proposed method is established. Next, a version of the inexact Newton method with feasible inexact projections for solving constrained smooth and nonsmooth equations is presented. Using suitable assumptions, the linear or superlinear convergence of a sequence generated by the method is proved. Furthermore, to illustrate the practical behavior of the proposed method, some numerical experiments are reported. Under another perspective, the last version of the Newton method to be investigated is an extension of the nonsmooth Newton method itself from the Euclidean context to the Riemannian, objecting to find a singularity of a special class of locally Lipschitz continuous vector fields. In particular, this method retrieves the classical nonsmooth Newton method to solve a system of nonsmooth equations. The well-definedness of the sequence generated by the method is ensured and the convergence analysis of the method is made under local and semi-local assumptions.

Keywords: Newton method, Feasible inexact projection, Riemannian manifolds, Vector fields, Convergence analysis.

Resumo

Nesta tese, estudaremos três versões do método de Newton para resolver problemas em dois contextos, a saber, Euclidiano e Riemanniano. No contexto Euclidiano, apresentaremos o método de Newton com projeções inexatas viáveis para resolver equações generalizadas sujeitas à um conjunto de restrições. Sob hipóteses locais, a convergência linear ou superlinear de uma sequência gerada pelo método proposto é estabelecida. Em seguida, uma versão do método de Newton inexato com projeções inexatas viáveis para resolver equações restritas diferenciáveis e não-diferenciáveis é apresentada. Usando hipóteses adequadas, a convergência linear ou superlinear de uma sequência gerada pelo método é provada. Além disso, para ilustrar o comportamento prático do método, alguns experimentos numéricos são reportados. Sob uma outra perspectiva, a última versão do método de Newton a ser investigada é uma extensão do próprio método de Newton não-diferenciável do contexto Euclidiano para o Riemanniano, objetivando encontrar uma singularidade de uma classe especial de campos de vetores localmente Lipschitz contínuos. Em particular, este método recupera o clássico método de Newton não-diferenciável para resolver um sistema de equações não-diferenciáveis. A boa definição da sequência gerada pelo método é garantida e a análise de convergência do método é feita sob hipóteses locais e semi-locais.

Palavras-chave: Médodo de Newton, Projeção inexata viável, Variedades riemannianas, Campos de vetores, Análise de convergência.

Contents

Chapter 1

Introduction

This thesis investigates the local and/or semi-local behavior of three versions of the Newton method to solve problems in two contexts. Newton method and the inexact Newton method with feasible inexact projections are proposed for solving, respectively, generalized equations, and smooth and nonsmooth equations, both subject to a set of constraints, and defined on Euclidean space. Another method studied in this thesis is the nonsmooth Newton method for finding singularities of a special class of locally Lipschitz continuous vector fields on a complete Riemannian manifold. In particular, this method can be seen as an extension to the Riemannian setting of the method studied in [76].

In Chapter 2, we recall some notations, definitions and preliminary results used throughout this thesis.

Chapter 3 addresses the Newton method with feasible inexact projections (*Newton-InexP method*) for solving generalized equations subject to a set of constraints, i.e., for solving the problem of finding $x \in \mathbb{R}^n$ such that

$$x \in C, \qquad f(x) + F(x) \ni 0, \tag{1.1}$$

where $f : \Omega \to \mathbb{R}^m$ is a continuously differentiable function, $\Omega \subseteq \mathbb{R}^n$ is an open set, $C \subset \Omega$ is a closed convex set, and $F : \Omega \rightrightarrows \mathbb{R}^m$ is a set-valued mapping with closed nonempty graph. As far as we know, this is the first time the problem (1.1) has been studied, thus being one of our contributions. However, it is worth mentioning that the applications of the Newton method and its variations for solving the problem (1.1) when $C = \mathbb{R}^n$ have been investigated in many studies, including but not limited to [2,3,28,30,36,39,40]. *Constrained Variational Inequality Problem*, see [16], and in particular, *Split Variational Inequality Problem*, see [16,52], can be stated as special cases of the constrained generalized equation (1.1). Further details are given in Section 3.3. It is known that if F is the zero mapping, i.e., $F \equiv \{0\}$, then problem (1.1) reduces to a constrained system of nonlinear equations, i.e., to solve $f(x) = 0$ such that

$x \in C$. This class of problems has been addressed in several studies, and various methods have been proposed for solving them, see, for example, [7, 10, 49, 50, 60, 66, 69].

Newton method for solving unconstrained generalized equations, i.e., when $C = \mathbb{R}^n$ in the problem (1.1), which has its origin in the work of N. H. Josephy [58], is formulated as follows. For the current iterate $x_k \in \mathbb{R}^n$, the next iterate x_{k+1} is computed as a point satisfying the following inclusion

$$f(x_k) + f'(x_k)(x - x_k) + F(x) \ni 0, \qquad k = 0, 1, \ldots, \tag{1.2}$$

where f' is the derivative of the function f. Note that at each iteration, a partially linearized inclusion at the current iterate has to be solved. The method (1.2) can be seen as a model for various iterative procedures in numerical nonlinear programming. For instance, when $F \equiv \{0\}$, this method corresponds to the usual Newton method for solving a system of nonlinear equations. If F is the product of the negative orthant in $\mathbb{R}^s$ with the origin at $\mathbb{R}^{m-s}$, i.e., $F = \mathbb{R}^s_- \times \{0\}^{m-s}$, then (1.2) becomes the Newton method for solving a system of nonlinear equalities and inequalities, see [20]. On the other hand, if $C = \mathbb{R}^n$, the problem (1.1) may represent the Karush–Kuhn–Tucker optimality conditions for a nonlinear programming problem, and then (1.2) describes the well-known sequential quadratic programming method, see [31, p. 384] and [29, 56].

Motivated by the method described above, we propose the Newton-InexP method for solving the problem (1.1). Taking into account that the Newton iterates satisfying (1.2) can be infeasible for the constraint set, a procedure is applied in order to get them back to the feasible set. In this thesis, we introduce the concept of a feasible inexact projection, which we will be adopt in the proposed methods. We remark that the concept of feasible inexact projection also accepts an exact projection, which can be adopted when it is easily obtained. For instance, the exact projections onto a box constraint or Lorentz cone are very easily obtained; see [72, p. 520] and [46, Proposition 3.3], respectively. It is worth mentioning that a feasible inexact projection on C can be computed by any method that minimize efficiently a quadratic function subject to C, by introducing a suitable error criteria. For instance, if the set C is polyhedral, then some iterations of an interior point method or active set method can be performed for finding a feasible inexact projection, see [51, 72, 87]. If C is a simple compact convex set, then the Frank-Wolfe method has been used recently to find a feasible inexact projection, see, for example, [49, 50, 61].

Our aim in Chapter 4 is to study the inexact Newton method with feasible inexact projections (*inexact Newton-InexP method*) for solving smooth and nonsmooth equations subject to a set of constraints, i.e., to find $x \in \mathbb{R}^n$ that solves the following constrained equation

$$x \in C, \qquad f(x) = 0, \tag{1.3}$$

where $f : \Omega \to \mathbb{R}^n$ is a locally Lipschitz continuous function, $\Omega \subseteq \mathbb{R}^n$ is an open set and $C \subset \Omega$ is a nonempty closed convex set. It is worth pointing out that if $C = \mathbb{R}^n$, then problem (1.3) reduces to unconstrained smooth and nonsmooth equations. If f is a continuously differentiable function, then problem (1.3) reduces to a constrained smooth equation, which can be easily found in the literature, see, for example, [8,9,44,66,70]. Besides its own importance, one of the main motivations to study constrained equations is that they appear in applications when we need to solve real-life problems, for which, only the solutions belonging to a constraint set have physical meaning. For further details, see [59]. Moreover, important problems in mathematical programming can be reformulated equivalently as a constrained nonsmooth equation, for instance, the *inequality feasibility problem*, see [75]. It is worth mentioning also that the nonlinear complementarity problem, systems of equalities and inequalities and, in particular, the Karush-Kuhn-Tucker systems can be reformulated in an appropriate manner as a constrained nonsmooth equation, see [33, 42, 43, 53, 60].

J. M. Martínez and L. Qi [67] presented a version of the inexact Newton method for solving unconstrained nonsmooth equations, i.e., the following problem

$$f(x) = 0, \tag{1.4}$$

where $f : \Omega \to \mathbb{R}^n$ is a locally Lipschitz continuous function and $\Omega \subseteq \mathbb{R}^n$ is an open set. In particular, the inexact Newton method for solving the problem (1.4) has the following formal formulation. For the current iterate $x_k \in \mathbb{R}^n$, the next iterate is any point $x_{k+1} \in \mathbb{R}^n$ satisfying the relative residual error criteria

$$\|f(x_k) + V_k(x_{k+1} - x_k)\| \leq \eta_k \|f(x_k)\|, \tag{1.5}$$

where $\eta_k \in [0, 1)$ is the relative residual error tolerance and V_k is an element of the Clarke generalized Jacobian of f at x_k. For a definition of the Clarke generalized Jacobian, see Definition 2.1.7, which was presented by F. H. Clarke [18]. More versions of inexact Newton-type methods for solving the problem (1.4) include, but are not limited to, those in [13, 14, 34, 80, 81].

The problem (1.3) has been addressed in several studies, and several similar methods and/or variants of (1.5) have been proposed for solving it. See, for example, the exact/inexact Newton-like methods in [49, 50, 66], projected Levenberg–Marquardt-type methods in [6, 7], and trust-region methods in [8, 9]. In particular, the method proposed in [7] combines a Levenberg–Marquardt-type method with an inexact projection, which also accepts an infeasible inexact projection. In the present thesis, we propose a scheme for solving the problem (1.3), which we call the inexact Newton-InexP method, that also uses the concept of inexact projection. However, inexact projections used in this scheme are always feasible. In essence, the proposed method combines the inexact Newton method (1.5) with a procedure to obtain feasible inexact projections onto a set C and thus to ensure the feasibility of the iterates. An issue to consider is the inexact solution in (1.5), which has an advantage over the exact solution, see [25]. This advantage appears more explicitly in practical implementations of the method, because finding an exact solution of linear approximations of equation (1.4) can be computationally expensive for large-scale problems. Thus, in the present thesis, we consider that from the current iterate, the next iterate is any point in C satisfying the relative residual error criteria (1.5). We remark that if $C = \mathbb{R}^n$, the inexact Newton-InexP method becomes the classical inexact Newton method applied for solving unconstrained smooth and nonsmooth equations. From the theoretical viewpoint, i.e., in the convergence analysis presented, to guarantee local efficiency of the proposed method, we assume appropriate assumptions, such as regularity and semismoothness. Under the regularity assumption, we ensure that a sequence generated by the method is well-defined. The semismoothness assumption is of particular interest owing to the key role it plays in the convergence of our method; in particular, this property is essential for fast local convergence. To illustrate the robustness and efficiency of our method, we present some preliminary numerical experiments of the proposed method for solving constrained absolute value equation (CAVE). We also compare the performance of the proposed method with the inexact Newton method with feasible exact projections.

Chapter 5 presents the nonsmooth Newton method for finding a singularity of a special class of vector fields defined on a complete Riemannian manifold M, i.e., for finding a point $p \in M$ satisfying the equation

$$X(p) = 0, \tag{1.6}$$

where X is a locally Lipschitz continuous vector field defined on M.

It is well-known that the Newton method is the most popular method for finding a singularity of a differentiable vector field. Its origins go back to the work of M. Shub [79]; see also [47, 64, 82, 85, 88]. This method became popular owing to its attractive convergence properties under suitable assumptions. For instance, in the previously cited works, the (superlinear and/or quadratic) local convergence of a sequence generated by the Newton method has been established under the invertibility assumption of the covariant derivative of the vector field at its singularity, and/or Lipschitz-like conditions on the covariant derivative of the vector field. Recently, in [35] were established local properties of the Newton method under the invertibility assumption of the covariant derivative of the vector field at its singularity. Basically, in the Newton method the vector field is replaced by an approximation depending on the current iterate, and then the original problem is converted in an approximated problem, which can be solved more easily. The solution of this approximated problem is then taken as a new iterate and the process is repeated.

The success of the Newton method for finding a singularity of a differentiable vector field has motivated us to propose and analyze the nonsmooth Newton method for finding a singularity of a locally Lipschitz continuous vector field. To present our method, we first generalize some results of nonsmooth analysis, from the Euclidean context to the Riemannian setting. In particular, we discuss the concept and main properties of the locally Lipschitz continuous vector fields defined on complete Riemannian manifolds, such as the Clarke generalized covariant derivative and Rademacher theorem. In particular, this derivative can be viewed as a natural generalization to Riemannian setting of the Clarke generalized Jacobian. The concept of the Clarke generalized covariant derivative has already appeared in [48, 77]. In this thesis, we show its existence using a version of Rademacher theorem in the Riemannian setting, which is one of our contributions. In the following, we introduce in the Riemannian settings an important subclass of locally Lipschitz continuous vector fields, namely the semismooth and μ-order semismooth vector fields. As well as in the Euclidean context, these concepts play an important role in the convergence analysis of our method. The essence of the nonsmooth Newton method is similar to the classical case, however, in the approximated problem, we combine the exponential mapping on the manifold with an element of the Clarke generalized covariant derivative of the vector field. This is because the covariant derivative of a locally Lipschitz continuous vector field may not exist. It is worth pointing out that, when the vector field is continuously differentiable, our method reduces to the classical Newton method. From the theoretical viewpoint, we present a local and semi-local convergence analysis of the proposed method under mild assumptions.

We finish this thesis with some remarks and future work in Chapter 6. It is worth mentioning that the results of this thesis gave rise to three scientific papers. One of which is already published, namely [23], and the other two are under review, namely [21, 22]. The cited papers have been submitted to important journals of international circulation in the area of optimization.

Chapter 2

Notations and preliminary results

In this chapter, we review some notations, definitions and preliminary results used throughout this text. Initially, we recall some concepts of the Euclidean space and in the sequence, we discuss some basic concepts of the Riemannian geometry.

2.1 Euclidean space

The following notations, definitions, and results are used throughout of Chapters 3 and 4. For further details, see [18, 31, 34]. We begin with some concepts of analysis and of a set-valued mapping.

The *open* and *closed balls* of radius $\delta > 0$, centered at x are defined respectively by

$$B_\delta(x) := \{y \in \mathbb{R}^n : \ \|x - y\| < \delta\}, \qquad B_\delta[x] := \{y \in \mathbb{R}^n : \ \|x - y\| \leq \delta\}.$$

The *vector space consisting of all continuous linear mappings* $A : \mathbb{R}^n \to \mathbb{R}^m$ is denoted by $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, and the *norm* of A is defined by

$$\|A\| := \sup\{\|Ax\| : \ \|x\| \leq 1\}.$$

Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}^m$ be a differentiable function at all $x \in \Omega$. Then, the *derivative* of f at x is the linear mapping $f'(x) : \mathbb{R}^n \to \mathbb{R}^m$, which is continuous. The *graph* of the set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is the set

$$\operatorname{gph} F := \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^m : \ u \in F(x)\}.$$

The *domain* and *range* of the set-valued mapping F are, respectively, the sets

$$\operatorname{dom} F := \{x \in \mathbb{R}^n : \ F(x) \neq \varnothing\}, \qquad \operatorname{rge} F := \{u \in \mathbb{R}^m : \ u \in F(x) \text{ for some } x \in \mathbb{R}^n\}.$$

The *inverse* of the mapping F is the set-valued mapping $F^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ defined by $F^{-1}(u) := \{x \in \mathbb{R}^n : \; u \in F(x)\}$ and the *partial linearization* of $f + F$ at $x \in \Omega$ is the set-valued mapping $L_{f+F}(x, \cdot) : \Omega \rightrightarrows \mathbb{R}^m$ defined by

$$L_{f+F}(x, y) := f(x) + f'(x)(y - x) + F(y). \tag{2.1}$$

For the sets C and D in $\mathbb{R}^n$, the *distance from x to D* and the *excess of C beyond D* are defined respectively by

$$d(x, D) := \inf_{y \in D} \|x - y\|, \qquad e(C, D) := \sup_{x \in C} d(x, D), \tag{2.2}$$

where the convention is adopted that $d(x, D) = +\infty$ when $D = \varnothing$, $e(\varnothing, D) = 0$ when $D \neq \varnothing$, and $e(\varnothing, \varnothing) = +\infty$. In the following, we present the notion of metric regularity, which plays an important role in Chapter 3.

Definition 2.1.1 *Let $\Omega \subset \mathbb{R}^n$ be an open set. A set-valued mapping $G : \Omega \rightrightarrows \mathbb{R}^m$ is said to be metrically regular at $\bar{x} \in \Omega$ for $\bar{u} \in \mathbb{R}^m$ when $\bar{u} \in G(\bar{x})$, the graph of G is locally closed at $(\bar{x}, \bar{u})$, and there exist constants $\kappa \geq 0$, $a > 0$, and $b > 0$ such that $B_a[\bar{x}] \subset \Omega$ and*

$$d(x, G^{-1}(u)) \leq \kappa \, d(u, G(x)), \qquad \forall \; (x, u) \in B_a[\bar{x}] \times B_b[\bar{u}].$$

Moreover, if the mapping $B_b[\bar{u}] \ni u \mapsto G^{-1}(u) \cap B_a[\bar{x}]$ is single-valued, then G is called strongly metrically regular at $\bar{x} \in \Omega$ for $\bar{u} \in \mathbb{R}^m$, with associated constants $\kappa \geq 0$, $a > 0$, and $b > 0$.

When the mapping $B_b[\bar{u}] \ni u \mapsto G^{-1}(u) \cap B_a[\bar{x}]$ in Definition 2.1.1 is single-valued, then for the sake of simplicity we hereafter adopt the notation $w = G^{-1}(u) \cap B_a[\bar{x}]$ instead of $\{w\} = G^{-1}(u) \cap B_a[\bar{x}]$.

Remark 2.1.2 If G is strongly metrically regular at $\bar{x} \in \Omega$ for $\bar{u} \in \mathbb{R}^m$ with constants $\kappa \geq 0$, $a > 0$, and $b > 0$, then the mapping $B_b[\bar{u}] \ni u \mapsto G^{-1}(u) \cap B_a[\bar{x}]$ is single-valued and Lipschitz continuous on $B_b[\bar{u}]$ with Lipschitz constant κ [31, Proposition 3G.1, p. 193], i.e.,

$$\left\| G^{-1}(u) \cap B_a[\bar{x}] - G^{-1}(v) \cap B_a[\bar{x}] \right\| \leq \kappa \|u - v\|, \qquad \forall \; u, v \in B_b[\bar{u}].$$

Next, we present a generalization of the contraction mapping principle for set-valued mappings. For a prove of this, see [31, Theorem 5E.2, p. 313].

Theorem 2.1.3 *Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping and let $\bar{x} \in \mathbb{R}^n$. Suppose that there exist scalars $\rho > 0$ and $\lambda \in (0, 1)$ such that the set $\operatorname{gph} \Phi \cap (B_\rho[\bar{x}] \times B_\rho[\bar{x}])$ is closed and the following conditions hold:*

(i) $d(\bar{x}, \Phi(\bar{x})) \leq \rho(1-\lambda)$*;*

(ii) $e\left(\Phi(p) \cap B_\rho[\bar{x}], \Phi(q)\right) \leq \lambda\|p-q\|$ *for all* $p, q \in B_\rho[\bar{x}]$*.*

Then, Φ *has a fixed point in* $B_\rho[\bar{x}]$*. That is, there exists* $y \in B_\rho[\bar{x}]$ *such that* $y \in \Phi(y)$*.*

In the following, we define the concepts of locally Lipschitz continuous and directionally differentiable functions, which plays an important role in our study, more specifically, in Chapter 4.

Definition 2.1.4 *A function* $f : \Omega \to \mathbb{R}^m$ *is said to be Lipschitz continuous on a set* $\Omega \subseteq \mathbb{R}^n$*, if there is a constant* $L > 0$ *such that*

$$\|f(x) - f(y)\| \leq L\|x - y\|, \qquad \forall\ x, y \in \Omega.$$

Given $x \in \Omega$*, if there exists* $\delta > 0$ *such that* f *is Lipschitz continuous on* $B_\delta(x)$*, then* f *is said to be Lipschitz continuous at* x*. Moreover, if for all* $x \in \Omega$*,* f *is Lipschitz continuous at* x*, then* f *is said to be locally Lipschitz continuous on* Ω*.*

Remark 2.1.5 According to Rademacher theorem, see [32, Theorem 2, p. 81], locally Lipschitz continuous functions are differentiable almost everywhere.

Definition 2.1.6 *Let* $\Omega \subseteq \mathbb{R}^n$ *be an open set. The directional derivative of a function* $f : \Omega \to \mathbb{R}^m$ *at* $x \in \Omega$ *in the direction* $h \in \mathbb{R}^n$ *is defined by*

$$f'(x; h) := \lim_{t \to 0^+} \frac{f(x + th) - f(x)}{t},$$

whenever the limit exists. If $f'(x; h)$ *exists for every* h*, then* f *is said to be directionally differentiable at* x*.*

We end this section by defining the Clarke generalized Jacobian of a function, which has appeared in [18]. This Jacobian requires only local Lipschitz continuity of the function f and its well-definedness is ensured by Rademacher theorem.

Definition 2.1.7 *The Clarke generalized Jacobian of a locally Lipschitz continuous function* f *at* x *is a set-valued mapping* $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *defined as*

$$\partial f(x) := \operatorname{co}\left\{ H \in \mathbb{R}^{m \times n} :\ \exists\, \{x_k\} \subset \mathcal{D}_f,\ \lim_{k \to +\infty} x_k = x,\ H = \lim_{k \to +\infty} f'(x_k) \right\},$$

*where "*co*" represents the convex hull,* $\mathbb{R}^{m \times n}$ *is the set consisting of all* $m \times n$ *matrices, and* $\mathcal{D}_f$ *denotes the set of points at which* f *is differentiable.*

Remark 2.1.8 It is worth mentioning that if f is continuously differentiable at x, then $\partial f(x) = \{f'(x)\}$. Otherwise, $\partial f(x)$ could contain other elements different from $f'(x)$, even if f is differentiable at x, see [18, Example 2.2.3, p. 33]. Furthermore, the Clarke generalized Jacobian is a subset of $\mathbb{R}^{m\times n}$ nonempty, convex, compact in the usual sense. We also remind that the set-valued mapping ∂f is closed and upper semicontinuous, see [18, Proposition 2.6.2, p. 70].

2.2 Riemannian geometry

In this section, we recall some notations, definitions and basic properties of Riemannian manifolds used throughout of Chapter 5. They can be found in many books on Riemannian geometry, see, for example, [62, 78, 84]. We begin this section by presenting the concepts of charts and smoothness of a mapping defined between manifolds.

Definition 2.2.1 *A chart on a n-dimensional smooth manifold M is a pair (U, φ), where U is an open subset of M and the coordinate mapping $\varphi : U \to \widehat{U}$ is a smooth homeomorphism from U to an open subset $\widehat{U} = \varphi(U) \subseteq \mathbb{R}^n$.*

Definition 2.2.2 *Let N and M be manifolds of finite dimension and $F : N \to M$ be a continuous mapping. We say that F is smooth at $p \in N$, if there exist smooth charts (U, φ) containing p and (W, ψ) containing $F(p)$ such that $F(U) \subseteq W$ and the composite mapping $\psi \circ F \circ \varphi^{-1} : \varphi(U) \to \psi(W)$ is smooth at $\varphi(p)$.*

Remark 2.2.3 The definition of smoothness of a mapping $F : N \to M$ at a point is independent of the choice of charts, see [84, Proposition 6.7, p. 61]. A *diffeomorphism* of manifolds is a bijective smooth mapping $F : N \to M$ whose inverse F^{-1} is also smooth. According to [84, Proposition 6.10, p. 63] coordinate mappings are diffeomorphisms and, in particular, are continuously differentiable.

Let M be a n-dimensional smooth Riemannian manifold with *Riemannian metric* denoted by $\langle\cdot,\cdot\rangle$ and the corresponding *norm* by $\|\cdot\|$. The *length* of a piecewise smooth curve $\gamma : [a, b] \to M$ joining p to q in M, i.e., $\gamma(a) = p$ and $\gamma(b) = q$ is defined by

$$\ell(\gamma) := \int_b^a \|\gamma'(t)\|\, dt.$$

The *Riemannian distance* between p and q is defined as $d(p, q) = \inf_{\gamma\in\Gamma_{p,q}} \ell(\gamma)$, where $\Gamma_{p,q}$ is the set of all piecewise smooth curves in M joining points p and q. This distance induces the original topology on M, namely (M, d) is a complete metric space and the bounded and closed subsets are compact. The *open and closed balls* of radius $r > 0$, centred at p are defined respectively by

$$B_r(p) := \{q \in M : \ d(p,q) < r\}, \qquad\qquad B_r[p] := \{q \in M : \ d(p,q) \leq r\}.$$

Denote the *tangent space* at point p by T_pM, the *tangent bundle* by $TM := \bigcup_{p\in M} T_pM$ and a *vector field* by a mapping $X : M \to TM$ such that $X(p) \in T_pM$. Let γ be a curve joining the points p and q in M, and let ∇ be the Levi-Civita connection associated to $(M, \langle\cdot,\cdot\rangle)$. For each $t \in [a,b]$, ∇ induces a linear isometry between the tangent spaces $T_{\gamma(a)}M$ and $T_{\gamma(t)}M$, relative to $\langle\cdot,\cdot\rangle$, defined by $P_{\gamma,a,t}v = Y(t)$, where Y is the unique vector field on γ such that $\nabla_{\gamma'(t)}Y(t) = 0$ and $Y(a) = v$. This isometry is called *parallel transport* along the segment γ joining $\gamma(a)$ to $\gamma(t)$. It can be showed that $P_{\gamma,b_1,b_2} \circ P_{\gamma,a,b_1} = P_{\gamma,a,b_2}$ and $P_{\gamma,b,a} = P^{-1}_{\gamma,a,b}$. For simplicity and convenience, whenever there is no confusion, we consider the notation $P_{\gamma,p,q}$ instead of $P_{\gamma,a,b}$, where γ is a segment joining p to q with $\gamma(a) = p$ and $\gamma(b) = q$. We use the short notation P_{pq} instead of $P_{\gamma,p,q}$ whenever there exists an unique geodesic segment joining p to q.

