INEXACT METHODS FOR CONSTRAINED OPTIMIZATION PROBLEMS AND FOR CONSTRAINED MONOTONE NONLINEAR EQUATIONS

Doctoral Thesis by

TIAGO DA COSTA MENEZES

Supervised by

Prof. Dr. **Max Leandro Nobre Gonçalves**

Funded by
CAPES

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

MAIO 2021

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

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

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

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

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

1. Identificação do material bibliográfico

[ ] Dissertação [X] Tese

2. Nome completo do autor

Tiago da Costa Menezes

3. Título do trabalho

INEXACT METHODS FOR CONSTRAINED OPTIMIZATION PROBLEMS AND FOR CONSTRAINED MONOTONE NONLINEAR EQUATIONS

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

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

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

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

Documento assinado eletronicamente por **Max Leandro Nobre Gonçalves**, **Professor do Magistério Superior**, em 24/05/2021, às 10:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **TIAGO DA COSTA MENEZES**, **Discente**, em 24/05/2021, às 23:24, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2086500** e o código CRC **45F06376**.

Referência: Processo nº 23070.019746/2021-00 SEI nº 2086500

Tiago da Costa Menezes

INEXACT METHODS FOR CONSTRAINED OPTIMIZATION PROBLEMS AND FOR CONSTRAINED MONOTONE NONLINEAR EQUATIONS

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 Doutor em Matemática.
Área de concentração: Otimização
Orientador: Prof. Dr. Max Leandro Nobre Gonçalves

Goiânia
2021

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

Menezes, Tiago da Costa
INEXACT METHODS FOR CONSTRAINED OPTIMIZATION PROBLEMS AND FOR CONSTRAINED MONOTONE NONLINEAR EQUATIONS [manuscrito] / Tiago da Costa Menezes. - 2021.
xi, 72 f.

Orientador: Prof. Dr. Max Leandro Nobre Gonçalves.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2021.
Bibliografia.
Inclui siglas, abreviaturas, símbolos, gráfico, tabelas, algoritmos.

1. Convex-constrained optimization problem. 2. Nonlinear equations. 3. Approximate projections. 4. Inexact variable metric method. 5. Gauss-Newton method. I. Gonçalves, Max Leandro Nobre, orient. II. Título.

CDU 51

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE TESE

Ata nº 05 da sessão de Defesa de Tese de **Tiago da Costa Menezes**, que confere o título de Doutor em Matemática, **na área de Otimização.**

Ao vigésimo dia do mês de maio do ano de dois mil e vinte um, a partir das quatorze horas, através de web-vídeo-conferência, realizou-se a sessão pública de Defesa de Tese intitulada **"INEXACT METHODS FOR CONSTRAINED OPTIMIZATION PROBLEMS AND FOR CONSTRAINED MONOTONE NONLINEAR EQUATIONS"**. Os trabalhos foram instalados pelo presidente da banca, Professor Doutor Max Leandro Nobre Gonçalves IME/UFG com a participação dos demais membros da Banca Examinadora: Professor Orizon Pereira Ferreira - IME/UFG membro titular interno, Professor Paulo Sergio Marques dos Santos DMAT/UFDPar membro titular externo, Professor Doutor Douglas Soares Gonçalves - DMAT/UFSC membro titular externo e Professora Doutora Sandra Augusta Santos - DMA/UNICAMP membro titular externo. Durante a arguição os membros da banca **não fizeram sugestão de alteração do título do trabalho**. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese, tendo sido o candidato aprovado pelos seus membros. Proclamados os resultados pelo Professor Doutor Max Leandro Nobre Gonçalves IME/UFG, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, Ao vigésimo dia do mês de maio do ano de dois mil e vinte um.

TÍTULO SUGERIDO PELA BANCA

INEXACT METHODS FOR CONSTRAINED OPTIMIZATION PROBLEMS AND FOR CONSTRAINED MONOTONE NONLINEAR EQUATIONS

Documento assinado eletronicamente por **Max Leandro Nobre Gonçalves**, **Professor do Magistério Superior**, em 20/05/2021, às 15:56, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Orizon Pereira Ferreira**, **Professora do Magistério Superior**, em 20/05/2021, às 16:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Sandra Augusta Santos**, **Usuário Externo**, em 20/05/2021, às 16:08, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Paulo Sérgio Marques dos Santos**, **Usuário Externo**, em 21/05/2021, às 10:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Douglas Soares Gonçalves**, **Usuário Externo**, em 22/05/2021, às 10:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2013727** e o código CRC **1CE470C8**.