Remark 2.2.4 For any n-dimensional smooth manifold M; the tangent bundle TM has a natural topology and smooth structure that make it into a $2n$-dimensional smooth manifold. With respect to this structure, the projection $\pi : TM \to M$ is smooth, see [63, Proposition 3.18, p. 66].

The *standard Riemannian distance d_{TM} on the tangent bundle TM* can be defined as follows: given $u, v \in TM$, then d_{TM} is defined by

$$d_{TM}(u,v) := \inf\left\{\sqrt{\ell^2(\gamma) + \|P_{\gamma,\pi u,\pi v}u - v\|^2} : \ \gamma \in \Gamma_{\pi u,\pi v}\right\}, \tag{2.3}$$

where $\Gamma_{\pi u,\pi v}$ is the set of all piecewise smooth curves in M joining the points πu to πv, whose derivative is never zero, see [15, Appendix, p. 240].

A vector field Y along a smooth curve γ in M is said to be *parallel* when $\nabla_{\gamma'}Y = 0$. If γ' itself is parallel, we say that γ is a *geodesic*. The geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second-order nonlinear ordinary differential equation, so the geodesic γ is determined by its position p and velocity v at p. It is easy to check that $\|\gamma'\|$ is constant. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. A geodesic segment joining p to q in M is said to be *minimal* if its length is equal to $d(p,q)$ and, in this case, it will be denoted by γ_{pq}. A Riemannian manifold is *complete* if its geodesics $\gamma(t)$ are defined for any value of $t \in \mathbb{R}$. The Hopf-Rinow theorem asserts that any pair of points in a complete Riemannian manifold M can be joined by a (not necessarily unique) minimal geodesic segment.

From now on, *M denotes a n-dimensional smooth and complete Riemannian manifold.* Owing to the completeness of the Riemannian manifold M, the *exponential mapping* at p,

$\exp_p : T_pM \to M$ can be given by $\exp_p v = \gamma(1)$, where γ is the geodesic defined by its position p and velocity v at p and $\gamma(t) = \exp_p(tv)$ for any value of t. The inverse of the exponential mapping (if exists) is denote by $\exp_p^{-1}$. Let $p \in M$, the *injectivity radius* of M at p is defined by

$$r_p := \sup \left\{ r > 0 : \ \exp_{p_{|_{B_r(0_p)}}} \text{ is a diffeomorphism} \right\},$$

where 0_p denotes the origin of the T_pM and $B_r(0_p) := \{v \in T_pM : \ \|v - 0_p\| < r\}$. A neighborhood $\mathcal{W}$ of $p \in M$ is said to be *normal neighborhood* of p if there exists a neighborhood $\mathcal{U}$ of the origin in T_pM such that $\exp_p : \mathcal{U} \to \mathcal{W}$ is a diffeomorphism. Furthermore, if $\mathcal{W}$ is a normal neighborhood of each of its points, then $\mathcal{W}$ is said to be *totally normal neighborhood*.

Remark 2.2.5 For $\bar{p} \in M$, the above definition implies that if $0 < \delta < r_{\bar{p}}$, then $\exp_{\bar{p}} B_\delta(0_{\bar{p}}) = B_\delta(\bar{p})$ is a totally normal neighborhood. Hence, for all $p, q \in B_\delta(\bar{p})$, there exists a unique geodesic segment γ joining p to q, which is given by $\gamma_{pq}(t) = \exp_p(t \exp_p^{-1} q)$, for all $t \in [0, 1]$ and $d(p, q) = \| \exp_p^{-1} q\|$.

In the following, we present a quantity, which plays an important role in Chapter 5, it was defined in [24].

Definition 2.2.6 *Let $p \in M$ and r_p be the radius of injectivity of M at p. Define the quantity K_p by*

$$K_p := \sup \left\{ \frac{d(\exp_q u, \exp_q v)}{\|u - v\|} : \ q \in B_{r_p}(p), \ u, v \in T_qM, \ u \neq v, \ \|v\| \leq r_p, \ \|u - v\| \leq r_p \right\}.$$

In the following remark, we show that an estimative for the value of K_p can be found for Riemannian manifolds with non-negative sectional curvature.

Remark 2.2.7 The number K_p measures how fast the geodesics spread apart in M. In particular, when $u = 0$ or more generally when u and v are on the same line through 0, $d(\exp_q u, \exp_q v) = \|u - v\|$. Hence, $K_p \geq 1$, for all $p \in M$. When M has non-negative sectional curvature, the geodesics spread apart less than the rays [26, Chapter 5], i.e., $d(\exp_p u, \exp_p v) \leq \|u - v\|$ and, in this case, $K_p = 1$ for all $p \in M$.

The *directional derivative* of X at p in the direction $v \in T_pM$ is defined by

$$\nabla X(p, v) := \lim_{t \to 0^+} \frac{1}{t} \left[P_{\exp_p(tv)p} X(\exp_p(tv)) - X(p) \right] \in T_pM, \tag{2.4}$$

whenever the limit exists, where $P_{\exp_p(tv)p}$ denotes the parallel transport along $\gamma(t) = \exp_p(tv)$. If this directional derivative exists for every v, then X is said to be *directionally* *differentiable* at p. Denote by $\mathcal{X}(M)$ the space of the differentiable vector fields on M. For each $X \in \mathcal{X}(M)$, the *covariant derivative* of X determined by the Levi-Civita connection ∇ defines at each $p \in M$ a linear mapping $\nabla X(p) : T_pM \to T_pM$ given by

$$\nabla X(p)v := \nabla_Y X(p),$$

where Y is a vector field such that $Y(p) = v$. Furthermore, $\nabla X(p, v) = \nabla X(p)v$, see [83, Proposition 3, p. 234]. To state the following result, we need to define the norm of a linear mapping.

Definition 2.2.8 *Let $p \in M$, the norm of a linear mapping $A : T_pM \to T_pM$ is defined by*

$$\|A\| := \sup \{\|Av\| : \ v \in T_pM, \ \|v\| = 1\}.$$

We end this section with the well-known Banach lemma. For a prove of it see [73, Lemma 2.3.2, p. 45].

Lemma 2.2.9 *Let A, B be linear operators in T_pM. If A is nonsingular and holds $\|A^{-1}\|\|B - A\| < 1$, then B is nonsingular, and*

$$\|B^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}(B - A)\|}.$$

Chapter 3

Newton method with feasible inexact projections for solving constrained generalized equations

In this chapter, we propose a version of the Newton method for solving constrained generalized equations, i.e., for solving the problem (1.1). Basically, the proposed method can be seen as a combination of the classical Newton method applied for solving unconstrained generalized equations with a procedure to obtain a feasible inexact projection. Our goal is to present an analysis of the behavior of a sequence generated by this method. For this purpose, using the contraction mapping principle, we establish a local convergence analysis of our method under appropriate assumptions, namely metric regularity or strong metric regularity and Lipschitz continuity. Furthermore, concrete examples of constrained generalized equations are presented.

3.1 Newton-InexP method and its convergence analysis

In this section, we present the Newton-InexP method for solving the problem (1.1). We also study the local convergence of a sequence generated by this method. Our analysis is performed under assumptions of metric regularity and strong metric regularity for an approximation of the set-valued mapping $f + F$ and assuming the Lipschitz continuity of the derivative f'. To ensure the feasibility of the Newton iterates, our method incorporates a procedure to obtain a feasible inexact projection onto the feasible set. Next, we introduce the concept of a feasible inexact projection, which will play an important role throughout the thesis.

Definition 3.1.1 *Let $C \subset \mathbb{R}^n$ be a closed convex set, $x \in C$, and $\theta \geq 0$. The feasible inexact projection mapping relative to x with error tolerance θ, denoted by $P_C(\cdot, x, \theta) : \mathbb{R}^n \rightrightarrows C$, is the set-valued mapping defined as follows:*

$$P_C(y, x, \theta) := \left\{ w \in C : \ \langle y - w, z - w \rangle \leq \theta \|y - x\|^2, \quad \forall\, z \in C \right\}.$$

Each point $w \in P_C(y, x, \theta)$ is called a feasible inexact projection of y onto C with respect to x and with error tolerance θ.

Since $C \subset \mathbb{R}^n$ is a closed convex set, [12, Proposition 2.1.3, p. 201] implies that for each $y \in \mathbb{R}^n$ we have $P_C(y) \in P_C(y, x, \theta)$, where P_C denotes the exact projection mapping onto C. Therefore, $P_C(y, x, \theta) \neq \varnothing$ for all $y \in \mathbb{R}^n$ and $x \in C$. If $\theta = 0$ in Definition 3.1.1, then $P_C(y, x, 0) = \{P_C(y)\}$ for all $y \in \mathbb{R}^n$ and $x \in C$. *We use $P_C(y, x, 0) = P_C(y)$ instead of* $P_C(y, x, 0) = \{P_C(y)\}$.

Remark 3.1.2 It is worth mentioning that the concept of inexact projection has been considered before; see, for example, [7]. However, in general, those inexact projections are infeasible and, thus, different from the above concept.

Conditional gradient procedure (*CondG procedure*); see, for instance, [45, 61], which is based on Frank-Wolfe method, is an example of the procedure for obtaining feasible inexact projections onto special compact sets C. For a general overview of this method, see [12]. For the sake of completeness, we present the CondG procedure in the following. For this, we assume the existence of a linear optimization oracle (or simply LO oracle) capable of minimizing linear functions over the constraint set C. Next, we formally describe the CondG procedure algorithm with $y \in \mathbb{R}^n$, $x \in C$ and $\epsilon \geq 0$ as the input data.

Algorithm 3.1.3 CondG procedure $x_+ = \mathrm{CondG}\,(y, x, \epsilon)$

Step 0. Set $w_1 = x$ and $k = 1$.

Step 1. Use a LO oracle to compute an optimal solution z_k and the optimal value g_k^* as

$$z_k := \arg\min_{z \in C} \langle y - w_k,\, z - w_k \rangle, \qquad g_k^* := \langle y - w_k,\, z_k - w_k \rangle. \tag{3.1}$$

Step 2. If $g_k^* \leq \epsilon$, set $x_+ := w_k$ and **stop**; otherwise, compute $\alpha_k \in (0, 1]$ and w_{k+1} as

$$\alpha_k := \min \left\{ 1, \frac{g_k^*}{\|z_k - w_k\|^2} \right\}, \qquad w_{k+1} := w_k + \alpha_k (z_k - w_k). \tag{3.2}$$

Step 3. Set $k \leftarrow k+1$, and go to **Step 1**.

In the following, we describe the main features of the CondG procedure; for further details, see [61].

Remark 3.1.4 Let $y \in \mathbb{R}^n$ and $\varphi : \mathbb{R}^n \to \mathbb{R}$ be defined by $\varphi(z) := \|y - z\|^2/2$. It is worth mentioning that the above CondG procedure can be viewed as a specialized version of the classic conditional gradient method, see [12, p. 215], applied to problem $\min_{z \in C} \varphi(z)$. In particular, it can be easily seen in (3.1) that z_k is equivalent to $\min_{z\in C}\langle \varphi'(w_k),\, z - w_k\rangle$. The stepsize α_k given in (3.2) is obtained using exact minimization, i.e., $\alpha_k = \arg\min_{\alpha \in [0,1]} \varphi(w_k + \alpha(z_k - w_k))$. Note that, if the CondG procedure computes the optimal value g_k^* satisfying $g_k^* \le \epsilon$, then it obtains $w_k \in C$ and the procedure terminates. Otherwise, computes α_k, which is well-defined and belongs to $(0, 1]$ due to $g_k^* > \epsilon \ge 0$. Since z_k, $w_k \in C$ we have $w_{k+1} \in C$, thus the CondG procedure generates a sequence in C. Finally, note that $\mathrm{CondG}\,(y, x, 0) = P_C(y)$, and therefore $x_+ = \mathrm{CondG}\,(y, x, \epsilon)$ can be seen as an approximation of the projection $P_C(y)$ onto C.

The next result plays an important role in the subsequent analysis, in particular, for this and the next chapter. It presents a basic property of the feasible inexact projection, the proof is similar to [50, Lemma 4]. For the sake of completeness, we decide to present the proof here.

Lemma 3.1.5 *Let $y, \tilde{y} \in \mathbb{R}^n$, $x, \tilde{x} \in C$, and $\theta \ge 0$. Then, for any $w \in P_C(y, x, \theta)$, we have*

$$\|w - P_C(\tilde{y}, \tilde{x}, 0)\| \le \|y - \tilde{y}\| + \sqrt{2\theta}\|y - x\|.$$

Proof. To simplify the notation, we set $\tilde{w} = P_C(\tilde{y}, \tilde{x}, 0)$, and take $w \in P_C(y, x, \theta)$. First, note that $\|y - \tilde{y}\|^2 = \|(y - w) - (\tilde{y} - \tilde{w})\|^2 + \|w - \tilde{w}\|^2 + 2\langle (y - \tilde{y}) - (w - \tilde{w}), w - \tilde{w}\rangle$, which implies that

$$\|w - \tilde{w}\|^2 \le \|y - \tilde{y}\|^2 + 2\langle y - w, \tilde{w} - w\rangle + 2\langle \tilde{y} - \tilde{w}, w - \tilde{w}\rangle.$$

Because $\tilde{w} = P_C(\tilde{y}, \tilde{x}, 0)$ and $w \in P_C(y, x, \theta)$, by using Definition 3.1.1 and the fact that $\tilde{w}, w \in C$, we can conclude that

$$\langle y - w,\, \tilde{w} - w\rangle \le \theta\|y - x\|^2, \qquad \langle \tilde{y} - \tilde{w},\, w - \tilde{w}\rangle \le 0.$$

Thus, the combination of these three previous inequalities yields

$$\|w - \tilde{w}\|^2 \le \|y - \tilde{y}\|^2 + 2\theta\|y - x\|^2,$$

and then $\|w - \tilde{w}\| \leq \|y - \tilde{y}\| + \sqrt{2\theta}\|y - x\|$, giving the desired inequality. ■

The conceptual Newton-InexP algorithm, for solving the problem (1.1), with $x_0 \in C$ and $\{\theta_k\} \subset [0, +\infty)$ as the input data, is formally described as follows.

Algorithm 3.1.6 Newton-InexP method

Step 0. Let $x_0 \in C$ and $\{\theta_j\} \subset [0, +\infty)$ be given, and set $k = 0$.

Step 1. If $f(x_k) + F(x_k) \ni 0$, then **stop**; otherwise, compute $y_k \in \mathbb{R}^n$ such that

$$f(x_k) + f'(x_k)(y_k - x_k) + F(y_k) \ni 0. \tag{3.3}$$

Step 2. If $y_k \in C$, set $x_{k+1} = y_k$; otherwise, use a procedure to obtain $P_C(y_k, x_k, \theta_k) \in C$ a feasible inexact projection of y_k onto C relative to x_k with relative error tolerance θ_k; and set

$$x_{k+1} \in P_C\left(y_k, x_k, \theta_k\right). \tag{3.4}$$

Step 3. Set $k \leftarrow k + 1$, and go to **Step 1**.

Remark 3.1.7 In **Step 1** of Algorithm 3.1.6, we check if the current iterate x_k is a solution of problem (1.1). Otherwise, we compute a point y_k satisfying the inclusion (3.3). Since the point y_k in **Step 1** may be infeasible for the constraint set C, the Newton-InexP method applies a procedure to obtain a feasible inexact projection, and consequently the new iterate x_{k+1} on C. In particular, the point x_{k+1} obtained in (3.4) is an approximate feasible solution for the projection subproblem $\min_{z \in C}\{\|z - y_k\|^2/2\}$, satisfying the condition $\langle y_k - x_{k+1}, z - x_{k+1}\rangle \leq \theta_k \|y_k - x_k\|^2$ for any $z \in C$. As we will see, the specific choice of the tolerance θ_k is essential to establish the local convergence of the Newton-InexP method. Whenever $F \equiv \{0\}$ and the CondG procedure is used to obtain $x_{k+1} \in P_C\left(y_k, x_k, \theta_k\right)$, the Algorithm 3.1.6 reduces to the one proposed in [50].

In the following, we state our main theorem for the Newton-InexP method. The proof constitutes a combination of the results that will be studied in the sequel.

Theorem 3.1.8 *Let $\Omega \subset \mathbb{R}^n$ be an open set, $f : \Omega \to \mathbb{R}^m$ be continuously differentiable in Ω, and $F : \Omega \rightrightarrows \mathbb{R}^m$ be a set-valued mapping with closed graph. Assume that $C \subset \Omega$ is a closed convex set, $x_* \in C$, $f(x_*) + F(x_*) \ni 0$, there exists $L > 0$ such that*

$$\|f'(x) - f'(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in \Omega, \tag{3.5}$$

and the set-valued mapping $\Omega \ni y \mapsto L_{f+F}(x_*, y)$ *is metrically regular at* x_* *for* 0, *with constants* $\kappa > 0$, $a > 0$, *and* $b > 0$. *Let* $r := \sup\{t \in \mathbb{R} : \ B_t(x_*) \subset \Omega\}$, $\{\theta_k\} \subset [0, 1/2)$ *and*

$$r_* := \min\left\{r, \ \frac{2\left(1-\sqrt{2\tilde{\theta}}\right)}{\left(3-\sqrt{2\tilde{\theta}}\right)\kappa L}, \ a, \ \sqrt{\frac{2b}{3L}}\right\}, \qquad \tilde{\theta} := \sup_k \theta_k < \frac{1}{2}. \tag{3.6}$$

Then, for every $x_0 \in C \cap B_{r_*}(x_*)\backslash\{x_*\}$, *there exists a sequence* $\{x_k\}$ *generated by the Newton-InexP method, associated to* $\{\theta_k\}$ *and starting at* x_0, *which is contained in* $B_{r_*}(x_*) \cap C$ *and converges to* x_* *with the following rate of convergence:*

$$\|x_{k+1} - x_*\| \leq \left[\left(1+\sqrt{2\theta_k}\right)\frac{\kappa L\|x_k - x_*\|}{2(1-\kappa L\|x_k - x_*\|)} + \sqrt{2\theta_k}\right]\|x_k - x_*\|, \tag{3.7}$$

for all $k = 0, 1, \ldots$. *As a consequence, if* $\lim_{k\to+\infty}\theta_k = 0$ *then* $\{x_k\}$ *converges to* x_* *superlinearly. In particular, if* $\theta_k = 0$ *for all* $k = 0, 1 \ldots$, *then*

$$\|x_{k+1} - x_*\| \leq \frac{3\kappa L}{2}\|x_k - x_*\|^2, \qquad k = 0, 1, \ldots, \tag{3.8}$$

and $\{x_k\}$ *converges to* x_* *quadratically. Furthermore, if the mapping* $L_{f+F}(x_*, \cdot)$ *is strongly metrically regular at* x_* *for* 0, *then* x_* *is the unique solution of* (1.1) *in* $B_{r_*}(x_*)$, *and every sequence generated by the Newton-InexP method associated to* $\{\theta_k\}$ *and starting at* $x_0 \in C \cap B_{r_*}(x_*)\backslash\{x_*\}$ *satisfies* (3.7) *and converges to* x_*.

Remark 3.1.9 In particular, (3.7) implies that $\limsup_{k\to+\infty}[\|x_{k+1} - x_*\|/\|x_k - x_*\|] \leq \sqrt{2\hat{\theta}}$, where $\hat{\theta} = \limsup_{k\to+\infty}\theta_k$. Note that if $C = \mathbb{R}^n$, then $\theta_k \equiv 0$, and using [31, Theorem 3E.7, p. 178], we can conclude with some adjustment that Theorem 3.1.8 reduces to [28, Theorem 1]. If $y_k \in C$ in the Newton-InexP method for all $k = 0, 1, \ldots$, then the procedure to obtain a feasible inexact projection plays no role. In this case, the convergence rate is quadratic, as in (3.8).

Henceforth, we assume that all the assumptions of Theorem 3.1.8 hold except the strong metric regularity, which will be considered to hold only when explicitly stated.

3.2 Preliminary results

In this section, our goal is to prove some preliminary results necessary in order to prove Theorem 3.1.8. We begin with a technical result that will be useful in our context.

Lemma 3.2.1 *The following inequality holds:* $\|f(q) - f(p) - f'(p)(q-p)\| \leq (L/2)\|q-p\|^2$, *for all* $p, q \in B_r(x_*)$. *Moreover, if* $\|p - x_*\| < r_*$, *then*

$$\|f(x_*) - f(p) - f'(p)(z-p) + f'(x_*)(z - x_*)\| < b, \qquad \forall\ z \in B_{r_*}(x_*).$$

Proof. Because $q+(1-\tau)(p-q) \in B_r(x_*)$ for all $\tau \in [0,1]$ and f is continuously differentiable in Ω, applying the fundamental theorem of calculus and the properties of the norm, we obtain

$$\|f(q) - f(p) - f'(p)(q-p)\| \leq \int_0^1 \|f'(p) - f'(q + (\tau - 1)(q-p))\| \; \|q-p\| \; d\tau.$$

On the other hand, by using (3.5) and performing the integration, the last inequality leads to the first inequality of the lemma. We proceed to prove the second inequality. For this purpose, let $0 < \|p - x_*\| < r_*$ and $0 < \|z - x_*\| < r_*$. By applying the triangle inequality, we have

$$\begin{aligned} &\|f(x_*) - f(p) - f'(p)(z-p) + f'(x_*)(z - x_*)\| \leq \\ &\qquad \|f(x_*) - f(p) - f'(p)(x_* - p)\| + \|f'(p) - f'(x_*)\| \, \|x_* - z\|. \end{aligned} \tag{3.9}$$

Hence, the first inequality of this lemma together with the Lipschitz condition in (3.5) implies that

$$\|f(x_*) - f(p) - f'(p)(x_* - p)\| + \|f'(p) - f'(x_*)\| \, \|x_* - z\| \leq \frac{L}{2}\|x_* - p\|^2 + L\|x_* - p\|\|x_* - z\|.$$

Therefore, by combining this inequality with (3.9), we conclude that

$$\|f(x_*) - f(p) - f'(p)(z-p) + f'(x_*)(z - x_*)\| \leq \frac{L}{2}\|x_* - p\|^2 + L\|x_* - p\|\|x_* - z\|.$$

Taking into account that $\|p - x_*\| < r_*$, $\|z - x_*\| < r_*$ and $r_* \leq \sqrt{2b/3L}$, the desired inequality follows from the last inequality. Thus, the proof of the lemma is complete. ■

To state the next result, for each fixed $x \in \mathbb{R}^n$ we define the following auxiliary set-valued mapping $\Phi_x : \Omega \rightrightarrows \mathbb{R}^n$:

$$\Phi_x(y) := L_{f+F}\left(x_*, f(x_*) - f(x) - f'(x)(y-x) + f'(x_*)(y - x_*)\right)^{-1}, \tag{3.10}$$

where $\mathbb{R}^m \ni u \mapsto L_{f+F}\left(x_*, u\right)^{-1} := \{z \in \mathbb{R}^n : \; u \in L_{f+F}(x_*, z)\}$ is the inverse of L_{f+F} defined in (2.1). Therefore, $y \in \Phi_x(y)$ if and only if x and y satisfy the following inclusion:

$$f(x) + f'(x)(y-x) + F(y) \ni 0,$$

i.e., y is the *next Newton iterate* from x. In the next lemma, we establish the existence of a fixed point of Φ_x for all x in a suitable neighborhood of x_*. Moreover, we present an important bound on the distance between x_* and this fixed point, and establish its uniqueness under strong metric regularity. The statement of this result is as follows.

Lemma 3.2.2 *If $0 < \|x - x_*\| < r_*$, then there exists a fixed point $y \in \Phi_x(y)$ such that*

$$\|y - x_*\| \leq \frac{\kappa L \|x - x_*\|^2}{2(1 - \kappa L \|x - x_*\|)}. \tag{3.11}$$

In particular, this implies that $y \in B_{r_*}(x_*)$. In addition, if $L_{f+F}(x_*, \cdot)$ is strongly metrically regular at x_* for 0, then for all $x \in B_{r_*}(x_*)$ the mapping Φ_x has only one fixed point in $B_{r_*}(x_*)$ satisfying (3.11)*.*

Proof. To prove the first part of the lemma, we will first prove the following two inequalities:

(i) $d\left(x_*, \Phi_x(x_*)\right) \leq \rho\left(1 - \kappa L \|x - x_*\|\right)$;

(ii) $e\left(\Phi_x(p) \cap B_\rho[x_*], \Phi_x(q)\right) \leq \kappa L \|x - x_*\| \, \|p - q\|, \qquad \forall\, p, q \in B_\rho[x_*]$,

where the scalar $\rho > 0$ is defined by

$$\rho := \frac{\kappa L \|x - x_*\|^2}{2(1 - \kappa L \|x - x_*\|)}. \tag{3.12}$$

In order to prove item (i), first note that the definition of the mapping Φ_x given in (3.10) implies that

$$d(x_*, \Phi_x(x_*)) = d\left(x_*, L_{f+F}(x_*, f(x_*) - f(x) - f'(x)(x_* - x))^{-1}\right).$$

Thus, taking into account that the second part of Lemma 3.2.1 with $p = x$ and $z = x_*$ implies that $\|f(x_*) - f(x) - f'(x)(x_* - x)\| < b$, and considering that $x_* \in B_a[x_*]$ and $0 \in L_{f+F}(x_*, x_*)$, we can apply Definition 2.1.1 to conclude that

$$d\left(x_*, \Phi_x(x_*)\right) \leq \kappa \left\|f(x_*) - f(x) - f'(x)(x_* - x)\right\|.$$

Since Lemma 3.2.1 with $p = x$ and $q = x_*$ also implies that

$$\|f(x_*) - f(x) - f'(x)(x_* - x)\| \leq (L/2)\|x_* - x\|^2,$$

combining the two last inequalities, we obtain $d(x_*, \Phi_x(x_*)) \leq (\kappa L/2)\|x - x_*\|^2$, which after some manipulation, yields

$$d(x_*, \Phi_x(x_*)) \leq \frac{\kappa L \|x - x_*\|^2}{2(1 - \kappa L \|x - x_*\|)} \left(1 - \kappa L \|x - x_*\|\right).$$

This inequality, together with definition given in (3.12), proves item (i). To prove item (ii), we take $p, q \in B_\rho[x_*]$. Owing to definition given in (3.12), taking into account that $r_* \leq 2/(3\kappa L)$ and $\|x - x_*\| < r_*$, we can verify that $\rho < r_*$. Thus, Lemma 3.2.1 implies that

$$\|f(x_*) - f(x) - f'(x)(p - x) + f'(x_*)(p - x_*)\| < b,$$

and

$$\|f(x_*) - f(x) - f'(x)(q-x) + f'(x_*)(q-x_*)\| < b.$$

Because $e(\varnothing, \Phi_x(q)) = 0$, we can assume without loss of generality that $\Phi_x(p) \cap B_a[x_*] \neq \varnothing$ for all $p \in B_\rho[x_*]$. Let $z \in \Phi_x(p) \cap B_a[x_*]$. Then, from Definition 2.1.1 with $\bar{x} = x_*$, $\bar{u} = 0$, $x = z$, $u = f(x_*) - f(x) - f'(x)(q-x) + f'(x_*)(q-x_*)$, and $G = L_{f+F}(x_*, \cdot)$, we have

$$d(z, \Phi_x(q)) \leq \kappa d\left(f(x_*) - f(x) - f'(x)(q-x) + f'(x_*)(q-x_*), L_{f+F}(x_*, z)\right).$$

Since $z \in \Phi_x(p)$ implies that $f(x_*) - f(x) - f'(x)(p-x) + f'(x_*)(p-x_*) \in L_{f+F}(x_*, z)$, by using the definition of distance given in (2.2), we obtain

$$d\left(f(x_*) - f(x) - f'(x)(q-x) + f'(x_*)(q-x_*), L_{f+F}(x_*, z)\right) \leq \|[f'(x) - f'(x_*)](p-q)\|.$$

Hence, combining the two last inequalities, we conclude that

$$d(z, \Phi_x(q)) \leq \kappa \|[f'(x) - f'(x_*)](p-q)\|.$$

Taking the supremum with respect to $z \in \Phi_x(p) \cap B_a[x_*]$ in the last inequality and using the definition of excess given in (2.2), we have

$$e\left(\Phi_x(p) \cap B_a[x_*], \Phi_x(q)\right) \leq \kappa \|[f'(x) - f'(x_*)](p-q)\|.$$

Since $\rho < r_* \leq a$, we have $e(\Phi_x(p) \cap B_\rho[x_*], \Phi_x(q)) \leq e(\Phi_x(p) \cap B_a[x_*], \Phi_x(q))$. Hence, from the last inequality and the properties of the norm, we obtain

$$e\left(\Phi_x(p) \cap B_\rho[x_*], \Phi_x(q)\right) \leq \kappa \|f'(x) - f'(x_*)\| \|p-q\|.$$

By using the fact that f' is Lipschitz continuous with constant $L > 0$, the latter inequality becomes

$$e\left(\Phi_x(p) \cap B_\rho[x_*], \Phi_x(q)\right) \leq \kappa L\|x - x_*\|\|p-q\|,$$

and thus item (ii) is proved. Because $r_* \leq 2/(3\kappa L)$ implies that $\kappa L\|x - x_*\| < 1$, we can apply Theorem 2.1.3 with $\Phi = \Phi_x$, $\bar{x} = x_*$, and $\lambda = \kappa L\|x - x_*\|$ to conclude that there exists $y \in B_\rho[x_*]$, i.e., the inequality (3.11) holds, with that $y \in \Phi_x(y)$. To prove that $y \in B_{r_*}(x_*)$, we use the fact that $r_* \leq 2/(3\kappa L)$ and (3.11) to conclude that

$$\|y - x_*\| \leq \frac{\kappa L r_*}{2(1 - \kappa L r_*)}\|x - x_*\| \leq \|x - x_*\| < r_*,$$

which implies the desired result. Therefore, the proof of the first part of the lemma is complete. Now, we assume that $L_{f+F}(x_*, \cdot)$ is strongly metrically regular at x_* for 0. Suppose that there exist $\hat{y}$ and $\tilde{y} \in B_\rho[x_*] \subset B_{r_*}(x_*)$ such that $\hat{y} \in \Phi_x(\hat{y})$ and $\tilde{y} \in \Phi_x(\tilde{y})$. We know that the mapping $z \mapsto L_{f+F}(x_*, z)^{-1} \cap B_a[x_*]$ is single-valued on $B_b[0]$, and thus the definition of Φ_x in (3.10) and the second part of Lemma 3.2.1 imply that $\hat{y} = \Phi_x(\hat{y})$ and $\tilde{y} = \Phi_x(\tilde{y})$. Using the definition of excess in (2.2), item (ii), and the fact that $r_* \leq 2/(3\kappa L)$, we obtain

$$\|\hat{y} - \tilde{y}\| = e(\Phi_x(\hat{y}) \cap B_\rho[x_*], \Phi_x(\tilde{y})) \leq \kappa L\|x - x_*\| \, \|\hat{y} - \tilde{y}\| < \|\hat{y} - \tilde{y}\| \, ,$$

which is a contradiction. Thus, $\hat{y} = \tilde{y}$, and the proof is concluded. ∎

The next lemma plays an important role in the convergence analysis. In particular, it will be used to prove the well-definedness of the sequence $\{x_k\} \subset B_{r_*}(x_*) \cap C$ and its convergence to a solution of problem (1.1).

Lemma 3.2.3 *If $\theta \geq 0$, $x \in C \cap B_{r_*}(x_*) \backslash \{x_*\}$ and $y \in \Phi_x(y)$ satisfies (3.11), then it holds that*

$$\|w - x_*\| \leq \left[\left(1 + \sqrt{2\theta}\right) \frac{\kappa L\|x - x_*\|}{2(1 - \kappa L\|x - x_*\|)} + \sqrt{2\theta}\right] \|x - x_*\|, \qquad \forall\, w \in P_C(y, x, \theta). \quad (3.13)$$

In addition, if $\theta < 1/2$, then $P_C(y, x, \theta) \subset B_{r_*}(x_*) \cap C$.

Proof. Take $w \in P_C(y, x, \theta)$. Then, applying Lemma 3.1.5 with $\tilde{y} = x_*$ and $\tilde{x} = x_*$, we have

$$\|w - P_C\,(x_*, x_*, 0)\| \leq \|y - x_*\| + \sqrt{2\theta}\,(\|x - x_*\| + \|y - x_*\|)\,. \quad (3.14)$$

On the other hand, because $\|x - x_*\| < r_*$, by applying Lemma 3.2.2 and some manipulations, we conclude that

$$\|w - P_C\,(x_*, x_*, 0)\| \leq \left[\left(1 + \sqrt{2\theta}\right) \frac{\kappa L\|x - x_*\|}{2(1 - \kappa L\|x - x_*\|)} + \sqrt{2\theta}\right] \|x - x_*\|.$$

Hence, owing to the fact that $P_C(x_*, x_*, 0) = x_*$, the last inequality and (3.14) yield (3.13). The conditions (3.6) imply that $(1 + \sqrt{2\theta})[(\kappa L\|x - x_*\|)/(2(1 - \kappa L\|x - x_*\|))] + \sqrt{2\theta} < 1$. Thus, it follows from (3.13) that

$$\|w - x_*\| < \|x - x_*\|, \qquad \forall\, w \in P_C(y, x, \theta),$$

and because $\|x - x_*\| < r_*$ we obtain that $P_C(y, x, \theta) \subset B_{r_*}(x_*)$. Because $P_C(y, x, \theta) \subset C$, the last statement of the lemma follows, which concludes the proof. ∎

Now, let us study the uniqueness of the solution for the problem (1.1) in the neighborhood $B_{r_*}(x_*)$.

Lemma 3.2.4 *If the mapping $L_{f+F}(x_*, \cdot)$ is strongly metrically regular at x_* for 0, then x_* is the unique solution of (1.1) in $B_{r_*}(x_*)$.*

Proof. Let $\hat{x}$ be a solution of (1.1) in $B_{r_*}(x_*)$. Thus, $\|\hat{x} - x_*\| < r_* \leq \sqrt{2b/3L}$, which together with the first part of Lemma 3.2.1 implies that

$$\|f(\hat{x}) - f(x_*) - f'(x_*)(\hat{x} - x_*)\| \leq \frac{L}{2}\|\hat{x} - x_*\|^2 < b. \tag{3.15}$$

Moreover, considering that $x_* \in B_{r_*}[x_*]$ and $r_* \leq a$, we can apply Definition 2.1.1 to conclude that

$$d(x_*, L_{f+F}(x_*, -f(\hat{x}) + f(x_*) + f'(x_*)(\hat{x} - x_*))^{-1}) \leq \\ \kappa d(-f(\hat{x}) + f(x_*) + f'(x_*)(\hat{x} - x_*), L_{f+F}(x_*, x_*)).$$

Thus, owing to the fact that $0 \in L_{f+F}(x_*, x_*)$, we can apply the first inequality in (3.15) and the definition of distance given in (2.2) to conclude that

$$d(x_*, L_{f+F}(x_*, -f(\hat{x}) + f(x_*) + f'(x_*)(\hat{x} - x_*))^{-1}) \leq \frac{\kappa L}{2}\|\hat{x} - x_*\|^2.$$

On the other hand, because the mapping $L_{f+F}(x_*, \cdot)$ is strongly metrically regular at x_* for 0, the mapping $z \mapsto L_{f+F}(x_*, z)^{-1} \cap B_a[x_*]$ is single-valued on $B_b[0]$. Furthermore, we know that $0 \in f(\hat{x}) + F(\hat{x}) = f(\hat{x}) - f(x_*) - f'(x_*)(\hat{x} - x_*) + L_{f+F}(x_*, \hat{x})$. Hence, we conclude that $\hat{x} = L_{f+F}(x_*, -f(\hat{x}) + f(x_*) + f'(x_*)(\hat{x} - x_*))^{-1}$, and we obtain from the last inequality that

$$\|\hat{x} - x_*\| \leq \frac{\kappa L}{2}\|\hat{x} - x_*\|^2.$$

If $\|\hat{x} - x_*\| \neq 0$, then last inequality implies that $\|\hat{x} - x_*\| \geq 2/(\kappa L) > 2/(3\kappa L) \geq r_*$, which is absurd, because $\|\hat{x} - x_*\| < r_*$. Therefore, $\|\hat{x} - x_*\| = 0$, and thus x_* is the unique solution of problem (1.1) in $B_{r_*}(x_*)$. ■

Our final task in this section is to prove Theorem 3.1.8. The proof comprises a convenient combination of Lemmas 3.2.2, 3.2.3, and 3.2.4.

3.2.1 Proof of Theorem 3.1.8

Proof. First, we will show by induction on k that there exists a sequence $\{x_k\}$ generated by the Newton-InexP method for solving the problem (1.1), associated to $\{\theta_k\}$ and starting in x_0, which satisfies the following two conditions:

$$\begin{aligned} & x_{k+1} \in B_{r_*}(x_*) \cap C, \\ & \|x_{k+1} - x_*\| \leq \left[\left(1 + \sqrt{2\theta_k}\right) \frac{\kappa L\|x_k - x_*\|}{2(1 - \kappa L\|x_k - x_*\|)} + \sqrt{2\theta_k}\right] \|x_k - x_*\|, \end{aligned} \tag{3.16}$$

for all $k = 0, 1, \ldots$. Take $x_0 \in C \cap B_{r_*}(x_*) \backslash \{x_*\}$ and $k = 0$. Because $\|x_0 - x_*\| < r_*$, applying the first part of Lemma 3.2.2 with $x = x_0$, we obtain that there exists $y_0 \in \Phi_{x_0}(y_0)$ such that

$y_0 \in B_{r_*}(x_*)$. If $y_0 \in C$, then $x_1 = y_0 \in B_{r_*}(x_*) \cap C$, and by using (3.11) we can conclude that (3.16) holds for $k = 0$. Otherwise if $y_0 \notin C$, then take $x_1 \in P_C(y_0, x_0, \theta_0)$. Moreover, by using the first part of Lemma 3.2.3 with $x = x_0$, we obtain that (3.16) holds for $k = 0$. Furthermore, the conditions (3.6) imply that $(1+\sqrt{2\theta_0})[(\kappa L\|x_0 - x_*\|)/(2(1-\kappa L\|x_0 - x_*\|))] + \sqrt{2\theta_0} < 1$, and so the second part of Lemma 3.2.3 give us that $x_1 \in B_{r_*}(x_*) \cap C$. Therefore, there exits x_1 satisfying (3.16) for $k = 0$. Assume for induction that the two assertions in (3.16) hold for $k = 0, 1, \ldots, j-1$. Because $x_j \in B_{r_*}(x_*) \cap C$, we can apply Lemma 3.2.2 with $x = x_j$ to conclude that there exists $y_j \in \Phi_{x_j}(y_j)$ such that $y_j \in B_{r_*}(x_*)$. If $y_j \in C$, then $x_{j+1} = y_j \in B_{r_*}(x_*) \cap C$, and (3.11) implies that (3.16) holds for $k = j$. Otherwise, if $y_j \notin C$ then take $x_{j+1} \in P_C(y_j, x_j, \theta_j)$. Hence, using first part of Lemma 3.2.3 we obtain that the inequality in (3.16) holds for $k = j$. Because the conditions (3.6) implies that $(1+\sqrt{2\theta_j})[(\kappa L\|x_j - x_*\|)/(2(1-\kappa L\|x_j - x_*\|))] + \sqrt{2\theta_j} < 1$, the second part of Lemma 3.2.3 yields that $x_{j+1} \in B_{r_*}(x_*) \cap C$. Thus, there exists x_{j+1} satisfying (3.16) for $k = j$, and the induction step is complete. Therefore, there exists a sequence $\{x_k\}$ generated by the Newton-InexP method solving the problem (1.1), associated to $\{\theta_k\}$ and starting in x_0, and it satisfies the two conditions in (3.16). Now, we proceed to prove that the sequence $\{x_k\}$ converges to x_*. Indeed, because $\|x_k - x_*\| < r_*$ for all $k = 0, 1, \ldots$, $\tilde{\theta} = \sup_k \theta_k < 1/2$ and $r_* \leq [2(1-\sqrt{2\tilde{\theta}})]/[(3-\sqrt{2\tilde{\theta}})\kappa L]$, we conclude from the inequality in (3.16) that

$$\|x_{k+1} - x_*\| < \|x_k - x_*\|.$$

This implies that the sequence $\{\|x_k - x_*\|\}$ converges. Let us say that $t_* = \lim_{k\to+\infty} \|x_k - x_*\| \leq \|x_0 - x_*\| < r_*$. Because $\{x_k\} \subset B_{r_*}(x_*) \cap C$, we can conclude that $t_* < r_*$. On the other hand, by combining the inequality in (3.16) with the second condition in (3.6), we obtain

$$\|x_{k+1} - x_*\| \leq \left[\left(1+\sqrt{2\tilde{\theta}}\right)\frac{\kappa L\|x_k - x_*\|}{2(1-\kappa L\|x_k - x_*\|)} + \sqrt{2\tilde{\theta}}\right]\|x_k - x_*\|,$$

for all $k = 0, 1, \ldots$. Thus, taking the limit in this inequality as k goes to $+\infty$, we have

$$t_* \leq \left[\left(1+\sqrt{2\tilde{\theta}}\right)\frac{\kappa L t_*}{2(1-\kappa L t_*)} + \sqrt{2\tilde{\theta}}\right] t_*.$$

If $t_* \neq 0$, we obtain from the last inequality that $[2(1-\sqrt{2\tilde{\theta}})]/[(3-\sqrt{2\tilde{\theta}})\kappa L] \leq t_*$, which contradicts the first assertion in (3.6), because $t_* < r_*$. Hence, $t_* = 0$, and consequently the sequence $\{x_k\}$ converges to x_*. In particular, if $\lim_{k\to+\infty}\theta_k = 0$, then by taking the limit in (3.7) as k goes to $+\infty$ we obtain that $\limsup_{k\to+\infty}[\|x_{k+1} - x_*\|/\|x_k - x_*\|] = 0$, i.e., the sequence $\{x_k\}$ converges to x_* superlinearly. On the other hand, if $\theta_k = 0$ for all $k = 0, 1, \ldots$, then $\tilde{\theta} = 0$. Hence, from (3.7) and the first equality in (3.6), we have

$$\|x_{k+1} - x_*\| \leq \frac{3\kappa L}{2}\|x_k - x_*\|^2$$

for all $k = 0, 1, \ldots$, and consequently $\{x_k\}$ converges to x_* quadratically. Furthermore, if the mapping $L_{f+F}(x_*, \cdot)$ is strongly metrically regular at x_* for 0, then Lemma 3.2.4 implies that x_* is the unique solution of problem (1.1) in $B_{r_*}(x_*)$. To prove the last statement of the theorem, take $x_0 \in C \cap B_{r_*}(x_*) \backslash \{x_*\}$. Then, the second part of Lemma 3.2.2 implies that there exist a *unique* $y_0 \in B_{\rho_0}(x_*)$ such that $y_0 \in \Phi_{x_0}(y_0)$, i.e., there exists a unique solution y_0 of (3.3) for $k = 0$, where

$$\rho_0 := \frac{\kappa L \|x_0 - x_*\|^2}{2(1 - \kappa L \|x_0 - x_*\|)}.$$

Furthermore, Lemma 3.2.3 implies that every $x_1 \in P_C(y_0, x_0, \theta_0)$ satisfies (3.7) for $k = 0$. Thus, proceeding by induction, we can prove that the every sequence $\{x_k\}$ generated by the Newton-InexP method, associated to $\{\theta_k\}$ and starting in x_0, satisfies (3.7), and by using similar argument as above, we can prove that such a sequence converges to x_*. Therefore, the proof of the theorem is complete. ■

3.3 Application and some examples

In this section, we present an application of Theorem 3.1.8 when F is a maximal monotone operator. To this end, we begin by presenting a class of mappings f and F for which the set-valued mapping defined in (2.1) is strongly metrically regular. The next result is a version of [40, Remark 4] for strongly metrically regular mappings, and its proof will be included here for the sake of completeness. See also [86, Lemma 2.2].

Proposition 3.3.1 *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone mapping and $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. Assume that $x_* \in \mathbb{R}^n$ and $\beta > 0$ satisfy the following condition:*

$$\langle f'(x_*)\, p, p \rangle \geq \beta \|p\|^2, \qquad \forall\ p \in \mathbb{R}^n. \tag{3.17}$$

Then, rge $L_{f+F}(x_*, \cdot) = \mathbb{R}^n$, and for any $\bar{x} \in \mathbb{R}^n$ and $\bar{u} \in L_{f+F}(x_*, \bar{x})$, the set-valued mapping $L_{f+F}(x_*, \cdot) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is strongly metrically regular at $\bar{x} \in \mathbb{R}^n$ for $\bar{u} \in \mathbb{R}^n$, with constants $\kappa = 1/\beta$, $a = +\infty$, and $b = +\infty$.

Proof. First, we will prove that rge $L_{f+F}(x_*, \cdot) = \mathbb{R}^n$. For this, let $0 < \mu < 2\beta/\|f'(x_*)\|^2$, take $\hat{x} \in \mathbb{R}^n$, and define the mapping $\mathbb{R}^n \ni y \mapsto \Phi(y) := (I + \mu F)^{-1}(\mu\hat{x} + y - \mu[f(x_*) + f'(x_*)(y - x_*)])$. Because F is a maximal monotone mapping, according to [31, Theorem 6C.4, p. 387] the mapping $(I + \mu F)^{-1}$ is single-valued and Lipschitz continuous on $\mathbb{R}^n$ with constant 1. Thus, for any $y, z \in \mathbb{R}^n$, we have

$$\begin{aligned}\|\Phi(y) - \Phi(z)\|^2 &\leq \|y - z - \mu f'(x_*)(y - z)\|^2 \\ &= \|y - z\|^2 - 2\mu\langle f'(x_*)(y - z), y - z\rangle + \mu^2\|f'(x_*)(y - z)\|^2.\end{aligned}$$

Using the inequality (3.17) in the last relation, we obtain that

$$\|\Phi(y) - \Phi(z)\|^2 \leq \left(1 - 2\beta\mu + \mu^2\|f'(x_*)\|^2\right) \|y - z\|^2.$$

Considering that $0 < \mu < 2\beta/\|f'(x_*)\|^2$, we have $\lambda^2 := (1 - 2\beta\mu + \mu^2\|f'(x_*)\|^2) < 1$. Thus, we conclude that $\|\Phi(y) - \Phi(z)\| \leq \lambda\|y - z\|$ for all y, $z \in \mathbb{R}^n$. Therefore, by the Banach contraction principle, see [31, Theorem 1A.3, p. 17], there exists $x \in \mathbb{R}^n$ such that $x = \Phi(x)$, which implies that $\hat{x} = L_{f+F}(x_*, x)$, and thus $\operatorname{rge} L_{f+F}(x_*, \cdot) = \mathbb{R}^n$. We proceed to prove that the graph of $L_{f+F}(x_*, \cdot)$ is locally closed at $(\bar{x}, \bar{u})$, i.e., there exists a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{u})$ such that the intersection $\operatorname{gph} L_{f+F}(x_*, \cdot) \cap \mathcal{U}$ is closed. Indeed, let $\{(\bar{x}_k, \bar{u}_k)\} \subset \operatorname{gph} L_{f+F}(x_*, \cdot) \cap \mathcal{U}$ be a sequence such that $\lim_{k\to+\infty} \bar{x}_k = \bar{x}$ and $\lim_{k\to+\infty} \bar{u}_k = \bar{u}$. By the definition of the graph of a set-valued mapping, we have $\bar{u}_k \in L_{f+F}(x_*, \bar{x}_k)$ for all $k = 0, 1, \ldots$. Hence, by using definition given in (2.1) we obtain that $\bar{u}_k \in f(x_*) + f'(x_*)(\bar{x}_k - x_*) + F(\bar{x}_k)$ for all $k = 0, 1, \ldots$. Because F is a maximal monotone mapping, according to [4, Proposition 6.1.3, p. 185] it has closed graph, and thus we can take the limit in the last inclusion to conclude that $\bar{u} \in f(x_*) + f'(x_*)(\bar{x} - x_*) + F(\bar{x})$. This implies that $\bar{u} \in L_{f+F}(x_*, \bar{x})$, and the desired statement follows. Now, we will prove that the mapping $\mathbb{R}^n \ni x \mapsto L_{f+F}(x_*, x)$ is metrically regular at $\bar{x} \in \mathbb{R}^n$ for $\bar{u} \in \mathbb{R}^n$ with constants $\kappa = 1/\beta$, $a = +\infty$, and $b = +\infty$. For this, take arbitrary $x, u \in \mathbb{R}^n$. Because $\operatorname{rge} L_{f+F}(x_*, \cdot) = \mathbb{R}^n$, there exists $y \in \mathbb{R}^n$ such that $u \in L_{f+F}(x_*, y)$. Thus, we can take $w_y \in F(y)$ such that $u = f(x_*) + f'(x_*)(y - x_*) + w_y$. Moreover, for every arbitrary $v \in L_{f+F}(x_*, x)$, we can find $w_x \in F(x)$ such that $v = f(x_*) + f'(x_*)(x - x_*) + w_x$. Thus, the monotonicity of F implies that

$$\begin{aligned}
\langle f'(x_*)(x - y), x - y\rangle &\leq \langle f'(x_*)(x - y), x - y\rangle + \langle w_x - w_y, x - y\rangle \\
&= \langle f(x_*) + f'(x_*)(x - x_*) + w_x - f(x_*) - f'(x_*)(y - x_*) - w_y, x - y\rangle. \\
&= \langle v - u, x - y\rangle \\
&\leq \|v - u\|\|x - y\|.
\end{aligned}$$

On the other hand, (3.17) yields that $\beta\|x - y\|^2 \leq \langle f'(x_*)(x - y), x - y\rangle$, which combined with the last inequality gives

$$\beta\|x - y\|^2 \leq \|v - u\|\|x - y\|.$$

Since $u \in L_{f+F}(x_*, y)$, it follows that if $x = y$ then $x \in L_{f+F}(x_*, u)^{-1}$. In this case, we can conclude that $d\left(x, L_{f+F}(x_*, u)^{-1}\right) = 0 \leq d\left(u, L_{f+F}(x_*, x)\right)/\beta$. Thus, we assume that $x \neq y$, and so

$$\|x - y\| \leq \frac{1}{\beta}\|v - u\|, \qquad \forall\, v \in L_{f+F}(x_*, x).$$

Because $u \in L_{f+F}(x_*, y)$, the definition of distance given in (2.2) and the latter inequality imply that

$$d\left(x, L_{f+F}(x_*, u)^{-1}\right) \leq \frac{1}{\beta} d\left(u, L_{f+F}(x_*, x)\right), \qquad \forall\, x, u \in \mathbb{R}^n.$$

To conclude the proof, it remains to prove that the mapping $z \mapsto L_{f+F}(x_*, z)^{-1}$ is single-valued from $\mathbb{R}^n$ to $\mathbb{R}^n$. Take $z \in \mathbb{R}^n$, $x_1 \in L_{f+F}(x_*, z)^{-1}$, and $x_2 \in L_{f+F}(x_*, z)^{-1}$. For $i = 1, 2$, find $v_i \in F(x_i)$ such that $z = f(x_*) + f'(x_*)(x_i - x_*) + v_i$. Thus, (3.17) and the monotonicity of F imply that

$$
\begin{aligned}
\beta\|x_1 - x_2\|^2 &\leq \langle f'(x_*)(x_1 - x_2), x_1 - x_2\rangle \\
&\leq \langle f'(x_*)(x_1 - x_2), x_1 - x_2\rangle + \langle v_1 - v_2, x_1 - x_2\rangle \\
&= \langle f(x_*) + f'(x_*)(x_1 - x_*) + v_1 - (f(x_*) + f'(x_*)(x_2 - x_*) + v_2), x_1 - x_2\rangle \\
&= 0
\end{aligned}
$$

Yielding that $x_1 = x_2$, so that $L_{f+F}(x_*, \cdot)^{-1}$ is a single-valued mapping. Therefore, the proof is concluded. ■

From now on, $N_D : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *denotes the normal cone mapping of a closed convex set* $D \subset \mathbb{R}^n$, which is defined by

$$
N_D(x) := \begin{cases} \{z : \ \langle z, y - x\rangle \leq 0, \ \forall \ y \in D\} & \text{if } \ x \in D, \\ \varnothing & \text{otherwise.} \end{cases}
$$

In the following result, we present a particular instance of Theorem 3.1.8.