Referência: Processo nº 23070.019746/2021-00 SEI nº 2013727

Dedicado a:

Meus pais, José Filho e Maria Bernadete

Meu irmão, Felipe

Agradecimentos

Primeiramente, agradeço a Deus por estar presente em todos os momentos e me dado forças nos momentos de dúvida.

Ao meu orientador Professor Max Leandro Nobre Gonçalves por ter aceitado me orientar, pelos ensinamentos, pela paciência e dedicação durante o desenvolvimento de todo o trabalho.

Aos professores do IME/UFG que contribuíram na minha formação durante o doutorado. Também agradeço a todos os professores do grupo de otimização pelo convívio e sugestões nos seminários.

Aos Professores Orizon Pereira Ferreira, Paulo Sergio Marques dos Santos, Douglas Soares Gonçalves e a Professora Sandra Augusta Santos, por participarem da banca examinadora, pelas valiosas correções e sugestões.

Aos amigos que conviveram comigo durante os anos de doutorado, pelos momentos de descontração e estudos, e também a todos os meus amigos de Goiânia pela excelente hospitalidade.

A todos meus familiares e amigos pelo apoio e torcida, em especial aos meus pais e irmão pelo incentivo e amor incondicional.

À CAPES, agradeço pelo apoio financeiro.

Abstract

In this work, we propose and analyze some methods to solve constrained optimization problems and constrained monotone nonlinear systems of equations. Our first algorithm is an inexact variable metric method for solving convex-constrained optimization problems. At each iteration of the method, the search direction is obtained by inexactly minimizing a strictly convex quadratic function over the closed convex feasible set. Here, we propose a new inexactness criterion for the search direction subproblems. Under mild assumptions, we prove that any accumulation point of the sequence generated by the method is a stationary point of the problem under consideration. Our second method consists of a Gauss-Newton algorithm with approximate projections for solving constrained nonlinear least squares problems. The local convergence of the method including results on its rate is discussed by using a general majorant condition. By combining the latter method and a nonmonotone line search strategy, we also propose a global version of this algorithm and analyze its convergence results. Our third approach corresponds to a framework, which is obtained by combining a safeguard strategy on the search directions with a notion of approximate projections, to solve constrained monotone nonlinear systems of equations. The global convergence of our framework is obtained under appropriate assumptions and some examples of methods which fall into this framework are presented. Numerical experiments illustrating the practical behaviors of the methods are reported and comparisons with existing algorithms are also presented.

Keywords: Convex-constrained optimization problem; Nonlinear equations; Approximate projections; Inexact variable metric method; Gauss-Newton method; Local and global convergence.

Resumo

Neste trabalho, propomos e analisamos alguns métodos para resolver problemas de otimização com restrições e sistemas de equações não lineares monótonas com restrições. Nosso primeiro algoritmo é um método inexato de métrica variável para resolver problemas de otimização com restrições convexas. A cada iteração deste método, a busca direcional é obtida minimizando inexatamente uma função quadrática estritamente convexa sobre o conjunto convexo fechado viável. Aqui, propusemos um novo critério de inexatidão para os subproblemas de busca direcional. Sob suposições apropriadas, provamos que qualquer ponto de acumulação da sequência gerada pelo novo método é um ponto estacionário do problema sob consideração. Nosso segundo método consiste em um método Gauss-Newton com projeções aproximadas para resolver problemas de quadrados mínimos não lineares com restrições. A convergência local do método, incluindo resultados sobre sua taxa de convergência, é discutida usando uma condição majorante geral. Ao combinar o último método e uma estratégia de busca linear não monótona, também propusemos uma versão global deste algoritmo e analisamos seus resultados de convergência. Nossa terceira abordagem corresponde a um "framework", o qual é obtido combinando uma estratégia de salvaguarda na busca direcional com uma noção de projeções aproximadas, para resolver sistemas de equações não lineares monótonas com restrições. A convergência global de nosso "framework" é obtida sob suposições apropriadas e alguns exemplos de métodos que se enquadram nesta estrutura são apresentados. Experimentos numéricos são relatados para ilustrar os desempenhos dos métodos e comparações com algoritmos existentes também são apresentadas.

Palavras-chave: Problema de otimização com restrição convexa; Equações não lineares; Projeções aproximadas; Método inexato de métrica variável; Método de Gauss-Newton; Convergência local e global.