Theorem 3.3.2 *Let C and D be convex sets in $\mathbb{R}^n$ such that C is closed and $C \subset D$, and let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function. Assume that $x_* \in C$, $h(x_*) + N_D(x_*) \ni 0$, and there exist $\beta > 0$ and $L > 0$ such that*

$$
\langle h'(x_*)\, p, p\rangle \geq \beta\|p\|^2, \qquad \|h'(x) - h'(y)\| \leq L\|x - y\|, \qquad \forall\ p, x, y \in \mathbb{R}^n.
$$

Let $\{\theta_k\} \subset [0, 1/2)$ be such that $\tilde{\theta} := \sup_k \theta_k < 1/2$ and $r_* := [2(1 - \sqrt{2\tilde{\theta}})\beta]/[(3 - \sqrt{2\tilde{\theta}})L]$. Then, every sequence $\{x_k\}$ generated by the Newton-InexP method to solve (1.1), associated to $\{\theta_k\}$ and starting in $x_0 \in C \cap B_{r_*}(x_*)\backslash\{x_*\}$, converges to x_*, and the rate of convergence is as follows:

$$
\|x_{k+1} - x_*\| \leq \left[\left(1 + \sqrt{2\theta_k}\right) \frac{L\|x_k - x_*\|}{2(\beta - L\|x_k - x_*\|)} + \sqrt{2\theta_k}\right] \|x_k - x_*\|, \qquad k = 0, 1, \ldots.
$$

As a consequence, if $\lim_{k \to +\infty} \theta_k = 0$, then $\{x_k\}$ converges to x_* superlinearly. In particular, if $\theta_k = 0$ for all $k = 0, 1 \ldots$, then

$$
\|x_{k+1} - x_*\| \leq \frac{3L}{2\beta}\|x_k - x_*\|^2, \qquad k = 0, 1, \ldots,
$$

and the sequence $\{x_k\}$ converges to x_* quadratically.

Proof. Because the normal cone mapping N_D is maximal monotone, see, for example, [4, Corollary 6.3.1, p. 192], we can use Proposition 3.3.1 to obtain that the set-valued mapping $L_{h+N_D}(x_*, \cdot) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is strongly metrically regular at $x_* \in \mathbb{R}^n$ for $0 \in \mathbb{R}^n$, with constants $\kappa = 1/\beta$, $a = +\infty$, and $b = +\infty$. On the other hand, because $x_* \in C$ is such that $h(x_*)+N_D(x_*) \ni 0$, h has a Lipschitz continuous derivative on $\mathbb{R}^n$ and $x_0 \in C \cap B_{r_*}(x_*) \backslash \{x_*\}$. Therefore, we can apply Theorem 3.1.8 to obtain the desired result. ■

We end this section by presenting two examples from the literature that can be seen as particular cases of constrained generalized equations. We begin by presenting the so-called *Constrained Variational Inequality Problem* (CVIP), see, for example, [16].

Example 3.3.3 Let U and Ω be closed convex sets in $\mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous function. The CVIP is defined as:

$$\text{find} \quad x_* \in U \cap \Omega \quad \text{such that} \quad \langle h(x_*), x - x_* \rangle \geq 0, \qquad \forall\, x \in U. \tag{3.18}$$

The problem (3.18) can be rewritten equivalently as:

$$\text{find} \quad x_* \in U \cap \Omega \quad \text{such that} \quad h(x_*) + N_U(x_*) \ni 0.$$

Then, (3.18) can be seen as a special instance of the constrained generalized equation (1.1). Observe that the classical variational inequality problem it is not equivalent to the above CVIP, since in (3.18) the point x_* must belongs to $U \cap \Omega$.

In the next example, we describe the *Split Variational Inequality Problem* (SVIP), which can be rewritten as a special case of the CVIP. See [16, 52] for an extensive discussion on this problem.

Example 3.3.4 Let $U \subset \mathbb{R}^n$ and $\Omega \subset \mathbb{R}^m$ be nonempty, closed convex sets, and consider $A : \mathbb{R}^n \to \mathbb{R}^m$ a linear mapping. Given functions $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^m \to \mathbb{R}^m$, the SVIP is formulated as follows: Find a point $x_* \in U$ such that

$$\langle f(x_*), x - x_* \rangle \geq 0, \qquad \forall\, x \in U,$$

and such that the point $y_* = Ax_* \in \Omega$ satisfies

$$\langle g(y_*), y - y_* \rangle \geq 0, \qquad \forall\, y \in \Omega.$$

By taking $\mathbb{R}^{nm} := \mathbb{R}^n \times \mathbb{R}^m$, $D := U \times \Omega$ and $V := \{w = (x, y) \in \mathbb{R}^{nm} : \ Ax = y\}$ the SVIP is equivalent to the following CVIP:

$$\text{find} \quad w_* \in D \cap V \quad \text{such that} \quad \langle h(w_*), w - w_* \rangle \geq 0, \qquad \forall\, w \in D, \tag{3.19}$$

where $w = (x, y)$ and $h : \mathbb{R}^{nm} \to \mathbb{R}^{nm}$ is defined by $h(x, y) := (f(x), g(y))$, see [16, Lemma 5.1]. Therefore, from Example 3.3.3 and (3.19), the SVIP is equivalent to the following constrained generalized equation:

$$\text{find} \quad w_* \in D \cap V \quad \text{such that} \quad h(w_*) + N_D(w_*) \ni 0,$$

where $D := U \times \Omega$, $V := \{w = (x, y) \in \mathbb{R}^{nm} : \ Ax = y\}$ and $h : \mathbb{R}^{nm} \to \mathbb{R}^{nm}$ is defined by $h(x, y) = (f(x), g(y))$.

It is worth noting that SVIP is quite general and includes several problems as special cases. For instance, *Split Minimization Problem* and *Common Solutions to Variational Inequalities Problem*, see, for example, [1, 16, 17, 71].

Chapter 4

Inexact Newton method with feasible inexact projections for solving constrained equations

In this chapter, we propose an inexact Newton method with feasible inexact projections for solving constrained smooth and nonsmooth equations, i.e, for solving the problem (1.3). Our goal is to show that, under the assumption of smoothness or semismoothness of the function that defines the equation and its regularity at the solution, a sequence generated by the method converges to a solution with linear, superlinear, or quadratic rate. Two applications for the main theorems are provided: one is for semismooth functions and the other is for functions whose derivatives satisfy a radial Hölder condition. To illustrate the practical behavior of the proposed method, some numerical experiments are reported. In particular, we compare the efficiency and robustness of the inexact Newton method with feasible inexact projections (*Inexact Newton-InexP method*) with the inexact Newton method with feasible exact projections (*Inexact Newton-ExP method*) for solving one class of problems.

4.1 Inexact Newton-InexP method and its convergence analysis

In this section, we present the inexact Newton-InexP method for solving the problem (1.3), where the function f is locally Lipschitz continuous. Basically, the inexact Newton-InexP method combines the inexact version of Newton method for solving unconstrained equations (see, for instance, [34, 67]) with a procedure to obtain a feasible inexact projection.

In the following, we formally describe the inexact Newton-InexP algorithm for solving the problem (1.3), with $x_0 \in C$, $\theta > 0$, $\eta > 0$, $\{\theta_k\} \subset [0, \theta)$, and $\{\eta_k\} \subset [0, \eta)$ as the input data.

Algorithm 4.1.1 Inexact Newton-InexP method

Step 0. Let $\theta > 0$, $\eta > 0$, $x_0 \in C$, $\{\theta_k\} \subset [0, \theta)$, and $\{\eta_k\} \subset [0, \eta)$ be given and set $k = 0$.

Step 1. If $f(x_k) = 0$, then **stop**; otherwise, choose an element $V_k \in \partial f(x_k)$ and compute $y_k \in \mathbb{R}^n$ such that

$$\|f(x_k) + V_k(y_k - x_k)\| \leq \eta_k \|f(x_k)\|. \tag{4.1}$$

Step 2. If $y_k \in C$, set $x_{k+1} = y_k$; otherwise, use a procedure to obtain $P_C(y_k, x_k, \theta_k) \in C$ a feasible inexact projection of y_k onto C relative to x_k with relative error tolerance θ_k; and set

$$x_{k+1} \in P_C\left(y_k, x_k, \theta_k\right).$$

Step 3. Set $k \leftarrow k + 1$, and go to **Step 1**.

Below, we describe the main features of the inexact Newton-InexP method.

Remark 4.1.2 In inexact Newton-InexP method, we first check whether the current iterate x_k is a solution of problem (1.3); otherwise, we compute y_k satisfying the relative residual error criteria (4.1). The forcing sequence $\{\eta_k\}$ is used to control the level of accuracy. In particular, as we will show, the specific choice of this sequence is essential to establish the local convergence of the inexact Newton-InexP method. It is worth pointing out that if $\eta_k = 0$ for all $k = 0, 1, \ldots$ (i.e., exact version of the Newton-InexP method), then y_k is obtained by solving for y the system $f(x_k) + V_k(y - x_k) = 0$. Note that to ensure the well-definedness of y_k the Clarke generalized Jacobian must be nonempty, see [18, Proposition 2.6.2, p. 70], and all $V_k \in \partial f(x_k)$ must be nonsingular, for any $k = 0, 1, \ldots$. As the point y_k can be infeasible with respect to the set of constraints C, the inexact Newton-InexP method uses a procedure to obtain a feasible inexact projection, and consequently the new iterate x_{k+1} belongs to C. The choice of the tolerance θ_k is also important in obtaining the local convergence of the inexact Newton-InexP method. Finally, we remark that if f is a continuously differentiable function, $\eta_k = 0$ for all $k = 0, 1, \ldots$, and the procedure to obtain $P_C(y_k, x_k, \theta_k)$ is the CondG procedure, then our method is equivalent to the method proposed in [50]. On the other hand, if f is a nonsmooth function, $C = \mathbb{R}^n$ and $\eta_k = \theta_k = 0$ for all $k = 0, 1, \ldots$, our method reduces to Newton method proposed in [76].

Next, we state and prove our first local convergence result for a sequence generated by the inexact Newton-InexP method. In this case, we assume that $f : \Omega \to \mathbb{R}^n$ is a locally Lipschitz continuous function, but not continuously differentiable.

Theorem 4.1.3 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, $C \subset \Omega$ be a closed convex set, and $f : \Omega \to \mathbb{R}^n$ be a locally Lipschitz continuous function. Suppose that $x_* \in C$ and $f(x_*) = 0$. Let $\Gamma > 0$ and $0 < r \leq r_* := \sup\{t \in \mathbb{R} : \ B_t(x_*) \subset \Omega\}$ such that*

$$\|f(x) - f(x_*)\| \leq \Gamma\|x - x_*\|, \qquad \forall\ x \in B_r(x_*). \tag{4.2}$$

Assume that each $V_{x_*} \in \partial f(x_*)$ is nonsingular and let $\lambda_{x_*} \geq \max\{\|V_{x_*}^{-1}\| : \ V_{x_*} \in \partial f(x_*)\}$. Moreover, there exist $\epsilon > 0$ and $0 < \delta \leq \min\{r, 1\}$ such that for all $x \in B_\delta(x_*)$, $V_x \in \partial f(x)$ is nonsingular and there hold

$$\|V_x^{-1}\| \ \leq \ \frac{\lambda_{x_*}}{1 - \epsilon\lambda_{x_*}}, \tag{4.3}$$

$$\|f(x_*) - f(x) - V_x(x_* - x)\| \ \leq \ \epsilon\|x - x_*\|^{1+\mu}, \qquad 0 \leq \mu \leq 1. \tag{4.4}$$

Let $0 < \theta < 1/2$. Furthermore, assume that $\eta > 0$ and $\epsilon > 0$ satisfy the following conditions

$$\eta < \frac{1 - \sqrt{2\theta}}{\lambda_{x_*}\Gamma\left(1 + \sqrt{2\theta}\right)}, \qquad \epsilon < \frac{1}{2\lambda_{x_*}}\left[\left(1 - \sqrt{2\theta}\right) - \eta\lambda_{x_*}\Gamma\left(1 + \sqrt{2\theta}\right)\right]. \tag{4.5}$$

Then, every sequence $\{x_k\}$ generated by Algorithm 4.1.1 starting in $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, with $0 \leq \eta_k < \eta$ and $0 \leq \theta_k < \theta$, for all $k = 0, 1, \ldots$, belongs to $B_\delta(x_*) \cap C$, satisfies

$$\|x_{k+1} - x_*\| \leq \left[\frac{\lambda_{x_*}[\eta_k\Gamma + \epsilon\|x_k - x_*\|^\mu]}{1 - \epsilon\lambda_{x_*}}\left(1 + \sqrt{2\theta_k}\right) + \sqrt{2\theta_k}\right]\|x_k - x_*\|, \tag{4.6}$$

for all $k = 0, 1, \ldots$, and converges linearly to x_*. As a consequence, if $\lim_{k\to+\infty}\theta_k = 0$ and $\lim_{k\to+\infty}\eta_k = 0$, then $\{x_k\}$ converges superlinearly to x_*. Furthermore, letting $\eta_k < \min\{\eta\|f(x_k)\|^\mu, \eta\}$ and $\theta_k < \min\{\theta\|f(x_k)\|^{2\mu}, \theta\}$, the convergence of $\{x_k\}$ to x_* is of the order of $1 + \mu$.

Proof. We show by induction on k that if $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, then every sequence $\{x_k\}$ generated by Algorithm 4.1.1 belongs to $B_\delta(x_*) \cap C$ and satisfies (4.6). Indeed, set $k = 0$, and take $\theta_0 \geq 0$, $\eta_0 \geq 0$, $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$ and $V_0 := V_{x_0} \in \partial f(x_0)$. Owing to $\|x_0 - x_*\| < \delta$, we obtain that V_{x_0} is nonsingular and then y_0 given in (4.1) is well-defined for $k = 0$. As $f(x_*) = 0$, we have

$$y_0 - x_* = V_{x_0}^{-1}\big(\,[f(x_0) + V_{x_0}(y_0 - x_0)] + [f(x_*) - f(x_0) - V_{x_0}(x_* - x_0)]\,\big).$$

Taking the norm on both sides of the last inequality and using the triangular inequality, we conclude that

$$\|y_0 - x_*\| \leq \left\|V_{x_0}^{-1}\right\|\left[\left\|f(x_0) + V_{x_0}(y_0 - x_0)\right\| + \left\|f(x_*) - f(x_0) - V_{x_0}(x_* - x_0)\right\|\right].$$

Thus, using (4.1) with $k=0$ and the assumptions (4.3) and (4.4) with $x=x_0$, we obtain that

$$\|y_0 - x_*\| \leq \frac{\lambda_{x_*}}{1-\epsilon\lambda_{x_*}} \left[\eta_0\|f(x_0)\| + \epsilon\|x_0 - x_*\|^{1+\mu}\right].$$

As $f(x_*)=0$, from (4.2) we have $\|f(x_0)\| \leq \Gamma\|x_0 - x_*\|$. Hence, the last inequality becomes

$$\|y_0 - x_*\| \leq \frac{\lambda_{x_*}[\eta_0\Gamma + \epsilon\|x_0 - x_*\|^{\mu}]}{1-\epsilon\lambda_{x_*}}\|x_0 - x_*\|. \tag{4.7}$$

Taking any $x_1 \in P_C(y_0, x_0, \theta_0)$ and applying Lemma 3.1.5 with $y = y_0$, $x = x_0$, $\theta = \theta_0$, $\tilde{y} = x_*$ and $\tilde{x} = x_*$, we have

$$\|x_1 - x_*\| \leq \|y_0 - x_*\| + \sqrt{2\theta_0}\,\|y_0 - x_0\| \leq \|y_0 - x_*\|\left(1+\sqrt{2\theta_0}\right) + \sqrt{2\theta_0}\,\|x_0 - x_*\|.$$

Combining (4.7) with the last inequality, we obtain that

$$\|x_1 - x_*\| \leq \frac{\lambda_{x_*}[\eta_0\Gamma + \epsilon\|x_0 - x_*\|^{\mu}]}{1-\epsilon\lambda_{x_*}}\,\|x_0 - x_*\|\left(1+\sqrt{2\theta_0}\right) + \sqrt{2\theta_0}\,\|x_0 - x_*\|,$$

which is equivalent to (4.6) with $k=0$. Since $\delta \leq 1$, $\eta_0 < \eta$, and $\theta_0 < \theta$, by using (4.5), we obtain

$$\frac{\lambda_{x_*}[\eta_0\Gamma + \epsilon\|x_0 - x_*\|^{\mu}]}{1-\epsilon\lambda_{x_*}}\left(1+\sqrt{2\theta_0}\right) + \sqrt{2\theta_0} < \frac{\lambda_{x_*}[\eta\Gamma + \epsilon]}{1-\epsilon\lambda_{x_*}}\left(1+\sqrt{2\theta}\right) + \sqrt{2\theta} < 1.$$

Thus, because $x_0 \in B_\delta(x_*)$, we obtain from (4.6) with $k=0$ that $\|x_1 - x_*\| < \|x_0 - x_*\| < \delta$. As $P_C(y_0, x_0, \theta_0)$ belongs to C and $x_1 \in P_C(y_0, x_0, \theta_0)$, we conclude that x_1 belongs to $B_\delta(x_*) \cap C$, which completes the induction step for $k=0$. The general induction step is completely analogous. Therefore, every sequence $\{x_k\}$ generated by Algorithm 4.1.1 is contained in $B_\delta(x_*) \cap C$ and satisfies (4.6). We proceed to prove that the sequence $\{x_k\}$ converges to x_*. As $\delta \leq 1$, $0 \leq \theta_k < \theta$, and $0 \leq \eta_k < \eta$ for all $k = 0, 1, \ldots$, it follows from (4.6) and (4.5) that

$$\|x_{k+1} - x_*\| < \left[\frac{\lambda_{x_*}[\eta\Gamma + \epsilon]}{1-\epsilon\lambda_{x_*}}\left(1+\sqrt{2\theta}\right) + \sqrt{2\theta}\right]\|x_k - x_*\| < \|x_k - x_*\|, \tag{4.8}$$

for all $k = 0, 1, \ldots$. This implies that the sequence $\{\|x_k - x_*\|\}$ converges. Let us say that $\bar{t} := \lim_{k\to+\infty}\|x_k - x_*\| \leq \delta$. Thus, taking the limit in (4.8) as k goes to $+\infty$, we have

$$\bar{t} \leq \left[\frac{\lambda_{x_*}[\eta\Gamma + \epsilon]}{1-\epsilon\lambda_{x_*}}\left(1+\sqrt{2\theta}\right) + \sqrt{2\theta}\right]\bar{t},$$

If $\bar{t} \neq 0$, then (4.5) implies that $\bar{t} < \bar{t}$, which is absurd. Hence, $\bar{t} = 0$ and $\{x_k\}$ converges linearly to x_*. Now, we assume that $\lim_{k\to+\infty}\theta_k = 0$ and $\lim_{k\to+\infty}\eta_k = 0$. Thus, for $\mu = 0$, it follows from (4.6) that

$$\lim_{k\to+\infty}\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = \frac{\epsilon\lambda_{x_*}}{1-\epsilon\lambda_{x_*}},$$

and, by taking into account that $\epsilon > 0$ is any number satisfying (4.5), we conclude that $\{x_k\}$ converges superlinearly to x_*. For $0 < \mu \leq 1$ it follows straight from (4.6) that $\{x_k\}$ converges superlinearly to x_*. Finally, we assume that $\eta_k < \min\{\eta\|f(x_k)\|^{\mu}, \eta\}$ and $\theta_k < \min\{\theta\|f(x_k)\|^{2\mu}, \theta\}$. Considering that $\{x_k\}$ belongs to $B_\delta(x_*)$, $f(x_*) = 0$, and $\delta \leq r$, it follows from (4.2) that $\|f(x_k)\| \leq \Gamma\|x_k - x_*\|$ for all $k = 0, 1, \ldots$. Hence, $\eta_k < \eta\Gamma^{\mu}\|x_k - x_*\|^{\mu}$ and $\theta_k < \theta\Gamma^{2\mu}\|x_k - x_*\|^{2\mu}$ for all $k = 0, 1, \ldots$. Then, (4.6) implies that

$$\|x_{k+1} - x_*\| < \left[\frac{\lambda_{x_*}[\eta\Gamma^{1+\mu} + \epsilon]}{1 - \epsilon\lambda_{x_*}}\left(1 + \Gamma^{\mu}\sqrt{2\theta}\|x_k - x_*\|^{\mu}\right) + \Gamma^{\mu}\sqrt{2\theta}\right]\|x_k - x_*\|^{1+\mu},$$

for all $k = 0, 1, \ldots$. Therefore, $\{x_k\}$ converges to x_* with order $1 + \mu$, and the proof of the theorem is complete. ■

In the following remark, we present a particular case of Theorem 4.1.3, i.e., when the projection and Newton method are exact.

Remark 4.1.4 Note that the mapping $(0, 1/2) \ni \theta \mapsto (1 - \sqrt{2\theta})/\lambda_{x_*}\Gamma(1 + \sqrt{2\theta})$ is decreasing. Thus, from the first inequality in (4.5), we conclude that if θ approaches the upper bound $1/2$, then η approaches the lower bound 0. Therefore, in Algorithm 4.1.1, the most inexact is the projection, the least inexact has to be the Newton direction. Moreover, it follows from (4.6) that if $\theta_k \equiv 0$ and $\eta_k \equiv 0$ in Theorem 4.1.3, then for $0 < \mu \leq 1$, the convergence rate is $1 + \mu$ as follows

$$\|x_{k+1} - x_*\| \leq \frac{\epsilon\lambda_{x_*}}{1 - \epsilon\lambda_{x_*}}\|x_k - x_*\|^{1+\mu}, \qquad k = 0, 1, \ldots.$$

Hence, ϵ in the second inequality in (4.5) is related to the bound for convergence rate.

Next, we state and prove our second local convergence result for a sequence generated by the inexact Newton-InexP method. In this case, we assume that $f : \Omega \to \mathbb{R}^n$ is a continuously differentiable function.

Theorem 4.1.5 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, $C \subset \Omega$ be a closed convex set, and $f : \Omega \to \mathbb{R}^n$ be a continuously differentiable function. Suppose that $x_* \in C$ and $f(x_*) = 0$. Let $\Gamma > 0$ and $0 < r \leq r_* := \sup\{t \in \mathbb{R} : B_t(x_*) \subset \Omega\}$ such that*

$$\|f(x) - f(x_*)\| \leq \Gamma\|x - x_*\|, \qquad \forall\, x \in B_r(x_*). \tag{4.9}$$

Assume that $f'(x_*)$ is nonsingular and there exist $0 < \mu \leq 1$, $K > 0$, and $0 < \hat{\delta} \leq r$ such that for all $x \in B_{\hat{\delta}}(x_*)$, $f'(x)$ is nonsingular and there hold

$$\|f'(x)^{-1}\| \leq \frac{\|f'(x_*)^{-1}\|}{1 - K\|f'(x_*)^{-1}\|\|x - x_*\|^{\mu}}, \tag{4.10}$$

$$\|f(x_*) - f(x) - f'(x)(x_* - x)\| \leq \frac{\mu K}{1 + \mu}\|x - x_*\|^{1+\mu}. \tag{4.11}$$

Furthermore, let $0 < \theta < 1/2$, $\eta > 0$, and $\delta > 0$ satisfying the following conditions

$$\eta < \frac{1-\sqrt{2\theta}}{\Gamma\|f'(x_*)^{-1}\|\left(1+\sqrt{2\theta}\right)}, \tag{4.12}$$

$$\delta < \min\left\{\hat{\delta}, \left[\frac{(1+\mu)\left[\left(1-\sqrt{2\theta}\right)-\eta\Gamma\|f'(x_*)^{-1}\|\left(1+\sqrt{2\theta}\right)\right]}{K\left[1+2\mu-\sqrt{2\theta}\right]\|f'(x_*)^{-1}\|}\right]^{1/\mu}\right\}. \tag{4.13}$$

Then, every sequence $\{x_k\}$ generated by Algorithm 4.1.1 starting in $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, with $0 \le \eta_k < \eta$ and $0 \le \theta_k < \theta$ for all $k = 0, 1, \ldots$, is contained in $B_\delta(x_*) \cap C$, satisfies

$$\|x_{k+1} - x_*\| \le \left[\frac{\|f'(x_*)^{-1}\|[\eta_k\Gamma(1+\mu)+\mu K\|x_k - x_*\|^\mu]}{(1+\mu)[1-K\|f'(x_*)^{-1}\|\|x_k - x_*\|^\mu]}\left(1+\sqrt{2\theta_k}\right)+\sqrt{2\theta_k}\right]\|x_k - x_*\|, \tag{4.14}$$

for all $k = 0, 1, \ldots$, and converges linearly to x_*. As a consequence, if $\lim_{k\to+\infty}\theta_k = 0$ and $\lim_{k\to+\infty}\eta_k = 0$, then $\{x_k\}$ converges superlinearly to x_*. Furthermore, letting $\eta_k < \min\{\eta\|f(x_k)\|^\mu, \eta\}$ and $\theta_k < \min\{\theta\|f(x_k)\|^{2\mu}, \theta\}$, the convergence of $\{x_k\}$ to x_* is of the order of $1+\mu$.

Proof. First, note that as f is continuously differentiable at x, we have $\partial f(x) = \{f'(x)\}$. We show by induction on k that if $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, then every sequence $\{x_k\}$ generated by Algorithm 4.1.1 is contained in $B_\delta(x_*) \cap C$ and satisfies (4.14). To this end, take $\theta_0 \ge 0$, $\eta_0 \ge 0$, $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, and set $k = 0$. Owing to $\|x_0 - x_*\| < \delta$, we obtain that $f'(x_0)$ is nonsingular. Consequently, (4.1) with $k = 0$ and $V_0 = f'(x_0)$, implies that y_0 is well-defined. Because $f(x_*) = 0$, after some algebraic manipulations, we have

$$y_0 - x_* = f'(x_0)^{-1}\big(\,[f(x_0) + f'(x_0)(y_0 - x_0)] + [f(x_*) - f(x_0) - f'(x_0)(x_* - x_0)]\,\big).$$

Taking the norm on both sides of the last inequality and using the triangular inequality, we conclude that

$$\|y_0 - x_*\| \le \left\|f'(x_0)^{-1}\right\|\left[\left\|f(x_0) + f'(x_0)(y_0 - x_0)\right\| + \left\|f(x_*) - f(x_0) - f'(x_0)(x_* - x_0)\right\|\right].$$

Using (4.1) with $k = 0$ and $V_0 = f'(x_0)$, and the assumptions (4.10) and (4.11) with $x = x_0$, we obtain that

$$\|y_0 - x_*\| \le \frac{\|f'(x_*)^{-1}\|}{1-K\|f'(x_*)^{-1}\|\|x_0 - x_*\|^\mu}\left[\eta_0\|f(x_0)\| + \frac{\mu K}{1+\mu}\|x_0 - x_*\|^{1+\mu}\right]. \tag{4.15}$$

Owing to $f(x_*) = 0$, from (4.9) we conclude that $\|f(x_0)\| \le \Gamma\|x_0 - x_*\|$. Hence, (4.15) becomes

$$\|y_0 - x_*\| \le \frac{\|f'(x_*)^{-1}\|[\eta_0\Gamma(1+\mu)+\mu K\|x_0 - x_*\|^\mu]}{(1+\mu)[1-K\|f'(x_*)^{-1}\|\|x_0 - x_*\|^\mu]}\|x_0 - x_*\|. \tag{4.16}$$

On the other hand, letting $x_1 \in P_C(y_0, x_0, \theta_0)$, Lemma 3.1.5 with $y = y_0$, $x = x_0$, $\theta = \theta_0$, $\tilde{y} = x_*$, and $\tilde{x} = x_*$, implies that

$$\|x_1 - x_*\| \leq \|y_0 - x_*\| + \sqrt{2\theta_0}\,\|y_0 - x_0\| \leq \|y_0 - x_*\| \left(1 + \sqrt{2\theta_0}\right) + \sqrt{2\theta_0}\,\|x_0 - x_*\|.$$

Thus, combining the inequality (4.16) with the last inequality, we conclude that

$$\|x_1 - x_*\| \leq \\ \frac{\|f'(x_*)^{-1}\|[\eta_0\Gamma(1+\mu) + \mu K\|x_0 - x_*\|^{\mu}]}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\|x_0 - x_*\|^{\mu}]}\|x_0 - x_*\| \left(1 + \sqrt{2\theta_0}\right) + \sqrt{2\theta_0}\,\|x_0 - x_*\|,$$

which it is equivalent to (4.14) for $k = 0$. As $\eta_0 < \eta$ and $\theta_0 < \theta$, by using (4.12) and (4.13), we have

$$\frac{\|f'(x_*)^{-1}\|[\eta_0\Gamma(1+\mu) + \mu K\|x_0 - x_*\|^{\mu}]}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\|x_0 - x_*\|^{\mu}]}\left(1 + \sqrt{2\theta_0}\right) + \sqrt{2\theta_0} < \\ \frac{\|f'(x_*)^{-1}\|[\eta\Gamma(1+\mu) + \mu K\delta^{\mu}]}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\delta^{\mu}]}\left(1 + \sqrt{2\theta}\right) + \sqrt{2\theta} < 1.$$

Then, because $x_0 \in B_\delta(x_*)$, we obtain from (4.14) with $k = 0$, that $\|x_1 - x_*\| < \|x_0 - x_*\| < \delta$. As $P_C(y_0, x_0, \theta_0)$ belongs to C and $x_1 \in P_C(y_0, x_0, \theta_0)$, we conclude that x_1 belongs to $B_\delta(x_*) \cap C$, which completes the induction step for $k = 0$. The general induction step is completely analogous. Therefore, every sequence $\{x_k\}$ generated by Algorithm 4.1.1 is contained in $B_\delta(x_*) \cap C$ and satisfies (4.14). Now, we proceed to prove that the sequence $\{x_k\}$ converges to x_*. As $0 \leq \theta_k < \theta$ and $0 \leq \eta_k < \eta$ for all $k = 0, 1, \ldots$, it follows from (4.14) that

$$\|x_{k+1} - x_*\| < \left[\frac{\|f'(x_*)^{-1}\|[\eta\Gamma(1+\mu) + \mu K\|x_k - x_*\|^{\mu}]}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\|x_k - x_*\|^{\mu}]}\left(1 + \sqrt{2\theta}\right) + \sqrt{2\theta}\right]\|x_k - x_*\|,$$

for all $k = 0, 1, \ldots$. On the other hand, using (4.12) and (4.13) in the last inequality we obtain that $\|x_{k+1} - x_*\| < \|x_k - x_*\|$ for all $k = 0, 1, \ldots$. This implies that the sequence $\{\|x_k - x_*\|\}$ converges. Let us say that $\bar{t} := \lim_{k\to+\infty}\|x_k - x_*\| \leq \delta$. Thus, taking the limit in the last inequality as k goes to $+\infty$, we obtain

$$\bar{t} \leq \left[\frac{\|f'(x_*)^{-1}\|[\eta\Gamma(1+\mu) + \mu K\bar{t}^{\mu}]}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\bar{t}^{\mu}]}\left(1 + \sqrt{2\theta}\right) + \sqrt{2\theta}\right]\bar{t},$$

If $\bar{t} \neq 0$, then (4.12) and (4.13) imply that $\bar{t} < \bar{t}$, which is absurd. Hence, $\bar{t} = 0$ and consequently $\{x_k\}$ converge linearly to x_*. Assuming that $\lim_{k\to+\infty}\theta_k = 0$ and $\lim_{k\to+\infty}\eta_k = 0$, it follows from (4.14) that

$$\lim_{k\to+\infty}\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0.$$

Hence, the sequence $\{x_k\}$ converges superlinearly to x_*. Now, we assume that $\eta_k < \min\{\eta\|f(x_k)\|^\mu, \eta\}$ and $\theta_k < \min\{\theta\|f(x_k)\|^{2\mu}, \theta\}$. Considering that $\{x_k\}$ belongs to $B_\delta(x_*)$, $f(x_*) = 0$, and $\delta < r$, it follows from (4.9) that $\|f(x_k)\| \leq \Gamma\|x_k - x_*\|$ for all $k = 0, 1, \ldots$. Thus, $\eta_k < \eta\Gamma^\mu\|x_k - x_*\|^\mu$ and $\theta_k < \theta\Gamma^{2\mu}\|x_k - x_*\|^{2\mu}$ for all $k = 0, 1, \ldots$. Then, (4.14) implies that

$$\|x_{k+1} - x_*\| <$$
$$\left[\frac{\|f'(x_*)^{-1}\|[\eta\Gamma^{1+\mu}(1+\mu) + \mu K]}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\|x_k - x_*\|^\mu]}\left(1 + \Gamma^\mu\sqrt{2\theta}\|x_k - x_*\|^\mu\right) + \Gamma^\mu\sqrt{2\theta}\right]\|x_k - x_*\|^{1+\mu},$$

for all $k = 0, 1, \ldots$. Therefore, $\{x_k\}$ converges to x_* with order $1 + \mu$, which complete the proof of the theorem. ■

In the following remark, we present a particular case of Theorem 4.1.5, where the projection and Newton method are exact.

Remark 4.1.6 In Theorem 4.1.5, if we take $\theta_k \equiv 0$ and $\eta_k \equiv 0$, then for $0 < \mu \leq 1$, the convergence rate is $1 + \mu$ as follows

$$\|x_{k+1} - x_*\| \leq \frac{\mu K\|f'(x_*)^{-1}\|}{(1+\mu)[1 - K\|f'(x_*)^{-1}\|\|x_k - x_*\|^\mu]}\|x_k - x_*\|^{1+\mu}, \qquad k = 0, 1, \ldots.$$

4.2 Special cases

In this section, we present two special cases: one of Theorem 4.1.3 and one of Theorem 4.1.5. We begin by presenting the special case of Theorem 4.1.3.

4.2.1 Under semismooth condition

In this section, we present a local convergence theorem for the inexact Newton-InexP method for solving constrained semismooth equations. The semismoothness plays an important role, since the Newton method is still applicable and converges locally with superlinear rate to a regular solution. Let us first to present the concept of regularity.

Definition 4.2.1 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set. A function $f : \Omega \to \mathbb{R}^n$ locally Lipschitz continuous is said to be regular at $x_* \in \Omega$ if every $V_{x_*} \in \partial f(x_*)$ is nonsingular. If f is regular at all points of Ω, the function f is said to be regular on Ω.*

In the following, our first task is to prove that locally Lipschitz continuous functions satisfy the inequality (4.3) near a regular point for every $0 < \epsilon < 1/\lambda_{x_*}$. First, we remind that $\partial f(x)$

is a nonempty and compact set for all $x \in \Omega$, see [18, Proposition 2.6.2, p. 70]. The statement of the result is as follows.

Lemma 4.2.2 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}^n$ be a locally Lipschitz continuous function. If f is regular at $x_* \in \Omega$, then for every $0 < \epsilon < 1/\lambda_{x_*}$, where $\lambda_{x_*} \geq \max\{\|V_{x_*}^{-1}\| : V_{x_*} \in \partial f(x_*)\}$, there exists $\delta > 0$ such that f is regular on $B_\delta(x_*)$ and there holds*

$$\|V_x^{-1}\| \leq \frac{\lambda_{x_*}}{1 - \epsilon\lambda_{x_*}}, \qquad \forall\, x \in B_\delta(x_*), \quad \forall\, V_x \in \partial f(x). \tag{4.17}$$

Proof. As f is regular at $x_* \in \Omega$ and $\partial f(x_*)$ is a nonempty and compact set, $\lambda_{x_*} > 0$ is well-defined. On the other hand, it follows from [18, Proposition 2.6.2, p. 70] that ∂f is upper semicontinuous at x_*. Thus, for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$\partial f(x) \subset \left\{V_x \in \mathbb{R}^{n\times n} : \|V_x - V_{x_*}\| < \epsilon \text{ for some } V_{x_*} \in \partial f(x_*)\right\}, \qquad \forall\, x \in B_\delta(x_*).$$

Hence, for each $V_x \in \partial f(x)$ and $0 < \epsilon < 1/\lambda_{x_*}$, there exists $V_{x_*} \in \partial f(x_*)$ that is nonsingular such that $\|V_{x_*}^{-1}\|\|V_x - V_{x_*}\| < \epsilon\lambda_{x_*} < 1$. Thus, applying the Banach lemma, see [31, Lemma 5A.4, p. 282], we conclude that V_x is nonsingular and

$$\|V_x^{-1}\| \leq \frac{\|V_{x_*}^{-1}\|}{1 - \|V_{x_*}^{-1}\|\|V_x - V_{x_*}\|}.$$

Therefore, considering that $\|V_{x_*}^{-1}\| \leq \lambda_{x_*}$ the inequality (4.17) follows, and the proof of the lemma is complete. ■

In the following, we present a class of functions satisfying the inequality (4.4), namely the semismooth functions. There are several equivalent definitions for semismooth functions, here we use that given in [31, p. 411]. For an extensive study on semismooth functions, see, for example, [34].

Definition 4.2.3 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set. A function $f : \Omega \to \mathbb{R}^n$ that is locally Lipschitz continuous on Ω and directionally differentiable in every direction is said to be semismooth at $x_* \in \Omega$ when for every $\epsilon > 0$ there exists $\delta > 0$ such that*

$$\|f(x_*) - f(x) - V_x(x_* - x)\| \leq \epsilon\|x - x_*\|, \qquad \forall\, x \in B_\delta(x_*), \quad \forall\, V_x \in \partial f(x),$$

and is said to be μ-order semismooth at $x_* \in \Omega$, for $0 < \mu \leq 1$ when there exist $\epsilon > 0$ and $\delta > 0$ such that

$$\|f(x_*) - f(x) - V_x(x_* - x)\| \leq \epsilon\|x - x_*\|^{1+\mu}, \qquad \forall\, x \in B_\delta(x_*), \quad \forall\, V_x \in \partial f(x).$$

Next, we state and prove the local convergence result of the inexact Newton-InexP method for solving constrained semismooth equations, which is a consequence of Lemma 4.2.2 and Theorem 4.1.3.

Theorem 4.2.4 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, $C \subset \Omega$ be a closed convex set, and $f : \Omega \to \mathbb{R}^n$ be semismooth and regular at $x_* \in \Omega$. Let $\Gamma > 0$ and $0 < r \le r_* := \sup\{t \in \mathbb{R} : \ B_t(x_*) \subset \Omega\}$ such that*

$$\|f(x) - f(x_*)\| \le \Gamma \|x - x_*\|, \qquad \forall\ x \in B_r(x_*).$$

Take $\theta > 0$ and $\eta > 0$ such that

$$\theta < \frac{1}{2}, \qquad\qquad \eta < \frac{1-\sqrt{2\theta}}{\lambda_{x_*}\Gamma\left(1+\sqrt{2\theta}\right)}.$$

Assume that $x_* \in C$ and $f(x_*) = 0$. Then, there exists $0 < \delta \le r$ such that every sequence $\{x_k\}$ generated by Algorithm 4.1.1 starting in $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, with $0 \le \theta_k < \theta$ and $0 \le \eta_k < \eta$ for all $k = 0, 1, \ldots$, belongs to $B_\delta(x_*) \cap C$ and converges linearly to x_*. If $\lim_{k\to+\infty} \theta_k = 0$ and $\lim_{k\to+\infty} \eta_k = 0$, then $\{x_k\}$ converges superlinearly to x_*. In addition, if f is μ-order semismooth at x_*, $\eta_k < \min\{\eta\|f(x_k)\|^\mu, \eta\}$, and $\theta_k < \min\{\theta\|f(x_k)\|^{2\mu}, \theta\}$, then the convergence of $\{x_k\}$ to x_* is of the order of $1 + \mu$.

Proof. Because the function f is semismooth and regular at $x_* \in \Omega$, we can take $\lambda_{x_*} \ge \max\{\|V_{x_*}^{-1}\| : \ V_{x_*} \in \partial f(x_*)\}$. Take $0 < \epsilon < 1/\lambda_{x_*}$. Then, from Lemma 4.2.2 and Definition 4.2.3, there exists $0 < \delta \le \min\{r, 1\}$ satisfying the inequalities (4.3) and (4.4) for $\mu = 0$. In addition, if f is μ-order semismooth, we conclude also from Lemma 4.2.2 and Definition 4.2.3 that there exists $0 < \delta \le \min\{r, 1\}$ satisfying the inequalities (4.3) and (4.4) for $0 < \mu \le 1$. Therefore, f satisfies all conditions of Theorem 4.1.3 and by reducing $\epsilon > 0$ so that it satisfies the second inequality in (4.5) the desired result follows. ■

In the following remark, we show that with some adjustments Theorem 4.2.4 reduces to some well-known results.

Remark 4.2.5 It is worth mentioning that if $C = \mathbb{R}^n$ and $\theta_k = 0$ for all $k = 0, 1, \ldots$, then with some adjustments Theorem 4.2.4 reduces to [67, Theorem 3]; see also [34, Theorem 7.5.5, p. 694]. If $C = \mathbb{R}^n$, $\eta_k = \theta_k = 0$ for all $k = 0, 1, \ldots$, then Theorem 4.2.4 reduces to [76, Theorem 3.2], see also [34, Theorem 7.5.3, p. 693]. Finally, if $C = \mathbb{R}^n$, f is a continuously differentiable function, $f'(x_*)$ is nonsingular, and $\theta_k = \eta_k = 0$ for all $k = 0, 1, \ldots$, then the theorem above reduces to the first part of [12, Proposition 1.4.1, p. 90].

4.2.2 Under radial Hölder condition on the derivative

In this section, we present a local convergence theorem for the inexact Newton-InexP method under the radial Hölder condition on the derivative. We begin by presenting the definition of the radial Hölder condition.

Definition 4.2.6 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}^n$ be a continuously differentiable function. The derivative f' satisfies the radial Hölder condition at $x_* \in \Omega$ if there exist $K > 0$ and $0 < \mu \leq 1$ such that*

$$\|f'(x) - f'(x_* + \tau(x - x_*))\| \leq K(1 - \tau^\mu)\|x - x_*\|^\mu,$$

for all $x \in \Omega$ and $\tau \in [0, 1]$ such that $x_* + \tau(x - x_*) \in \Omega$.

Our first task is to prove that continuously differentiable functions with radially Hölder derivative satisfies the inequality (4.10) around regular points.

Lemma 4.2.7 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $f : \Omega \to \mathbb{R}^n$ be a continuously differentiable function. Assume that f' is nonsingular and radially Hölder at $x_* \in \Omega$, with constants $K > 0$ and $0 < \mu \leq 1$. Take*

$$0 < \hat{r} < \frac{1}{(K\|f'(x_*)^{-1}\|)^{1/\mu}}. \tag{4.18}$$

Then, $f'(x)$ is nonsingular for all $x \in B_{\hat{r}}(x_*)$, and there holds

$$\left\|f'(x)^{-1}\right\| \leq \frac{\|f'(x_*)^{-1}\|}{1 - K\|f'(x_*)^{-1}\|\|x - x_*\|^\mu}, \qquad \forall\, x \in B_{\hat{r}}(x_*).$$

Proof. Using that f' is nonsingular and radially Hölder at $x_* \in \Omega$, with constants $K > 0$ and $0 < \mu \leq 1$, and taking into account (4.18), we have

$$\|f'(x_*)^{-1}\|\|f'(x) - f'(x_*)\| \leq K\|f'(x_*)^{-1}\|\|x - x_*\|^\mu < K\|f'(x_*)^{-1}\|\hat{r}^\mu < 1,$$

for all $x \in B_{\hat{r}}(x_*)$. Therefore, the desired result follows by applying the Banach lemma, see [31, Lemma 5A.4, p. 282]. ■

The next lemma establishes that continuously differentiable functions with radially Hölder derivative satisfy (4.11); its proof follows the same idea as [12, Proposition 1.4.1, p. 90] and will be included here for the sake of completeness.

Lemma 4.2.8 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, $x_* \in \Omega$, $r_* := \sup\{t \in \mathbb{R} : \ B_t(x_*) \subset \Omega\}$ and $f : \Omega \to \mathbb{R}^n$ be a continuously differentiable function. Assume that f' is radially Hölder at x_*, with constants $K > 0$ and $0 < \mu \leq 1$. Then it holds that*

$$\|f(x_*) - f(x) - f'(x)(x_* - x)\| \leq \frac{\mu K}{1 + \mu}\|x - x_*\|^{1+\mu}, \qquad \forall\, x \in B_{r_*}(x_*).$$

Proof. Note that $\|f(x_*) - f(x) - f'(x)(x_* - x)\| = \|f(x) - f(x_*) - f'(x)(x - x_*)\|$. Because $x_* + \tau(x - x_*) \in B_{r_*}(x_*)$, for all $\tau \in [0, 1]$, the fundamental theorem of calculus implies that

$$\|f(x) - f(x_*) - f'(x)(x - x_*)\| \leq \int_0^1 \|f'(x_* + \tau(x - x_*)) - f'(x)\|\|x - x_*\|d\tau.$$

Owing to f' be radially Hölder at $x_* \in \Omega$ with constants $K > 0$ and $0 < \mu \leq 1$, we have

$$\|f(x) - f(x_*) - f'(x)(x - x_*)\| \leq \int_0^1 K(1 - \tau^{\mu})\|x - x_*\|^{1+\mu} d\tau.$$

Therefore, performing the integration the desired result follows. ■

Now, we are ready to present a local convergence theorem on the inexact Newton-InexP method for a continuously differentiable function f such that the derivative f' is radially Hölder. The statement of the result is as follows.

Theorem 4.2.9 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, $C \subset \Omega$ be a closed convex set, and $f : \Omega \to \mathbb{R}^n$ be a continuously differentiable function such that f' is nonsingular and radially Hölder at $x_* \in \Omega$, with constants $K > 0$ and $0 < \mu \leq 1$. Let $\Gamma > 0$ and $0 < r \leq r_* := \sup\{t \in \mathbb{R} : \ B_t(x_*) \subset \Omega\}$ such that*

$$\|f(x) - f(x_*)\| \leq \Gamma\|x - x_*\|, \qquad \forall\ x \in B_r(x_*).$$

Let $\theta > 0$ and $\eta > 0$ such that

$$\theta < \frac{1}{2}, \qquad\qquad \eta < \frac{1 - \sqrt{2\theta}}{\Gamma\|f'(x_*)\|\left(1 + \sqrt{2\theta}\right)}.$$

Assume that $x_* \in C$ and $f(x_*) = 0$. Then, there exists $0 < \delta \leq r$ such that every sequence $\{x_k\}$ generated by Algorithm 4.1.1 starting in $x_0 \in C \cap B_\delta(x_*)\backslash\{x_*\}$, with $0 \leq \theta_k < \theta$ and $0 \leq \eta_k < \eta$ for all $k = 0, 1, \ldots$, belongs to $B_\delta(x_*) \cap C$ and converges linearly to x_*. As a consequence, if $\lim_{k\to+\infty} \theta_k = 0$ and $\lim_{k\to+\infty} \eta_k = 0$, then $\{x_k\}$ converges superlinearly to x_*. In addition, if $\eta_k < \min\{\eta\|f(x_k)\|^{\mu}, \eta\}$ and $\theta_k < \min\{\theta\|f(x_k)\|^{2\mu}, \theta\}$, then the convergence of $\{x_k\}$ to x_* is of the order of $1 + \mu$.

Proof. Let $0 < \hat{r} < 1/[(K\|f'(x_*)^{-1}\|)^{1/\mu}]$ and $0 < \hat{\delta} \leq \min\{\hat{r}, r\}$. Then, from Lemmas 4.2.7 and 4.2.8, we conclude that f satisfies the conditions (4.10) and (4.11) in $B_{\hat{\delta}}(x_*)$. Therefore, f satisfies all conditions of Theorem 4.1.5 and by taking $\delta > 0$ satisfying (4.13) the desired result follows. ■

In the following remark, we show that with some adjustments, Theorem 4.2.9 has as particular instances some well-known results.

Remark 4.2.10 It is worth mentioning that if $C = \mathbb{R}^n$ and $\eta_k = \theta_k = 0$ for all $k = 0, 1, \ldots$, then Theorem 4.2.9 reduces to the second part of [12, Proposition 1.4.1, p. 90]. If the procedure to obtain the feasible inexact projection is the CondG procedure and $\eta_k = 0$ for all $k = 0, 1, \ldots$, then Theorem 4.2.9 reduces to [50, Theorem 7]. Finally, if the procedure to obtain the feasible inexact projection is the CondG procedure, then with some adjustments Theorem 4.2.9 reduces to [49, Corollary 2].