Contents

Chapter 1

Introduction

Let us first consider the following convex-constrained optimization problem

$$\min_{x \in C} \; f(x), \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function and $C \subseteq \mathbb{R}^n$ is a nonempty convex closed set. This is a classical problem in continuous optimization and different methods have been proposed in the literature for solving it; see, for example, [11,12,16,64,68]. A well-known one is the projected gradient method, which can be seen as the constrained extension of the gradient method (also known as steepest descent) for unconstrained optimization problem. The projected gradient method is quite simple to implement; however, it may be very slow in some applications. In order to overcome this drawback, work [12] proposed the spectral projected gradient (SPG) method, which has been shown an efficient approach for solving (1.1) mainly in large-scale, owing to its low memory requirements. Given an arbitrary initial point $x_0 \in C$, the SPG method generates a sequence of iterates by the rule

$$x_{k+1} = x_k + \alpha_k d_k, \quad k \geq 0, \tag{1.2}$$

where the step-size α_k is obtained by the nonmonotone line search strategies proposed in [44] and the search direction d_k is defined as $d_k = P_C(x_k - (1/\lambda_k)\nabla f(x_k)) - x_k$, where P_C denotes the orthogonal projection on C and λ_k is the Barzilai-Borwein scaling [8] defined by

$$\lambda_0 \in [\lambda_{min}, \lambda_{max}], \qquad \lambda_k = \min\left\{\lambda_{max}, \max\left\{\lambda_{min}, a_k/b_k\right\}\right\}, \tag{1.3}$$

with $0 < \lambda_{min} < \lambda_{max}$, $b_k := \langle x_k - x_{k-1}, x_k - x_{k-1}\rangle$ and $a_k := \langle x_k - x_{k-1}, \nabla f(x_k) - \nabla f(x_{k-1})\rangle$. The convergence results and/or numerical experiments illustrating the practical behavior of the SPG method were discussed in [12] and in many subsequent works including [6, 13, 15, 17, 36, 46, 54, 60, 76, 81, 82].

It is well-known that depending on the geometry of C, the orthogonal projection onto it neither has a closed-form nor can be easily computed. For this reason, [14] (see also [4]) proposed an inexact version of the SPG method in which approximate projections are allowed. Indeed, a more general approach, called Inexact Variable Metric method (IVM), was proposed. It differs from the SPG method by the fact that the search direction d_k in (1.2) is computed such that $x_k + d_k \in C$ and

$$Q_k(d_k) \leq \eta Q_k(\bar{d}_k), \tag{1.4}$$

where $\eta \in (0, 1]$,

$$\bar{d}_k := \operatorname{argmin}_{x_k+d\in C} Q_k(d) := \frac{1}{2}\langle d, B_k d\rangle + \langle \nabla f(x_k), d\rangle, \tag{1.5}$$

and $B_k \in \mathbb{R}^{n\times n}$ is a suitable symmetric positive definite matrix. If $B_k := \lambda_k I$ for every $k \geq 0$, where λ_k is as in (1.3), the inexact variable metric method corresponds to an inexact version of the SPG method. It is not hard to verify that $\bar{d}_k$ in (1.5) is equivalent to $\bar{d}_k = \bar{y}_k - x_k$, with

$$\bar{y}_k = \operatorname{argmin}_{y\in C}\ q_k(y) := \frac{1}{2}\langle B_k y, y\rangle + \langle \nabla f(x_k) - B_k x_k, y\rangle. \tag{1.6}$$

In its turn, $\bar{y}_k$ in (1.6) is equivalent to

$$\bar{y}_k := \operatorname{argmin}_{y\in C} \frac{1}{2}\|y - (x_k - B_k^{-1}\nabla f(x_k))\|^2_{B_k}, \tag{1.7}$$

where $\|\cdot\|^2_{B_k} := \langle B_k\cdot, \cdot\rangle$. Therefore, d_k in (1.4) can also be interpreted as an approximation of the search direction $\bar{d}_k$ of the projected (in the norm $\|\cdot\|_{B_k}$) quasi-Newton method.