4.3 Numerical experiments

In this section, we report some numerical experiments that show the computational feasibility of the inexact Newton-ExP method and inexact Newton-InexP method on one class of problems, which we call the CAVEs. It is worth mentioning that works dealing with the Newton method to solve absolute value equation (AVE) include [11, 65]. The CAVE is described as

$$\text{find} \quad x \in C \quad \text{such that} \quad Ax - |x| = b,$$

where $C := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i \leq d,\, x_i \geq -1,\, i = 1, 2, \ldots, n\}$, $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n \equiv \mathbb{R}^{n\times 1}$, $d \in \mathbb{R}$, and $|x|$ denotes the vectors whose i-th component is equal to $|x_i|$. In our implementation, the CAVEs have been generated randomly. We used the Matlab routine *sprand* to construct matrix A. In particular, this routine generates a sparse matrix with predefined dimension, density, and singular values. Initially, we defined the dimension n and randomly generated the vector of singular values from a uniform distribution on $(0, 1)$. To ensure that $\|A^{-1}\| < 1/3$, i.e., so that the assumptions of [11, Theorem 2] are fulfilled, we rescale the vector of singular values by multiplying it by 3 divided by the minimum singular value multiplied by a random number in the interval $(0, 1)$. To generate the vector b and the constant d, we chose a random solution x_* from a uniform distribution on $(0.1, 300)$ and computed $b = Ax_* - |x_*|$ and $d = \sum_{i=1}^n (x_*)_i$, where $(x_*)_i$ denotes the i-th component of the vector x_*. In both methods, $x_0 = (d/2n, d/2n, \ldots, d/2n)$ was defined as the starting point, the initialization data θ was taken equal to 10^{-1} and 10^{-8} for the methods with inexact and exact projection, respectively, and η was taken equal to $0.9999[(1 - \sqrt{2\theta})/0.5\Gamma(1 + \sqrt{2\theta})]$ with $\Gamma = \|A\| + 1$. We stopped the execution of Algorithm 4.1.1 at x_k, declaring convergence if

$$\|Ax_k - |x_k| - b\| < 10^{-6}.$$

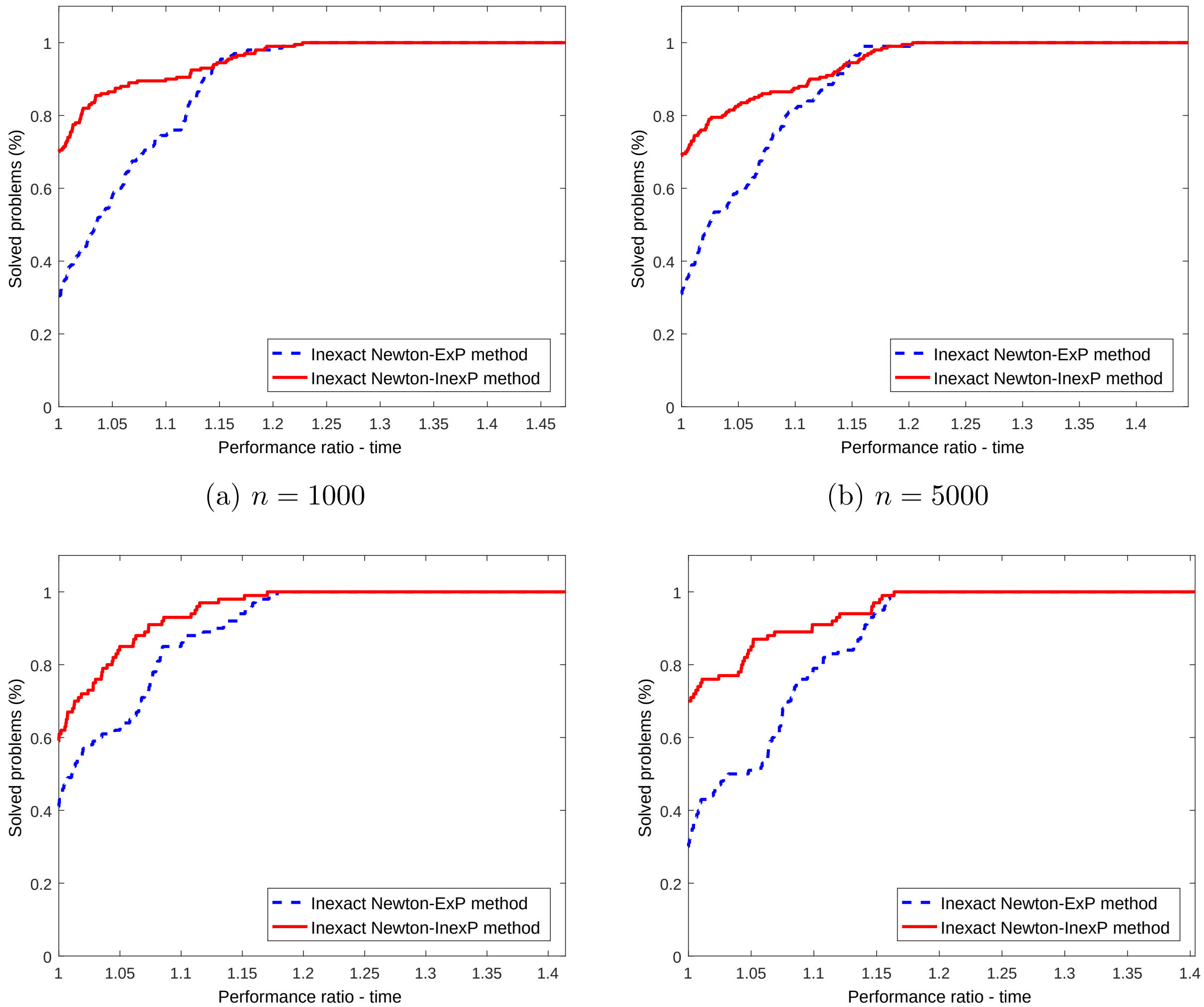
In case this stopping criterion was not respected, the method stopped when a maximum of 50 iterations had been performed. The procedure to obtain feasible projections used in our implementation was the CondG Procedure; see, for example, [61]. In particular, this procedure stopped when either the stopping criterion, i.e., the condition $\langle y_k - x_{k+1}, z - x_{k+1}\rangle \leq \theta_k \|y_k - x_k\|^2$ was satisfied for all $z \in C$ and $k = 0, 1, \ldots$ or a maximum of 100 iterations was performed. For this class of problems, an element of the Clarke generalized Jacobian (see [11, 65]) is given by

$$V = A - \operatorname{diag}(\operatorname{sgn}(x)), \qquad x \in \mathbb{R}^n,$$

where $\operatorname{diag}(\alpha_i)$ denotes a diagonal matrix with diagonal elements $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\operatorname{sgn}(x)$ denotes a vector with components equal to -1, 0 or 1 depending on whether the corresponding component of the vector x is negative, zero or positive.

The inexact Newton-ExP and inexact Newton-InexP methods requires the linear system $f(x_k) + V_k(y_k - x_k) = 0$ to be solved approximately, in the sense of (4.1). Matlab has several iterative methods for solving linear equations. For our class of problems, the routine *lsqr* was the most efficient; thus, in all tests, we used *lsqr* as an iterative method to solve linear equations approximately. In particular, this routine is an algorithm for sparse linear equations and sparse least squares; for further details, see, for example, [74]. We compare the efficiency and robustness of the methods using the performance profiles graphics, see [27]. The efficiency is related to the percentage of problems for which the method was the fastest, whereas robustness is related to the percentage of problems for which the method found a solution. In a performance profile, efficiency and robustness can be accessed on the extreme left (at 1 in domain) and right of the graphic, respectively. The numerical results were obtained using Matlab version R2016a on a 2.5GHz Intel® Core™ i5 2450M computer with 6GB of RAM and Windows 7 ultimate system and are freely available from https://orizon.mat.ufg.br/admin/pages/11432-codes.

Figure 4.1 reports a comparison, using performance profiles, between the inexact Newton-ExP and inexact Newton-InexP methods for solving CAVEs of dimensions 1000, 5000, 8000, and 10000. We generated 200 CAVEs with dimensions 1000 and 5000, and 100 CAVEs with dimensions 8000 and 10000. The density of the matrix A was taken equal to 0.003, as well as in [11]. This means that only about 0.3% of the elements of A are nonnull. To obtain the CPU time more accurately, we run each test problem 10 times and we define the corresponding CPU time as the median of these measurements. Analyzing Figure 4.1, we see that the inexact Newton-InexP method is more efficient than the inexact Newton-ExP method on the set of test problems. In particular, the efficiencies of the inexact Newton method with the exact and inexact projections are, respectively, 30.5% and 69.5% for problems of dimension 1000, 31.0% and 69.0% for problems of dimension 5000, 41.0% and 59.0% for problems of dimension 8000, and 30.0% and 70.0% for problems of dimension 10000. Thus, we can conclude that for this class of test problem the parameter θ and consequently η given in (4.5) limit the effectiveness of the method.



(a) $n = 1000$ (b) $n = 5000$

(c) $n = 8000$ (d) $n = 10000$

Figure 4.1: Performance profile comparing the inexact Newton-ExP method versus inexact Newton-InexP method for CAVEs using CPU time as performance measurement.

Table 4.1 lists, for each method, the percentage of problems solved "%", the average numbers of iterations "Iter", and the average times in seconds "Time". As can be seen, the robustness is 100.0% for both methods. The average numbers of iterations is approximately 7 and 6 for the exact and inexact versions, respectively. Moreover, with respect to the average time, it is possible to observe a certain trend, that is, as the dimension of the problem increases, the performance of the inexact Newton-InexP method becomes better compared with the inexact Newton-ExP method.

	Inexact Newton-ExP method			Inexact Newton-InexP method		
Dimension	**%**	**Iter**	**Time**	**%**	**Iter**	**Time**
1000	100.0	6.61	0.58	100.0	5.50	0.56
5000	100.0	6.70	11.67	100.0	5.67	11.48
8000	100.0	6.90	30.76	100.0	5.81	30.42
10000	100.0	6.88	46.66	100.0	5.77	45.27

Table 4.1: Performance of the inexact Newton-ExP method versus the inexact Newton-InexP method

The results discussed above allow us to conclude that the use of the inexact projection can make the inexact Newton method more efficient for solving some constrained problems. Thus, we can say that the inexact Newton-InexP method may be a robust and efficient tool for solving other classes of nonsmooth functions subject to a set of constraints.

Chapter 5

Nonsmooth Newton method for finding a singularity of a special class of vector fields on Riemannian manifolds

In this chapter, we extend some results of nonsmooth analysis from the Euclidean context to the Riemannian setting. In particular, we discuss the concept and some properties of a locally Lipschitz continuous vector field defined on a Riemannian manifold, such as Clarke generalized covariant derivative, upper semicontinuity of this covariant derivative and Rademacher theorem. We also present a version of the nonsmooth Newton method for finding a singularity of a special class of locally Lipschitz continuous vector fields. Under mild conditions, we establish the well-definedness and local convergence of a sequence generated by this method in a neighborhood of a singularity. In particular, a local convergence theorem for semismooth vector fields is presented. Under Kantorovich-type assumptions the convergence of the sequence generated by the nonsmooth Newton method to a solution is established, and its uniqueness in a suitable neighborhood of the starting point is verified. Furthermore, a class of examples of locally Lipschitz continuous vector field satisfying the assumptions of the convergence theorems is presented.

5.1 Nonsmooth analysis in Riemannian manifolds

The goal of this section is to extend some basic results of nonsmooth analysis from linear context to Riemannian setting. In particular, we study the basic properties of the locally

Lipschitz continuous vector fields in Riemannian setting, a generalization of Rademacher theorem and introduce the concept of Clarke generalized covariant derivative to this new context. A comprehensive study of nonsmooth analysis in a linear context can be found in [18]. We begin with the definition of a locally Lipschitz continuous vector field. This concept was introduced in [19] for gradient vector fields and its extension to general vector fields can be found in [15, p. 241].

Definition 5.1.1 *A vector field X on M is said to be Lipschitz continuous on $\Omega \subset M$, if there exists a constant $L > 0$ such that for $p, q \in M$ and all γ geodesic segment joining p to q, there holds*

$$\|P_{\gamma,p,q}X(p) - X(q)\| \leq L\,\ell(\gamma), \qquad \forall\ p, q \in \Omega.$$

Given $p \in M$, if there exists $\delta > 0$ such that X is Lipschitz continuous on $B_\delta(p)$, then X is said to be Lipschitz continuous at p. Moreover, if for all $p \in M$, X is Lipschitz continuous at p, then X is said to be locally Lipschitz continuous on M.

Let d_{TM} be the Riemannian distance on tangent bundle TM. Let us define the concept of *Lipschitz continuity* of a vector field defined on a Riemannian manifold as a mapping between the metric spaces (M, d) and (TM, d_{TM}). The formal definition is as follows.

Definition 5.1.2 *A vector field X on M is said to be metrically Lipschitz continuous on $\Omega \subset M$, if there exists a constant $L > 0$ such that*

$$d_{TM}(X(p), X(q)) \leq L\,d(p, q), \qquad \forall\ p, q \in \Omega.$$

Given $p \in M$, if there exists $\delta > 0$ such that X is metrically Lipschitz continuous on $B_\delta(p)$, then X is said to be metrically Lipschitz continuous at p. Moreover, if for all $p \in M$, X is metrically Lipschitz continuous at p, then X is said to be locally metric Lipschitz continuous on M.

It is an immediate consequence from the last definition that all metrically Lipschitz continuous vector fields are continuous. In the following result, we present a relationship between the Definitions 5.1.1 and 5.1.2.

Theorem 5.1.3 *If X is Lipschitz continuous with constant $L > 0$, then X is also metrically Lipschitz continuous with constant $\sqrt{1 + L^2}$. As a consequence, if X is locally Lipschitz continuous on M, then X is also locally metric Lipschitz continuous on M.*

Proof. Because M is a complete manifold, $\pi X(p) = p$ and $\pi X(q) = q$, it follows from definition (2.3) that

$$d_{TM}(X(p), X(q)) \leq \sqrt{d^2(p, q) + \|P_{\gamma,p,q}X(p) - X(q)\|^2}, \qquad \forall\ p, q \in M, \tag{5.1}$$

where γ is the minimal geodesic segment joining p to q. Considering that X is a Lipschitz continuous vector field with constant $L > 0$ from Definition 5.1.1, we have $\|P_{\gamma,p,q}X(p) - X(q)\| \leq L\, d(p,q)$ for all $p, q \in M$. Hence, inequality (5.1) becomes

$$d_{TM}(X(p), X(q)) \leq \sqrt{1+L^2}\; d(p,q),$$

for all $p, q \in M$. Consequently, by using Definition 5.1.2, we conclude that X is metrically Lipschitz continuous with constant $\sqrt{1+L^2}$. Therefore, the proof of the first part is concluded. The proof of the second part is similar. ■

In the next definition, we present the notion of sets of measure zero to manifolds, which has appeared in [63, 78].

Definition 5.1.4 *A subset $E \subseteq M$ has measure zero in M if for every smooth chart (U, φ) for M, the subset $\varphi(E \cap U) \subseteq \mathbb{R}^n$ has n-dimensional measure zero.*

Let X be a locally Lipschitz continuous vector field on M. Throughout this chapter, $\mathcal{D}_X$ is the set defined by

$$\mathcal{D}_X := \{p \in M : X \text{ is differentiable at } p\}.$$

Locally Lipschitz continuous vector fields are in general non-differentiable, however, they are almost everywhere differentiable with respect to the Riemannian measure (see the concept of Riemannian measure in [78, p. 61]), i.e., the set $M \backslash \mathcal{D}_X$ has measure zero. This result follows from Rademacher theorem, which is one of our contributions. A version of this theorem for locally Lipschitz continuous vector fields is given below.

Theorem 5.1.5 *If X is a locally Lipschitz continuous vector field on M, then X is almost everywhere differentiable on M.*

Proof. As M is a n-dimensional smooth manifold then the tangent bundle TM is $2n$-dimensional smooth manifold. First note that Theorem 5.1.3 implies that X is a continuous vector field. Let (U, φ) and (W, ψ) be smooth charts for M and TM, respectively, such that $X(U) \subseteq W$ and consider the composite mapping $\psi \circ X \circ \varphi^{-1} : \varphi(U) \to \psi(W)$. We proceed to prove that the mapping $\psi \circ X \circ \varphi^{-1}$ is locally Lipschitz continuous on $\varphi(U)$. According to [84, Proposition 6.10, p. 63], we obtain that the coordinate mappings $\varphi^{-1} : \varphi(U) \to U$ and $\psi : W \to \psi(W)$ are diffeomorphisms and, in particular, continuously differentiable. Take $z \in \varphi(U)$ and $\rho > 0$ such that $B_\rho[z] \subset \varphi(U)$. Since $B_\rho[z]$ is a compact set and the derivative of φ^{-1} is a continuous mapping in $B_\rho[z]$, from Mean Value Inequality (see [5, Theorem 2.14]) there exists $L_1 > 0$ such that

$$d(\varphi^{-1}(x), \varphi^{-1}(y)) \leq L_1\, \hat{d}(x,y), \qquad \forall\, x, y \in B_\rho[z],$$

where $\hat{d}$ is the Euclidean distance in $\mathbb{R}^n$. On the other hand, Theorem 5.1.3 implies that X is a locally metric Lipschitz continuous vector field on $\varphi(U)$, then shrinking $\rho > 0$ if necessary, we conclude that there exists $L_2 > 0$ such that

$$d_{TM}\left(X \circ \varphi^{-1}(x), X \circ \varphi^{-1}(y)\right) \leq L_2\, d\left(\varphi^{-1}(x), \varphi^{-1}(y)\right), \qquad \forall\, x, y \in B_\rho[z].$$

Because $X(\varphi^{-1}(B_\rho[z]))$ is a compact set and the derivative of ψ is a continuous mapping in $X(\varphi^{-1}(B_\rho[z]))$ again using Mean Value Inequality there exists $L_3 > 0$ such that

$$\tilde{d}(\psi \circ X \circ \varphi^{-1}(x), \psi \circ X \circ \varphi^{-1}(y)) \leq L_3\, d_{TM}(X \circ \varphi^{-1}(x), X \circ \varphi^{-1}(y)), \qquad \forall\, x, y \in B_\rho[z],$$

where $\tilde{d}$ is the Euclidean distance in $\mathbb{R}^{2n}$. Combining the three last inequalities, we obtain that

$$\tilde{d}(\psi \circ X \circ \varphi^{-1}(x), \psi \circ X \circ \varphi^{-1}(y)) \leq \widetilde{L}\, \hat{d}(x, y), \qquad \forall\, x, y \in B_\rho[z],$$

where $\widetilde{L} = L_1 L_2 L_3 > 0$. Hence, the mapping $\psi \circ X \circ \varphi^{-1}$ is locally Lipschitz continuous on $\varphi(U) \subseteq \mathbb{R}^n$. Therefore, from Rademacher theorem, see [32, Theorem 2, p. 81], we obtain that $\psi \circ X \circ \varphi^{-1}$ is almost everywhere differentiable on $\varphi(U)$. Since the charts (U, φ) and (W, ψ) are arbitrary, we conclude that X is almost everywhere differentiable on M. ■

Based on the definition presented in [48], we introduce the concept of *Clarke generalized covariant derivative* of a locally Lipschitz continuous vector field and explore some of its properties. For a comprehensive study about Clarke generalized Jacobian in linear space, see, for example, [18].

Definition 5.1.6 *The Clarke generalized covariant derivative of a locally Lipschitz continuous vector field X is a set-valued mapping $\partial X : M \rightrightarrows TM$ defined as*

$$\partial X(p) := \mathrm{co}\left\{ H \in \mathcal{L}(T_pM) :\ \exists\, \{p_k\} \subset \mathcal{D}_X,\ \lim_{k \to +\infty} p_k = p,\ H = \lim_{k \to +\infty} P_{p_k p} \nabla X(p_k) \right\},$$

where "co" *represents the convex hull and $\mathcal{L}(T_pM)$ denotes the vector space consisting of all linear operator from T_pM to T_pM.*

From Definition 5.1.6 and [35, Corollary 3.1], it is clear that if X is differentiable near p, and its covariant derivative is continuous at p, then $\partial X(p) = \{\nabla X(p)\}$. In the following proposition, we show important results of the Clarke generalized covariant derivative. In particular, that $\partial X(p)$ is a nonempty subset for all $p \in M$, and that the set-valued mapping ∂X is locally bounded and closed, which is a generalization of [18, Proposition 2.6.2, items (a), (b) and (c), p. 70]. These results will be very useful throughout this chapter. Similar results have already been extended to functions defined on a Riemannian manifold, see [54, Theorem 2.9].

Proposition 5.1.7 *Let X be locally Lipschitz continuous vector field on M. The following statements are valid for any $p \in M$:*

(i) $\partial X(p)$ is a nonempty, convex and compact subset of $\mathcal{L}(T_pM)$;

(ii) the set-valued mapping $\partial X : M \rightrightarrows TM$ is locally bounded, i.e., for all $\delta > 0$ there exists a $L > 0$ such that for all $q \in B_\delta(p)$ and $V \in \partial X(q)$, there holds $\|V\| \leq L$;

(iii) the mapping ∂X is upper semicontinuous at p, i.e., for every scalar $\epsilon > 0$ there exists a $0 < \delta < r_p$, and such that for all $q \in B_\delta(p)$,

$$P_{qp}\partial X(q) \subset \partial X(p) + B_\epsilon(0),$$

where $B_\epsilon(0) := \{V \in \mathcal{L}(T_pM) : \ \|V\| < \epsilon\}$. Consequently, the set-valued mapping ∂X is closed at p, i.e., if $\lim_{k\to+\infty} p_k = p$, $V_k \in \partial X(p_k)$ for all $k = 0, 1, \ldots,$ and $\lim_{k\to+\infty} P_{p_kp}V_k = V$, then $V \in \partial X(p)$.

Proof. To prove item (i), we define the following auxiliary set

$$\partial_B X(p) := \left\{ H \in \mathcal{L}(T_pM) : \ \exists \{p_k\} \subset \mathcal{D}_X, \ \lim_{k\to+\infty} p_k = p, \ H = \lim_{k\to+\infty} P_{p_kp}\nabla X(p_k) \right\}.$$

As T_pM is a finite dimensional space and $\partial X(p)$ is the convex hull in $\mathcal{L}(T_pM)$ of the set $\partial_B X(p)$, then $\partial X(p)$ must be convex. Our next goal is to prove that $\partial X(p)$ is compact. Owing to the convex hull of a compact set be compact, it is sufficient to prove that $\partial_B X(p)$ is bounded and closed. Our first task is to prove that $\partial_B X(p)$ is bounded. For this end, take $p \in \mathcal{D}_X$ and $v \in T_pM$. Because $\nabla X(p)v = \nabla X(p, v)$ using definition (2.4), the fact that X is a locally Lipschitz continuous vector field on M, and the definition of the exponential mapping, we obtain that

$$\|\nabla X(p)v\| = \lim_{t\to 0^+} \left\| \frac{1}{t}\left[P_{\exp_p(tv)p} X(\exp_p(tv)) - X(p) \right] \right\| \leq L\|v\|,$$

where $L > 0$ is the Lipschitz constant of X around p. Hence, from Definition 2.2.8 we conclude that $\|\nabla X(p)\| \leq L$, which implies that $\partial_B X(p)$ is a bounded set. To prove that $\partial_B X(p)$ is closed, let $\{H_\ell\}$ be a sequence in $\partial_B X(p)$ such that $\lim_{\ell\to+\infty} H_\ell = H$. Because $\{H_\ell\} \subset \partial_B X(p)$ there exists a sequence $\{p_{k,\ell}\}$ such that

$$\lim_{k\to+\infty} p_{k,\ell} = p \ \text{ and } \ \lim_{k\to+\infty} P_{p_{k,\ell}p}\nabla X(p_{k,\ell}) = H_\ell,$$

for each fixed ℓ. Therefore, $\lim_{k\to+\infty} p_{k,k} = p$ and $\lim_{k\to+\infty} P_{p_{k,k}p}\nabla X(p_{k,k}) = H$, and then $H \in \partial_B X(p)$. Consequently, $\partial_B X(p)$ is a compact set. To prove that $\partial X(p)$ is a nonempty set, first note that Theorem 5.1.5 implies that X is almost everywhere differentiable on M, i.e., the set $M\backslash\mathcal{D}_X$ has measure zero. According to [63, Proposition 6.8, p. 128], $\mathcal{D}_X$ is dense in M. Then, for any fixed point $p \in M$ there exists a sequence $\{p_k\} \subset \mathcal{D}_X$ that converges to p. Since ∇X is bounded in norm by the Lipschitz constant and the parallel transport is an isometry, the sequence $\{P_{p_k p}\nabla X(p_k)\}$ must have at least one accumulation point, and thus $\partial X(p)$ is indeed a nonempty set. To prove item (ii), take $\delta > 0$, $p \in M$ and $L > 0$ the Lipschitz constant of X around p. The same argument used to prove item (i) shows that $\|\nabla X(\bar{q})\| \leq L$ for all $\bar{q} \in B_\delta(p) \cap \mathcal{D}_X$. Let $q \in B_\delta(p)$ and $V \in \partial X(q)$. Then, there exist $H_1, \ldots, H_m \in \partial_B X(q)$ and $\alpha_1, \ldots, \alpha_m \in [0, 1]$ such that $V = \sum_{\ell=1}^m \alpha_\ell H_\ell$ and $\sum_{\ell=1}^m \alpha_\ell = 1$. As $H_1, \ldots, H_m \in \partial_B X(q)$ there exists a sequence $\{q_{k,\ell}\} \subset B_\delta(p) \cap \mathcal{D}_X$ with $\lim_{k\to+\infty} q_{k,\ell} = q$ such that

$$V = \sum_{\ell=1}^{m} \alpha_\ell \lim_{k\to+\infty} P_{q_{k,\ell}q}\nabla X(q_{k,\ell}).$$

Owing to $\{q_{k,\ell}\} \subset B_\delta(p) \cap \mathcal{D}_X$ we have $\|\nabla X(q_{k,\ell})\| \leq L$. Therefore, using that the parallel transport is an isometry, the properties of the norm and that $\sum_{\ell=1}^m \alpha_\ell = 1$, we conclude of the last equality that

$$\|V\| = \left\| \sum_{\ell=1}^{m} \alpha_\ell \lim_{k\to+\infty} P_{q_{k,\ell}q}\nabla X(q_{k,\ell}) \right\| \leq \sum_{\ell=1}^{m} \alpha_\ell \lim_{k\to+\infty} \left\| P_{q_{k,\ell}q}\nabla X(q_{k,\ell}) \right\| \leq L,$$

which is the desired inequality. To prove item (iii), suppose by contradiction that for a given $\epsilon > 0$ and all $0 < \delta < r_p$ there exists $q \in B_\delta(p)$ such that

$$P_{qp}\partial X(q) \not\subset \partial X(p) + B_\epsilon(0).$$

Hence, there exists a sequence $\{q_k\} \subset \mathcal{D}_X$ such that $\lim_{k\to+\infty} q_k = p$ and $P_{q_k p}\nabla X(q_k) \notin \partial X(p) + B_\epsilon(0)$. On the other hand, item (ii) implies that ∂X is a locally bounded set-valued mapping. As the parallel transport is an isometry we have $\{P_{q_k p}\nabla X(q_k)\}$ is a bounded sequence. Thus, we can extract $\{P_{q_{k_\ell} p}\nabla X(q_{k_\ell})\}$ a convergent subsequence of $\{P_{q_k p}\nabla X(q_k)\}$, let us say that $\{P_{q_{k_\ell} p}\nabla X(q_{k_\ell})\}$ converges to some H. From Definition 5.1.6 we obtain that $H \in \partial X(p)$, which is a contradiction. Therefore, ∂X is a upper semicontinuous mapping at p. The last part of item (iii) it is an immediate consequence of the first part, and the proof of the proposition is complete. ■

5.2 Nonsmooth Newton method

In this section, we present the nonsmooth Newton method for finding a singularity of locally Lipschitz continuous vector fields X defined on a Riemannian manifold M, i.e., for solving the problem (1.6). We study the local and semi-local properties of a sequence generated by the method. The nonsmooth Newton algorithm for solving the problem (1.6), with $p_0 \in M$

as the input data, is formally described as follows.

Algorithm 5.2.1 Nonsmooth Newton method

Step 0. Let $p_0 \in M$ be given, and set $k = 0$.

Step 1. If $X(p_k) = 0$, **stop**.

Step 2. Choose a $V_k \in \partial X(p_k)$ and compute

$$p_{k+1} = \exp_{p_k}(-V_k^{-1} X(p_k)). \tag{5.2}$$

Step 3. Set $k \leftarrow k + 1$, and go to **Step 1**.

This method is a natural extension to the Riemannian setting of the Newton method introduced in [76]. Note that to guarantee the well-definedness of the method, there are two issues which deserve attention in each iteration k. The Clarke generalized covariant derivative $\partial X(p_k)$ must be a nonempty subset, which has already been proven in item (i) of Proposition 5.1.7, and all $V_k \in \partial X(p_k)$ must be nonsingular. In the following section, we study the well-definedness and convergence of a sequence generated by the nonsmooth Newton method.

5.3 Local convergence analysis

In this section, we present the local convergence analysis of Algorithm 5.2.1. To this end, we assume that $p_* \in M$ is a solution of problem (1.6). First, we show that under some assumptions, the sequence generated by this algorithm starting from a suitable neighborhood of p_* is well-defined and converges to p_* with rate of the order of $1 + \mu$. We begin by introducing the concept of regularity.

Definition 5.3.1 *We say that a vector field X on M is regular at $p \in M$ if all $V_p \in \partial X(p)$ are nonsingular. If X is regular at every point of $\Omega \subseteq M$, we say that X is regular on Ω.*

In the following, we study the behavior of a sequence generated by the nonsmooth Newton method for a special class of vector field in a neighborhood of a regular point. For this purpose, we assume that X *is a locally Lipschitz continuous vector field on* M. Consider the following condition:

A1. Let $\bar{p} \in M$, $0 < \delta < r_{\bar{p}}$, X be regular on $B_\delta(\bar{p})$, $\lambda_{\bar{p}} \geq \max\{\|V_{\bar{p}}^{-1}\| : \ V_{\bar{p}} \in \partial X(\bar{p})\}$ and $\epsilon > 0$ satisfying $\epsilon\lambda_{\bar{p}} < 1$. Moreover, for all $p, q \in B_\delta(\bar{p})$ and $V_p \in \partial X(p)$ there hold

$$\|V_p^{-1}\| \leq \frac{\lambda_{\bar{p}}}{1-\epsilon\lambda_{\bar{p}}}, \tag{5.3}$$

$$\left\|X(q) - P_{pq}\left[X(p) + V_p \exp_p^{-1} q\right]\right\| \leq \epsilon\, d(p,q)^{1+\mu}, \qquad 0 \leq \mu \leq 1. \tag{5.4}$$

Let $0 < \delta < r_{\bar{p}}$ be given by above assumption and $N_X \colon B_\delta(\bar{p}) \rightrightarrows M$ be the *Newton iteration mapping* for the vector field X defined by

$$N_X(p) := \left\{\exp_p(-V_p^{-1}X(p)) : \ V_p \in \partial X(p)\right\}.$$

The above assumption guarantee, in particular, that X is regular in a neighborhood of $\bar{p}$ and, consequently, the Newton iteration mapping is well-defined. Therefore, one can apply a single Newton iteration on any $p \in B_\delta(\bar{p})$ to obtain $N_X(p)$, which may be not included in $B_\delta(\bar{p})$. Thus, this is enough to guarantee the well-definedness of only one iteration. In the following result, we establish that the Newtonian iterations may be repeated indefinitely in a suitable neighborhood of $\bar{p}$.

Lemma 5.3.2 *Suppose that $p_* \in M$ is a solution of problem* (1.6)*, X satisfies* **A1** *and the constants $\epsilon > 0$, $0 < \delta < r_{p_*}$, and $0 \leq \mu \leq 1$ satisfy $\epsilon\lambda_{p_*}(1 + \delta^\mu K_{p_*}) < 1$. Then, there exists $\hat{\delta} > 0$ such that X is regular on $B_{\hat{\delta}}(p_*)$ and*

$$d\left(\exp_p(-V_p^{-1}X(p)), p_*\right) \leq \frac{\epsilon\lambda_{p_*}K_{p_*}}{1-\epsilon\lambda_{p_*}} d(p,p_*)^{1+\mu}, \qquad \forall\ p \in B_{\hat{\delta}}(p_*), \quad \forall\ V_p \in \partial X(p). \tag{5.5}$$

Consequently, N_X is well-defined on $B_{\hat{\delta}}(p_*)$ and $N_X(p) \subset B_{\hat{\delta}}(p_*)$ for all $p \in B_{\hat{\delta}}(p_*)$.

Proof. Assume without loss of generality that X satisfies **A1** with $\bar{p} = p_*$ and $q = p_*$. Consider the constants $\epsilon > 0$, $0 < \delta < r_{p_*}$ and $0 \leq \mu \leq 1$. Since $X(p_*) = 0$ and the parallel transport is an isometry, we conclude that

$$\begin{aligned} \left\|V_p^{-1}X(p) + \exp_p^{-1} p_*\right\| &\leq \left\|V_p^{-1}\right\| \left\|X(p_*) - P_{pp_*}\left[X(p) + V_p \exp_p^{-1} p_*\right]\right\| \\ &\leq \frac{\epsilon\lambda_{p_*}}{1-\epsilon\lambda_{p_*}} d(p,p_*)^{1+\mu}, \end{aligned} \tag{5.6}$$

for all $p \in B_\delta(p_*)$ and $V_p \in \partial X(p)$. Hence, (5.6) implies that there exists $0 < \hat{\delta} < \delta$ such that $\|V_p^{-1}X(p) + \exp_p^{-1} p_*\| \leq r_{p_*}$ for all $p \in B_{\hat{\delta}}(p_*)$ and $V_p \in \partial X(p)$. Thus, considering that $\|\exp_p^{-1} p_*\| = d(p,p_*) < r_{p_*}$, we can use Definition 2.2.6 with $p = p_*$, $q = p$, $u = -V_p^{-1}X(p)$ and $v = \exp_p^{-1} p_*$ to obtain that

$$d\left(\exp_p(-V_p^{-1}X(p)), p_*\right) \leq K_{p_*}\left\|-V_p^{-1}X(p) - \exp_p^{-1} p_*\right\|,$$

for all $p \in B_{\hat\delta}(p_*)$ and $V_p \in \partial X(p)$. Therefore, the combination of the last inequality with (5.6) yields (5.5). Owing to $0 < \hat\delta < \delta$ and X be regular on $B_\delta(p_*)$, we conclude that N_X is well-defined on $B_{\hat\delta}(p_*)$. Moreover, since $\epsilon\lambda_{p_*}(1 + \delta^\mu K_{p_*}) < 1$ and $0 < \hat\delta < \delta$, we have from (5.5) that $d\left(\exp_p(-V_p^{-1}X(p)), p_*\right) < d(p, p_*)$ for all $p \in B_{\hat\delta}(p_*)$ and $V_p \in \partial X(p)$. Thus, we obtain that $N_X(p) \subset B_{\hat\delta}(p_*)$ for all $p \in B_{\hat\delta}(p_*)$, and the proof of the lemma is complete. ■

Now, we are ready to establish the main result of this section, its proof is a straight application of Lemma 5.3.2.

Theorem 5.3.3 *Suppose that $p_* \in M$ is a solution of problem* (1.6)*, X satisfies* **A1** *and the constants $\epsilon > 0$, $0 < \delta < r_{p_*}$, and $0 \le \mu \le 1$ satisfy $\epsilon\lambda_{p_*}(1 + \delta^\mu K_{p_*}) < 1$. Then, there exists $0 < \hat\delta < \delta$ such that for each $p_0 \in B_{\hat\delta}(p_*)\backslash\{p_*\}$ the sequence $\{p_k\}$ in Algorithm 5.2.1 is well-defined, belongs to $B_{\hat\delta}(p_*)$ and converges to p_* with order $1 + \mu$ as follows*

$$d\left(p_{k+1}, p_*\right) \le \frac{\epsilon\lambda_{p_*}K_{p_*}}{1 - \epsilon\lambda_{p_*}} d(p_k, p_*)^{1+\mu}, \qquad k = 0, 1, \ldots. \tag{5.7}$$

Proof. The definition of the Newton iteration mapping N_X implies that the sequence generated by Algorithm 5.2.1 is equivalently stated as

$$p_{k+1} \in N_X(p_k), \qquad k = 0, 1, \ldots. \tag{5.8}$$

Hence, by using (5.8), we can apply Lemma 5.3.2 to conclude that there exists $0 < \hat\delta < \delta$ such that if $p_0 \in B_{\hat\delta}(p_*)\backslash\{p_*\}$, then the sequence $\{p_k\}$ in Algorithm 5.2.1 is well-defined, belongs to $B_{\hat\delta}(p_*)$ and satisfies the inequality (5.7). Because $\{p_k\}$ belongs to $B_{\hat\delta}(p_*)$ and $\epsilon\lambda_{p_*}(1 + \delta^\mu K_{p_*}) < 1$, we obtain from (5.7) that

$$d\left(p_{k+1}, p_*\right) < \frac{\epsilon\lambda_{p_*}\hat\delta^\mu K_{p_*}}{1 - \epsilon\lambda_{p_*}} d(p_k, p_*) < d(p_k, p_*), \qquad k = 0, 1, \ldots.$$

Therefore, we conclude that $\{p_k\}$ converges to p_* with order $1 + \mu$ as (5.7). ■

Remark 5.3.4 Note that if $\mu = 0$ in Theorem 5.3.3, then the inequality (5.7) holds for any $\epsilon > 0$ satisfying $\epsilon\lambda_{p_*}(1 + K_{p_*}) < 1$, independently of the scalar $\hat\delta > 0$. Therefore, (5.7) implies that the sequence $\{p_k\}$ converges superlinearly to p_*.

5.3.1 Local convergence for semismooth vector fields

In this section, we present a local convergence theorem for the nonsmooth Newton method for finding a singularity of semismooth vector fields. Semismoothness in Euclidean setting was originally introduced by Mifflin [68] for scalar-valued functions and subsequently extended by Qi and Sun [76] for vector-valued functions. The extension of the concept of semismoothness to the Riemannian settings is presented in this section. As occur in the Euclidean context, semismooth vector fields are in general nonsmooth. However, as we shall show, the Newton method is still applicable and converges locally with superlinear rate to a regular solution. Before, we state formally the concept of semismoothness in the Riemannian setting, let us first show that locally Lipschitz continuous vector fields are regular near regular points. The statement of the result is as follows.

Lemma 5.3.5 *Let X be a locally Lipschitz continuous vector field on M. Assume that X is regular at $p_* \in M$ and let $\lambda_{p_*} \geq \max\{\|V_{p_*}^{-1}\| : \; V_{p_*} \in \partial X(p_*)\}$. Then, for every $\epsilon > 0$ satisfying $\epsilon\lambda_{p_*} < 1$ there exists $0 < \delta < r_{p_*}$ such that X is regular on $B_\delta(p_*)$ and*

$$\|V_p^{-1}\| \leq \frac{\lambda_{p_*}}{1 - \epsilon\lambda_{p_*}}, \qquad \forall\, p \in B_\delta(p_*), \quad \forall\, V_p \in \partial X(p). \tag{5.9}$$

Proof. Let $\epsilon > 0$ such that $\epsilon\lambda_{p_*} < 1$. Because X is a locally Lipschitz continuous vector field, it follows from item (iii) of Proposition 5.1.7 that there exists a $0 < \delta < r_{p_*}$ such that $P_{pp_*}\partial X(p) \subset \partial X(p_*) + \{V \in \mathcal{L}(T_{p_*}M) : \; \|V\| < \epsilon\}$ for all $p \in B_\delta(p_*)$, i.e.,

$$\partial X(p) \subset \{V \in \mathcal{L}(T_pM) : \; \|P_{pp_*}V - V_{p_*}\| < \epsilon \text{ for some } V_{p_*} \in \partial X(p_*)\}, \qquad \forall\, p \in B_\delta(p_*).$$

This inclusion implies that for each $p \in B_\delta(p_*)$ and $V_p \in \partial X(p)$, there exists $V_{p_*} \in \partial X(p_*)$ nonsingular such that $\|V_{p_*}^{-1}\|\|P_{pp_*}V_p - V_{p_*}\| < \epsilon\lambda_{p_*} < 1$. Thus, taking into account that the parallel transport is an isometry, it follows from Lemma 2.2.9 that V_p is nonsingular and

$$\|V_p^{-1}\| \leq \frac{\|V_{p_*}^{-1}\|}{1 - \|V_{p_*}^{-1}\|\|P_{pp_*}V_p - V_{p_*}\|}.$$

Therefore, considering that $\|V_{p_*}^{-1}\| \leq \lambda_{p_*}$ and $\|P_{pp_*}V_p - V_{p_*}\| < \epsilon$, (5.9) follows. ■

In the following, let us present a class of vector fields satisfying the condition **A1**, namely the semismooth vector fields and μ-order semismooth vector fields. There exist, in the Euclidean context, several equivalent definitions of the concept of semismoothness, see, for example, [76]; see also [34, Definition 7.4.2, p. 677]. In the present thesis, we extend to the Riemannian settings the concept of semismoothness adopted in [31, p. 411].

Definition 5.3.6 *A vector field X on M, which is locally Lipschitz continuous at p_* and directionally differentiable at $p \in M$, for all directions in T_pM, is said to be semismooth at $p_* \in M$ when for every $\epsilon > 0$ there exists $0 < \delta < r_{p_*}$ such that*

$$\left\|X(p_*) - P_{pp_*}\left[X(p) + V_p \exp_p^{-1} p_*\right]\right\| \leq \epsilon\, d(p, p_*),$$

for all $p \in B_\delta(p_*)$ and $V_p \in \partial X(p)$. The vector field X is said to be μ-order semismooth at $p_* \in M$, for $0 < \mu \leq 1$, when there exist $\epsilon > 0$ and $0 < \delta < r_{p_*}$ such that

$$\left\|X(p_*) - P_{pp_*}\left[X(p) + V_p \exp_p^{-1} p_*\right]\right\| \leq \epsilon\, d(p, p_*)^{1+\mu}, \tag{5.10}$$

for all $p \in B_\delta(p_*)$ and $V_p \in \partial X(p)$.

Next, we state and prove the local convergence theorem for the nonsmooth Newton method for finding a singularity of semismooth vector fields and μ-order semismooth vector fields.

Theorem 5.3.7 *Let X be a locally Lipschitz continuous vector field on M and $p_* \in M$ be a solution of problem* (1.6)*. Assume that X is semismooth and regular at p_*. Then, there exists a $\delta > 0$ such that for each $p_0 \in B_\delta(p_*)\backslash\{p_*\}$, $\{p_k\}$ generated by Algorithm 5.2.1, is well-defined, belongs to $B_\delta(p_*)$ and converges superlinearly to p_*. In addition, if X is μ-order semismooth at p_*, then the convergence of $\{p_k\}$ to p_* is of the order of $1+\mu$.*

Proof. Owing to X be semismooth and regular at $p_* \in M$, we can take $\lambda_{p_*} \geq \max\{\|V_{p_*}^{-1}\| : V_{p_*} \in \partial X(p_*)\}$. Consider $\epsilon > 0$ satisfying $\epsilon\lambda_{p_*}(1 + K_{p_*}) < 1$. Thus, from Lemma 5.3.5 and Definition 5.3.6, we can take $\delta > 0$ such that (5.3) and (5.4) hold for $\mu = 0$. Hence, condition **A1** holds with $\bar{p} = p_*$ and $q = p_*$ for all $p \in B_\delta(p_*)$ and $\mu = 0$. Therefore, applying Theorem 5.3.3, we obtain that there exists $0 < \hat{\delta} < \delta$ such that every sequence $\{p_k\}$ generated by Algorithm 5.2.1 with $p_0 \in B_{\hat{\delta}}(p_*)\backslash\{p_*\}$ belongs to $B_{\hat{\delta}}(p_*)$ and satisfies the inequality (5.7). Hence, we have

$$\frac{d(p_{k+1}, p_*)}{d(p_k, p_*)} \leq \frac{\epsilon\lambda_{p_*}K_{p_*}}{1 - \epsilon\lambda_{p_*}}, \qquad k = 0, 1, \ldots.$$

Since the last inequality holds for any ϵ such that $0 < \epsilon < 1/(\lambda_{p_*}(1 + K_{p_*}))$, we conclude that $\{p_k\}$ converges superlinearly to p_*. The proof of the second part is similar. Indeed, for a given $\epsilon > 0$ with $\epsilon\lambda_{p_*} < 1$, take $\delta > 0$ satisfying $\epsilon\lambda_{p_*}(1 + \delta^\mu K_{p_*}) < 1$ and such that (5.9) and (5.10) hold. Then, we can apply Theorem 5.3.3 and the proof follows. ■

We remark that with some adjustments Theorem 5.3.7 reduces to some well-known results.

Remark 5.3.8 It is well-known that the Newton method and its variants are quite efficient for finding zero on nonlinear functions in Euclidean settings. This is because they have an excellent convergence rate in a neighborhood of a zero. It was shown in [76] that for a class of nonsmooth functions, namely semismooth functions, the convergence of the nonsmooth Newton method still is guaranteed. The above theorem, allows us to conclude that the generalization of the nonsmooth Newton method from the linear context to Riemannian settings for finding singularities of semismooth vector fields still preserves its main convergence properties. It is worth mentioning that if X is continuously differentiable, then Theorem 5.3.7 reduces to the [35, Theorem 3.1]. If $M = \mathbb{R}^n$, then Theorem 5.3.7 reduces to first part of [76, Theorem 3.2]; see also [34, Theorem 7.5.3, p. 693]. Finally, if X is continuously differentiable and $M = \mathbb{R}^n$, then the theorem above reduces to the first part of [12, Proposition 1.4.1, p. 90].

5.4 Semi-local convergence analysis

In this section, we state and prove the Kantorovich-type theorem for the nonsmooth Newton method. This theorem ensures that the sequence generated by the method converges towards a singularity of the vector field by using semi-local conditions. It is worth mentioning that the theorem does not require a priori existence of a singularity, proving instead the existence of the singularity and its uniqueness on some region. The statement of the theorem is as follows.

Theorem 5.4.1 *Let X be a locally Lipschitz continuous vector field on M and $p_0 \in M$. Suppose that X satisfies* **A1** *with $\bar{p} = p_0$, $\mu = 0$ and $\delta > \bar{\delta}$. Moreover, $B_{\bar{\delta}}(p_0) \subset M$ is a totally normal neighborhood of the point p_0, and the constants $\lambda_{p_0} > 0$, $\epsilon > 0$ and $0 < \bar{\delta} < r_{p_0}$ are such that*

$$\epsilon\lambda_{p_0} < \frac{1}{2}, \qquad\qquad \frac{\lambda_{p_0}}{1 - 2\epsilon\lambda_{p_0}} \|X(p_0)\| \leq \bar{\delta}. \tag{5.11}$$

Then, the sequence $\{p_k\}$ in Algorithm 5.2.1 is well-defined, belongs to $B_{\bar{\delta}}(p_0)$ and converge towards the unique solution p_* of problem (1.6) *in $B_{\bar{\delta}}[p_0]$. Furthermore, the following error estimate holds*

$$d(p_k, p_*) \leq \frac{\epsilon\lambda_{p_0}}{1 - 2\epsilon\lambda_{p_0}} d(p_k, p_{k-1}), \qquad k = 1, 2, \ldots. \tag{5.12}$$

Proof. Firstly, let us prove by induction that the sequence $\{p_k\}$ in Algorithm 5.2.1 is well-defined, belongs to $B_{\bar{\delta}}(p_0)$ and satisfies

$$d(p_{k+1}, p_k) \leq \left(\frac{\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}}\right)^k \bar{\delta} \left(\frac{1 - 2\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}}\right), \qquad k = 0, 1, \ldots. \tag{5.13}$$

Let $V_0 \in \partial X(p_0)$ and note that the condition **A1** implies that V_0 is nonsingular and $\|V_0^{-1}\| \leq \lambda_{p_0}/(1 - \epsilon\lambda_{p_0})$. Hence, by using (5.2), we obtain that the iterate p_1 is well-defined. Furthermore, (5.2), the definition of the exponential mapping, properties of the norm and the inequalities in (5.11) imply that

$$d(p_1, p_0) = d\left(\exp_{p_0}(-V_0^{-1}X(p_0)), p_0\right) \leq \\ \left\|-V_0^{-1}X(p_0)\right\| \leq \frac{\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \|X(p_0)\| \leq \bar{\delta} \left(\frac{1 - 2\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}}\right) < \bar{\delta}.$$

Therefore, the iterate p_1 is well-defined, belongs to $B_{\bar{\delta}}(p_0)$ and (5.13) holds for $k = 0$. Assume by induction that the iterates $p_1, \ldots, p_{\ell-1}$ are well-defined, belongs to $B_{\bar{\delta}}(p_0)$ and (5.13) holds for $k = 1, \ldots, \ell - 1$. As $p_{\ell-1} \in B_{\bar{\delta}}(p_0)$ it follows from condition **A1** that $V_{\ell-1}$ is nonsingular and, consequently, the iterate p_ℓ is well-defined. Thus, using the triangular inequality and the induction assumption, we have

$$d(p_\ell, p_0) \le \sum_{j=1}^{\ell} d(p_j, p_{j-1}) \le \bar{\delta} \left(\frac{1 - 2\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right) \sum_{j=1}^{\ell} \left(\frac{\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right)^{j-1} < \bar{\delta}. \tag{5.14}$$

This implies that $p_\ell \in B_{\bar{\delta}}(p_0)$. Since $p_\ell \in B_{\bar{\delta}}(p_0)$ it follows from condition **A1** that V_ℓ is nonsingular and, consequently, the iterate $p_{\ell+1}$ is well-defined. Moreover, also follows from **A1** that $\|V_\ell^{-1}\| \le \lambda_{p_0}/(1 - \epsilon\lambda_{p_0})$. Thus, by using (5.2), the definition of the exponential mapping and properties of the norm, we have

$$d(p_{\ell+1}, p_\ell) = d\left(\exp_{p_\ell}(-V_\ell^{-1} X(p_\ell)), p_\ell\right) \le \left\| -V_\ell^{-1} X(p_\ell) \right\| \le \frac{\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \left\| X(p_\ell) \right\|. \tag{5.15}$$

On the other hand, considering that $B_{\bar{\delta}}(p_0)$ is a totally normal neighborhood and that the iterates $p_{\ell-1}, p_\ell \in B_{\bar{\delta}}(p_0)$, we conclude after some algebraic manipulations that

$$\|X(p_\ell)\| \le \left\| X(p_\ell) - P_{p_{\ell-1} p_\ell} \left[X(p_{\ell-1}) + V_{\ell-1} \exp^{-1}_{p_{\ell-1}} p_\ell \right] \right\| + \left\| X(p_{\ell-1}) + V_{\ell-1} \exp^{-1}_{p_{\ell-1}} p_\ell \right\|.$$

Taking into account that (5.2) implies $X(p_{\ell-1}) + V_{\ell-1} \exp^{-1}_{p_{\ell-1}} p_\ell = 0$ the last inequality becomes

$$\|X(p_\ell)\| \le \left\| X(p_\ell) - P_{p_{\ell-1} p_\ell} \left[X(p_{\ell-1}) + V_{\ell-1} \exp^{-1}_{p_{\ell-1}} p_\ell \right] \right\|.$$

Using **A1** with $q = p_\ell$, $p = p_{\ell-1}$ and $V_p = V_{\ell-1}$, it follows from the latter inequality that

$$\|X(p_\ell)\| \le \epsilon d(p_\ell, p_{\ell-1}).$$

Therefore, combining the inequality $\|X(p_\ell)\| \le \epsilon d(p_\ell, p_{\ell-1})$ with (5.15) and by using the induction assumption, we conclude that

$$d(p_{\ell+1}, p_\ell) \le \frac{\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} d(p_\ell, p_{\ell-1}) \le \left(\frac{\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right)^{\ell} \bar{\delta} \left(\frac{1 - 2\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right), \tag{5.16}$$

and the induction proof is complete. Hence, using (5.16) and the same argument used to prove (5.14), we obtain that $p_{\ell+1} \in B_{\bar{\delta}}(p_0)$. Therefore, the Newton iterates are well-defined, belongs to $B_{\bar{\delta}}(p_0)$ and satisfy (5.13). We proceed to prove that the sequence $\{p_k\}$ converges. Indeed, using the triangular inequality, and (5.13) for any k and $s \in \{0, 1, \ldots\}$, we have

$$d(p_{k+s+1}, p_k) \le \sum_{j=k}^{k+s} d(p_{j+1}, p_j) \le \bar{\delta} \left(\frac{1 - 2\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right) \sum_{j=k}^{k+s} \left(\frac{\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right)^{j} < \bar{\delta} \left(\frac{\epsilon\lambda_{p_0}}{1 - \epsilon\lambda_{p_0}} \right)^{k}.$$

Because $2\epsilon\lambda_{p_0} < 1$, we conclude that $\{p_k\}$ is a Cauchy sequence. This implies that the sequence $\{p_k\}$ converges, let us say to some $p_* \in B_{\bar{\delta}}[p_0]$. Hence, owing to X be a locally Lipschitz continuous vector field, item (ii) of Proposition 5.1.7 implies that $\{V_k\}$ is bounded.

Therefore, using that X is continuous, the equality (5.2), some properties of the norm and that the sequence $\{p_k\}$ converges to p_*, we have

$$0 \leq \|X(p_*)\| = \lim_{k\to+\infty} \|X(p_k)\| = \lim_{k\to+\infty} \left\|-V_k \exp_{p_k}^{-1} p_{k+1}\right\| \leq \lim_{k\to+\infty} \|V_k\| \, d(p_{k+1}, p_k) = 0,$$

consequently, $X(p_*) = 0$. Now, we are going to prove the uniqueness of the solution in $B_{\bar{\delta}}[p_0]$. For this purpose, assume that $q \in B_{\bar{\delta}}[p_0]$ is such that $X(q) = 0$. Take $V_* \in \partial X(p_*)$, by assumption X is regular on $B_\delta(p_0)$ and $p_* \in B_{\bar{\delta}}[p_0] \subset B_\delta(p_0)$, then V_* is nonsingular. As $X(p_*) = 0$ and $X(q) = 0$, using condition **A1** with $p = p_*$ and $V_p = V_*$, and some manipulations, we obtain that

$$d(p_*, q) = \left\|V_*^{-1} V_* \exp_{p_*}^{-1} q\right\| \leq \\ \left\|V_*^{-1}\right\| \left[\left\|X(q) - P_{p_* q}\left[X(p_*) + V_* \exp_{p_*}^{-1} q\right]\right\| \right] \leq \frac{\epsilon\lambda_{p_0}}{1-\epsilon\lambda_{p_0}} d(p_*, q). \quad (5.17)$$

Because $\epsilon\lambda_{p_0} < 1/2$, we conclude that $d(p_*, q) = 0$, i.e., $q = p_*$. Therefore, p_* is the unique solution of problem (1.6) in $B_{\bar{\delta}}[p_0]$. It remains to show the error estimate, i.e., the inequality (5.12). First note that using the same arguments to establish the first inequality in (5.16), we can also prove that $d(p_{i+1}, p_i) \leq [\epsilon\lambda_{p_0}/(1-\epsilon\lambda_{p_0})]\, d(p_i, p_{i-1})$ for all $i = 1, 2, \ldots$, and thereby obtain

$$d(p_{k+j}, p_{k+j-1}) \leq \left(\frac{\epsilon\lambda_{p_0}}{1-\epsilon\lambda_{p_0}}\right)^j d(p_k, p_{k-1}), \qquad j = 1, 2, \ldots.$$

Hence, for any $s \in \{0, 1, \ldots\}$, we can use the triangular inequality and the last inequality to conclude that

$$d(p_{k+s+1}, p_k) \leq \sum_{j=k}^{k+s} d(p_{j+1}, p_j) \leq d(p_k, p_{k-1}) \sum_{j=1}^{s+1} \left(\frac{\epsilon\lambda_{p_0}}{1-\epsilon\lambda_{p_0}}\right)^j < \frac{\epsilon\lambda_{p_0}}{1-2\epsilon\lambda_{p_0}} d(p_k, p_{k-1}).$$

Taking the limit as s goes to $+\infty$, we obtain the inequality (5.12), and the proof of the theorem is complete. ■

Remark 5.4.2 It is worth pointing out that for $0 < \mu \leq 1$, we can not obtain (5.14) in order to assure the well-definedness of iterate $p_{\ell+1}$. Besides, for $0 < \mu \leq 1$, we can not obtain the uniqueness of the solution from (5.17). Therefore, a new argument will be needed to extend Theorem 5.4.1 for $0 < \mu \leq 1$.

5.5 Some examples

In this section, we present a class of examples of locally Lipschitz continuous vector fields on the sphere satisfying the condition **A1**. For this purpose, we begin by presenting some basic definitions about the geometry of the sphere. For further details, see [37,38] and references therein.

Let $\langle \cdot, \cdot \rangle$ be the *usual inner product on* $\mathbb{R}^{n+1}$, with corresponding *norm* denoted by $\| \cdot \|$. The *n-dimensional Euclidean sphere* and its *tangent hyperplane at a point* p are denoted, respectively, by

$$\mathbb{S}^n := \left\{ p = (p_1, \ldots, p_{n+1}) \in \mathbb{R}^{n+1} : \ \|p\| = 1 \right\}, \qquad T_p\mathbb{S}^n := \left\{ v \in \mathbb{R}^{n+1} : \ \langle p, v \rangle = 0 \right\}.$$

Denotes by I the $(n+1)\times(n+1)$ identity matrix. The *projection onto the tangent hyperplane* $T_p\mathbb{S}^n$ is the linear mapping defined by $I - pp^T : \mathbb{R}^{n+1} \to T_p\mathbb{S}^n$, where p^T denotes the transpose of the vector p. Let Ω be an open set in $\mathbb{R}^{n+1}$ such that $\mathbb{S}^n \subset \Omega$, and $Y : \Omega \to \mathbb{R}^{n+1}$ be any semismooth mapping; several examples can be found in [31,34,57]. Then, we define the vector field $X : \mathbb{S}^n \to \mathbb{R}^{n+1}$ as follows

$$X(p) := (I - pp^T)Y(p).$$

Note that $X(p) \in T_p\mathbb{S}^n$ for all $p \in \mathbb{S}^n$. The Clarke generalized covariant derivative of X at p is given by

$$\partial X(p) := \left(I - pp^T\right) \partial Y(p) - p^T Y(p) I, \tag{5.18}$$

where $\partial Y(p)$ is the Clarke generalized covariant derivative of Y at p. Therefore, all $V_p \in \partial X(p)$ is a linear mapping $V_p : T_p\mathbb{S}^n \to T_p\mathbb{S}^n$ given by $V_p := \left(I - pp^T\right) \tilde{V}_p - p^T Y(p) I$, where $\tilde{V}_p \in \partial Y(p)$. Since Y is a locally Lipschitz continuous mapping, from Rademacher theorem, see [32, Theorem 2, p. 81], we conclude that Y is almost everywhere differentiable. As $I - pp^T$ is a differentiable mapping, we obtain that X is almost everywhere differentiable. Using the fundamental theorem of calculus in Riemannian setting (see [41]), the fact that $\partial Y(p)$ is locally bounded and continuity of Y, we conclude that X is also locally Lipschitz continuous vector field. Assume that X is regular at $\bar{p} \in \Omega$ and let $\lambda_{\bar{p}} \geq \max\{\|V_{\bar{p}}^{-1}\| : \ V_{\bar{p}} \in \partial X(\bar{p})\}$. Then, from Lemma 5.3.5 for every $\epsilon > 0$ satisfying $\epsilon \lambda_{\bar{p}} < 1$, there exists $0 < \delta < \pi$ (where π is the injectivity radius of $\mathbb{S}^n$) such that X is regular on $B_\delta(\bar{p})$ and for all $p \in B_\delta(\bar{p})$ and $V_p \in \partial X(p)$ the following holds

$$\|V_p^{-1}\| \leq \frac{\lambda_{\bar{p}}}{1 - \epsilon \lambda_{\bar{p}}}.$$

This implies that inequality (5.3) holds. On the other hand, because X is a composition of semismooth mappings, we conclude that X is semismooth, see [57, Proposition 1.74, p. 54]. Hence, from Definition 5.3.6 inequality (5.4) holds. Therefore, the projected vector field X satisfies the condition **A1**. In the following, we present a concrete example.

Example 5.5.1 Let $Y : \mathbb{R}^2 \to \mathbb{R}^2$ be a semismooth mapping defined by $Y(p) := Ap - |p| - b$ with matrix $A = \text{diag}(4,3)$ and vector $b = (b_1, b_2) \in \mathbb{R}^2$, where $\text{diag}(p_1, p_2)$ denotes a

2×2 diagonal matrix with (i,i)-th entry equal to p_i, $i = 1,2$. Take $\bar{p} = (0,1) \in \mathbb{S}^2$ and note that $Y(\bar{p}) = 0$ for $b = (0,2)$. Some calculus show that the Clarke generalized covariant derivative of Y at $\bar{p}$ is given by $\partial Y(\bar{p}) = \{\mathrm{diag}(d, 2) : \ d \in [3,5]\}$. Define $X(p) := (I - pp^T)Y(p)$ the vector field on $\mathbb{S}^2$. Therefore, using (5.18), we conclude that $\partial X(\bar{p}) = \{V_{\bar{p}} := \mathrm{diag}(d - 2 + b_2, -2 + b_2) : \ d \in [3,5]\}$. *Note that all $V_{\bar{p}} \in \partial X(\bar{p})$ are nonsingular as a linear mapping $V_{\bar{p}} : T_{\bar{p}}\mathbb{S}^2 \to T_{\bar{p}}\mathbb{S}^2$, where the tangent hyperplane at $\bar{p}$ is given by* $T_{\bar{p}}\mathbb{S}^2 := \{v := (v_1, 0) \in \mathbb{R}^2 : \ v_1 \in \mathbb{R}\}$. Hence, from Definition 5.3.1, we obtain that X is regular at $\bar{p} = (0,1)$. Let $\lambda_{\bar{p}} \geq \max\{\|V_{\bar{p}}^{-1}\| : \ V_{\bar{p}} \in \partial X(\bar{p})\}$. As X is a locally Lipschitz continuous vector field, using Lemma 5.3.5 for every $\epsilon > 0$ satisfying $\epsilon\lambda_{\bar{p}} < 1$, there exists $0 < \delta < \pi$ such that X is regular on $B_\delta(\bar{p})$ and for all $p \in B_\delta(\bar{p})$ and $V_p \in \partial X(p)$ the following holds $\|{V_p}^{-1}\| \leq \lambda_{\bar{p}}/(1 - \epsilon\lambda_{\bar{p}})$. Because X is a semismooth vector field, we conclude that the condition **A1** holds.

It is worth pointing out that in the literature there exist other examples, see, for example, [55].

Chapter 6

Final remarks

In this thesis, we have proposed and studied three versions of the Newton method to solve problems in two contexts. The convergence analysis for a sequence generated by these methods was done under local and/or semi-local assumptions.

In Chapter 3, we have proposed a method for solving constrained generalized equations, which we call the Newton-InexP method. As already mentioned, we have combined the classical Newton method for solving unconstrained generalized equations with feasible inexact projections. It is worth pointing out that Lemma 3.1.5 played a key role in the proof of the main theorems of Chapters 3 and 4. In particular, under assumptions of metric regularity and strong metric regularity, Theorem 3.1.8 establishes a local convergence analysis of a sequence generated by the method. In future work, we aim to make this analysis under an weaker assumption, namely strong metric subregularity, see, for example, [31]. Furthermore, it is well-known that in practical implementations the inexact versions of the Newton method have computational advantages compared with an exact one. Therefore, following the same idea of Chapter 3 it would be interesting to study the inexact Newton method with feasible inexact projections for solving constrained generalized equations. Inexact Newton method for solving unconstrained generalized equations is formulated as follows. For the current iterate $x_k \in \mathbb{R}^n$, the next iterate x_{k+1} is computed as a point satisfying

$$\big(f(x_k) + f'(x_k)(x - x_k) + F(x)\big) \cap R_k(x_k) \neq 0, \qquad k = 0, 1, \ldots,$$

where $R_k : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a sequence of set-valued mappings with closed graphs, which represents the inexactness. For further details, see, for example, [28, 30].

In Chapter 4, we have presented a method for solving constrained smooth and nonsmooth equations, which we call the inexact Newton-InexP method. As mentioned in the introduction, this method combines the classical exact/inexact Newton method for solving nonsmooth equations with feasible inexact projections. In Theorems 4.1.3 and 4.1.5, we have shown that under mild assumptions, the exact/inexact Newton-InexP method for solving constrained smooth and nonsmooth equations preserves the local convergence properties if feasible inexact projections with suitable error (relative tolerance) are used. In particular, under the standard nonsingularity condition, the superlinear/quadratic rate is preserved. In this sense, we expect that our results become the first step towards a study of the behavior of the Newton method and its variants (including, the Gauss–Newton method, Levenberg–Marquardt method and trust region method), with feasible inexact projections, under more reasonable regularity conditions. To show the practical behavior of the proposed method, we have tested it on some medium and large-scale CAVEs. The numerical experiments have shown that the dimension of the problem and the choice of the parameter θ, which influences in the computing of η given in Theorem 4.1.3 limit the efficiency of the proposed method. With respect to robustness, the numerical results have shown that the inexact Newton-InexP method works quite well for solving this class of problems since all test problems were resolved. In future work, we aim to investigate computationally the behavior of the inexact Newton-InexP method for other class of problems, for example, the inequality feasibility problem. Computational implementations of Algorithm 3.1.6 described in Chapter 3 also is a line of future research.

Because the extension of results and methods from the Euclidean context to Riemannian setting have been a promising possibility over the years, see, for example, [35, 48, 77, 82, 88], in Chapter 5, we have studied the main properties of nonsmooth analysis for this context. Firstly, we have extended to the Riemannian setting the concept and some properties of the locally Lipschitz continuous vector fields. It is worth mentioning that the Rademacher theorem, i.e., Theorem 5.1.5 is an essential tool to ensure the existence of the Clarke generalized covariant derivative. In addition, a version of the nonsmooth Newton method for finding a singularity of these vector fields was proposed. Under the regularity and semismoothness assumptions the well-definedness and local convergence of a sequence generated by the proposed method were established. Furthermore, a semi-local convergence analysis was presented, see Theorem 5.4.1. We expect that the results of this chapter can aid in the extensions of new results and methods of nonsmooth analysis to the Riemannian context, for example, the mean value theorem as well as the inexact and globalized versions of the nonsmooth Newton method.

Bibliography

[1] M. Abbas, M. AlShahrani, Q. H. Ansari, O. S. Iyiola, and Y. Shehu. Iterative methods for solving proximal split minimization problems. *Numer. Algorithms*, 78(1):193–215, 2018.

[2] F. J. Aragón Artacho, A. Belyakov, A. L. Dontchev, and M. López. Local convergence of quasi-Newton methods under metric regularity. *Comput. Optim. Appl.*, 58(1):225–247, 2014.

[3] F. J. Aragón Artacho, A. L. Dontchev, M. Gaydu, M. H. Geoffroy, and V. M. Veliov. Metric regularity of Newton's iteration. *SIAM J. Control Optim.*, 49(2):339–362, 2011.

[4] A. Auslender and M. Teboulle. *Asymptotic cones and functions in optimization and variational inequalities*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.

[5] D. Azagra, J. Ferrera, and F. López-Mesas. Nonsmooth analysis and Hamilton–Jacobi equations on Riemannian manifolds. *J. Funct. Anal.*, 220(2):304–361, 2005.

[6] R. Behling, A. Fischer, G. Haeser, A. Ramos, and K. Schönefeld. On the constrained error bound condition and the projected Levenberg–Marquardt method. *Optimization*, 66(8):1397–1411, 2017.

[7] R. Behling, A. Fischer, M. Herrich, A. Iusem, and Y. Ye. A Levenberg–Marquardt method with approximate projections. *Comput. Optim. Appl.*, 59(1-2):5–26, 2014.

[8] S. Bellavia, M. Macconi, and B. Morini. STRSCNE: a scaled trust-region solver for constrained nonlinear equations. *Comput. Optim. Appl.*, 28(1):31–50, 2004.

[9] S. Bellavia, M. Macconi, and S. Pieraccini. Constrained dogleg methods for nonlinear systems with simple bounds. *Comput. Optim. Appl.*, 53(3):771–794, 2012.

[10] S. Bellavia and B. Morini. Subspace trust-region methods for large bound-constrained nonlinear equations. *SIAM J. Numer. Anal.*, 44(4):1535–1555, 2006.

[11] J. Y. Bello Cruz, O. P. Ferreira, and L. F. Prudente. On the global convergence of the inexact semi-smooth Newton method for absolute value equation. *Comput. Optim. Appl.*, 65(1):93–108, 2016.

[12] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific Optimization and Computation Series. Athena Scientific, Belmont, MA, second edition, 1999.

[13] E. G. Birgin, N. Krejić, and J. M. Martínez. Globally convergent inexact quasi-Newton methods for solving nonlinear systems. *Numer. Algorithms*, 32(2-4):249–260, 2003.

[14] S. Bonettini and F. Tinti. A nonmonotone semismooth inexact Newton method. *Optim. Methods Softw.*, 22(4):637–657, 2007.

[15] R. D. Canary, D. Epstein, and A. Marden. *Fundamentals of hyperbolic geometry: selected expositions*, volume 328 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.

[16] Y. Censor, A. Gibali, and S. Reich. Algorithms for the split variational inequality problem. *Numer. Algorithms*, 59(2):301–323, 2012.

[17] Y. Censor, A. Gibali, S. Reich, and S. Sabach. Common solutions to variational inequalities. *Set-Valued Var. Anal.*, 20(2):229–247, 2012.

[18] F. H. Clarke. *Optimization and nonsmooth analysis*, volume 5 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 1990.

[19] J. X. da Cruz Neto, L. L. de Lima, and P. R. Oliveira. Geodesic algorithms in Riemannian geometry. *Balkan J. Geom. Appl.*, 3(2):89–100, 1998.

[20] J. W. Daniel. Newton's method for nonlinear inequalities. *Numer. Math.*, 21:381–387, 1973.

[21] F. R. de Oliveira and O. P. Ferreira. Inexact Newton method with feasible inexact projections for solving constrained smooth and nonsmooth equations. *preprint arXiv:1810.11640v3*, 2019.

[22] F. R. de Oliveira and O. P. Ferreira. Newton method for finding a singularity of a special class of locally Lipschitz continuous vector fields on Riemannian manifolds. *preprint arXiv:1810.11636v2*, 2019.

[23] F. R. de Oliveira, O. P. Ferreira, and G. N. Silva. Newton's method with feasible inexact projections for solving constrained generalized equations. *Comput. Optim. Appl.*, 72(1):159–177, 2019.

[24] J.-P. Dedieu, P. Priouret, and G. Malajovich. Newton's method on Riemannian manifolds: convariant alpha theory. *IMA J. Numer. Anal.*, 23(3):395–419, 2003.

[25] R. S. Dembo and T. Steihaug. Truncated-Newton algorithms for large-scale unconstrained optimization. *Math. Program.*, 26(2):190–212, 1983.

[26] M. P. do Carmo. *Riemannian geometry.* Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[27] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2):201–213, 2002.

[28] A. L. Dontchev. Local analysis of a Newton-type method based on partial linearization. In *The mathematics of numerical analysis (Park City, UT, 1995)*, volume 32 of *Lectures in Appl. Math.*, pages 295–306. Amer. Math. Soc., Providence, RI, 1996.

[29] A. L. Dontchev. Uniform convergence of the Newton method for Aubin continuous maps. *Serdica Math. J.*, 22(3):283–296, 1996.

[30] A. L. Dontchev and R. T. Rockafellar. Convergence of inexact Newton methods for generalized equations. *Math. Program.*, 139(1-2, Ser. B):115–137, 2013.

[31] A. L. Dontchev and R. T. Rockafellar. *Implicit functions and solution mappings.* Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2014. A view from variational analysis.

[32] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[33] F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. *Math. Program.*, 146(1-2, Ser. A):1–36, 2014.

[34] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems. Vol. II.* Springer Series in Operations Research. Springer-Verlag, New York, 2003.

[35] T. A. Fernandes, O. P. Ferreira, and J. Yuan. On the Superlinear Convergence of Newton's Method on Riemannian Manifolds. *J. Optim. Theory Appl.*, 173(3):828–843, 2017.

[36] O. P. Ferreira. A robust semi-local convergence analysis of Newton's method for cone inclusion problems in Banach spaces under affine invariant majorant condition. *J. Comput. Appl. Math.*, 279:318–335, 2015.

[37] O. P. Ferreira, A. N. Iusem, and S. Z. Németh. Projections onto convex sets on the sphere. *J. Optim. Theory Appl.*, 57(3):663–676, 2013.

[38] O. P. Ferreira, A. N. Iusem, and S. Z. Németh. Concepts and techniques of optimization on the sphere. *TOP*, 22(3):1148–1170, 2014.

[39] O. P. Ferreira and G. N. Silva. Kantorovich's theorem on Newton's method for solving strongly regular generalized equation. *SIAM J. Optim.*, 27(2):910–926, 2017.

[40] O. P. Ferreira and G. N. Silva. Local convergence analysis of Newton's method for solving strongly regular generalized equations. *J. Math. Anal. Appl.*, 458(1):481–496, 2018.

[41] O. P. Ferreira and B. F. Svaiter. Kantorovich's theorem on Newton's method in Riemannian manifolds. *J. Complexity*, 18(1):304–329, 2002.

[42] A. Fischer, M. Herrich, A. F. Izmailov, W. Scheck, and M. V. Solodov. A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points. *Comput. Optim. Appl.*, 69(2):325–349, 2018.

[43] A. Fischer, M. Herrich, A. F. Izmailov, and M. V. Solodov. Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. *Comput. Optim. Appl.*, 63(2):425–459, 2016.

[44] C. A. Floudas et al. *Handbook of Test Problems in Local and Global Optimization, in: Nonconvex Optimization and Its Applications, Volume 33, Kluwer Academic, Dordrecht, 1999.*

[45] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval. Res. Logist.*, 3(1-2):95–110, 1956.

[46] M. Fukushima, Z.-Q. Luo, and P. Tseng. Smoothing functions for second-order-cone complementarity problems. *SIAM J. Optim.*, 12(2):436–460 (electronic), 2001/02.

[47] D. Gabay. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, 37(2):177–219, 1982.

[48] E. Ghahraei, S. Hosseini, and M. R. Pouryayevali. Pseudo-Jacobian and characterization of monotone vector fields on Riemannian manifolds. *J. Convex Anal.*, 24(1):149–168, 2017.

[49] M. L. N. Gonçalves and F. R. Oliveira. An inexact Newton-like conditional gradient method for constrained nonlinear systems. *Appl. Num. Math.*, 132:22–34, 2018.

[50] M. L. N. Gonçalves and J. G. Melo. A Newton conditional gradient method for constrained nonlinear systems. *J. Comput. Appl. Math.*, 311:473–483, 2017.

[51] N. I. M. Gould and P. L. Toint. Numerical methods for large-scale non-convex quadratic programming. In *Trends in industrial and applied mathematics (Amritsar, 2001)*, volume 72 of *Appl. Optim.*, pages 149–179. Kluwer Acad. Publ., Dordrecht, 2002.

[52] H. He, C. Ling, and H.-K. Xu. A relaxed projection method for split variational inequalities. *J. Optim. Theory Appl.*, 166(1):213–233, 2015.

[53] M. Herrich. *Local convergence of Newton-type methods for nonsmooth constrained equations and applications.* PhD thesis, Technical University of Dresden, Germany, 2014.

[54] S. Hosseini and M. R. Pouryayevali. Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds. *Nonlinear Anal.*, 74(12):3884–3895, 2011.

[55] S. Hosseini and A. Uschmajew. A Riemannian Gradient Sampling Algorithm for Nonsmooth Optimization on Manifolds. *SIAM J. Optim.*, 27(1):173–189, 2017.

[56] A. F. Izmailov and M. V. Solodov. Inexact Josephy-Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization. *Comput. Optim. Appl.*, 46(2):347–368, 2010.

[57] A. F. Izmailov and M. V. Solodov. *Newton-type methods for optimization and variational problems*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.

[58] N. H. Josephy. *Newton's method for generalized equations and the PIES energy model.* PhD thesis, Department of Industrial Engineering, University of Wisconsin–Madison, 1979.

[59] C. Kanzow and A. Klug. An interior-point affine-scaling trust-region method for semismooth equations with box constraints. *Comput. Optim. Appl.*, 37(3):329–353, 2007.

[60] W. La Cruz. A projected derivative-free algorithm for nonlinear equations with convex constraints. *Optim. Methods Softw.*, 29(1):24–41, 2014.

[61] G. Lan and Y. Zhou. Conditional gradient sliding for convex optimization. *SIAM J. Optim.*, 26(2):1379–1409, 2016.

[62] S. Lang. *Differential and Riemannian manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 1995.

[63] J. M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[64] C. Li and J. Wang. Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. *IMA J. Numer. Anal.*, 26(2):228–251, 2006.

[65] O. L. Mangasarian. A generalized Newton method for absolute value equations. *Optim. Lett.*, 3(1):101–108, 2009.

[66] L. Marini, B. Morini, and M. Porcelli. Quasi-Newton methods for constrained nonlinear systems: complexity analysis and applications. *Comput. Optim. Appl.*, 71(1):147–170, 2018.

[67] J. M. Martínez and L. Qi. Inexact Newton methods for solving nonsmooth equations. *J. Comput. Appl. Math.*, 60(1):127–145, 1995.

[68] R. Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.*, 15(6):959–972, 1977.

[69] R. D. C. Monteiro and J.-S. Pang. A potential reduction Newton method for constrained equations. *SIAM J. Optim.*, 9(3):729–754, 1999.

[70] B. Morini, M. Porcelli, and P. L. Toint. Approximate norm descent methods for constrained nonlinear systems. *Math. Comp.*, 87(311):1327–1351, 2018.

[71] A. Moudafi. Split monotone variational inclusions. *J. Optim. Theory Appl.*, 150(2):275–283, 2011.

[72] J. Nocedal and S. Wright. *Numerical optimization.* Springer series in operations research. Springer, second edition, 2000.

[73] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Academic Press, New York-London, 1970.

[74] C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982.

[75] J.-S. Pang and L. Qi. Nonsmooth equations: motivation and algorithms. *SIAM J. Optim.*, 3(3):443–465, 1993.

[76] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Math. Program.*, 58(3, Ser. A):353–367, 1993.

[77] F. Rampazzo and H. J. Sussmann. Commutators of flow maps of nonsmooth vector fields. *J. Differential Equations*, 232(1):134–175, 2007.

[78] T. Sakai. *Riemannian geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1996. Translated from the 1992 Japanese original by the author.

[79] M. Shub. Some remarks on dynamical systems and numerical analysis. In *Dynamical systems and partial differential equations (Caracas, 1984)*, pages 69–91. Univ. Simon Bolivar, Caracas, 1986.

[80] M. J. Śmietański. Some superlinearly convergent inexact generalized Newton method for solving nonsmooth equations. *Optim. Methods Softw.*, 27(3):405–417, 2012.

[81] M. J. Śmietański. A perturbed version of an inexact generalized Newton method for solving nonsmooth equations. *Numer. Algorithms*, 63(1):89–106, 2013.

[82] S. T. Smith. Optimization techniques on Riemannian manifolds. In *Hamiltonian and gradient flows, algorithms and control*, volume 3 of *Fields Inst. Commun.*, pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[83] M. Spivak. *A comprehensive introduction to differential geometry. Vol. II.* Publish or Perish, Inc., Wilmington, Del., third edition, 1999.

[84] L. W. Tu. *An introduction to manifolds*. Universitext. Springer-Verlag New York, second edition, 2011.

[85] C. Udrişte. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.

[86] L. U. Uko. Generalized equations and the generalized Newton method. *Math. Program.*, 73(3, Ser. A):251–268, 1996.

[87] R. J. Vanderbei. *Linear programming: foundations and extensions*, volume 4 of *International Series in Operations Research & Management Science*. Kluwer Academic Publishers, Boston, MA, 1996.

[88] J. H. Wang. Convergence of Newton's method for sections on Riemannian manifolds. *J. Optim. Theory Appl.*, 148(1):125–145, 2011.