At first sight, a drawback of the inexact criterion in (1.4) is that it requires the optimal value of the problem in (1.5). It was presented in [4, 14] some applications in which it is possible to establish a sequence of lower bounds $C_l \leq\ Q_k(\bar{d}_k)$ that converges to the value $Q_k(\bar{d}_k)$ as l goes to infinity. Hence, criterion (1.4) is satisfied when the verifiable condition $Q_k(d_k) \leq \eta C_l$ holds. It is not clear, however, how the strategies in [4,14] can be employed or even how an inexact direction satisfying (1.4) can be obtained for other complex feasible sets (where the projection cannot be easily performed). Therefore, the first goal of this thesis is to present an inexact variable metric method with a different inexactness criterion for the subproblems (1.6). We present a concept of approximate solution for (1.6), which does not require the knowledge of its optimal value. The new criterion can be verified by finding the infimum of a linear function over the feasible set C. Such verification comes for free when the conditional gradient method (Frank-Wolfe) [27, 32] is used to solve the problem in (1.6). Under mild assumptions, we prove that any accumulation point of the sequence generated by the proposed method is a stationary point of (1.1). In order to illustrate the practical advantages of the new approach for inexact variable metric method, we report some

numerical experiments. In particular, we present an application where our concept of inexact solutions is quite appealing; more details about this application are given in Subsection 3.2.2.

Our second problem of interest is a particular case of (1.1), which corresponds to the convex-constrained nonlinear least squares problem

$$\min_{x\in C} \; f(x) := \frac{1}{2}\|F(x)\|^2, \tag{1.8}$$

where $\mathbb{U} \subseteq \mathbb{R}^n$ is an open set containing the nonempty convex closed set C and $F : \mathbb{U} \to \mathbb{R}^m$ is a continuously differentiable nonlinear function. This problem appears in many important applications (see, e.g., [3, 5, 10, 67]). It is worth pointing out that different algorithms have been proposed and studied in the literature for solving (1.8). Strategies based on sequential quadratic programming, quasi-Newton and trust-region methods have been used; see, for instance, [53, 57, 68]. Among the various approaches, one of the most popular is the Gauss-Newton method and its variations, capable of obtaining efficient computational results by exploring the structure of the function f (see [7, 9, 23, 34, 70]).

The second goal of this thesis is propose and analyze a Gauss-Newton methods with approximate projections for solving (1.8). The method to be proposed here basically consists of computing an approximate projection of the unconstrained Gauss-Newton step. The approximate projection is based on the inexactness criterion for the subproblems (1.6) with respect to the metric defined by $B_k = F'(x_k)^T F'(x_k)$, where A^T denotes the transposed matrix of A. From the theoretical viewpoint, we provide an estimate of the convergence radius, for which well-definedness and convergence of the method are ensured. Furthermore, results on its convergence rates are also established. Our analysis is done by using a majorant condition, which allows us to study convergence results of Newton and Gauss-Newton methods in a unified way; see, for example, [29, 30, 41]. Thus, our local analysis covers two large families of nonlinear functions, namely, one satisfying a Lipschitz condition and another one satisfying a Smale condition, which includes a substantial class of analytic functions. However, as it is well-known, globalization strategies produce, in general, more robust methods. Therefore, we also propose a global version of our local method. As in our first global algorithm, the globalization technique is based on the efficient nonmonotone line search in [44]. It is worth pointing out that the nonmonotone strategy has been shown to be more efficient due to the fact that enforcing monotonicity of the function values may make the method converge slower. Under suitable assumptions, this global version can be seen as an instance of our first method. We also report some numerical experiments for the algorithm on a set of box- and polyhedral-constrained nonlinear systems and compare their performances with the proximal Gauss-Newton method in [70], which, applied to (1.8), corresponds to our local method with exact projections. In the box-constrained case, we also compare performance of our global version with the inexact Gauss-Newton trust-region method in [68].

Finally, our third problem corresponds to the convex-constrained monotone nonlinear system of equations: finding $x_* \in C$ such that

$$F(x_*) = 0, \tag{1.9}$$

where C is a nonempty closed convex set and $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous and monotone nonlinear function, not necessarily differentiable. The monotonicity of $F : \mathbb{R}^n \to \mathbb{R}^n$ here means $\langle F(x) - F(y), x - y \rangle \geq 0$, for all $x, y \in \mathbb{R}^n$. Problems of this nature have many applications such as power engineering, chemical equilibrium systems and economic equilibrium problems, see e.g., [25, 28, 61, 79]. Recently, how to solve the constrained problem (1.9) has become an important subject of research. Due to the efficiency and low computational costs for large values of n, different attractive methods have been proposed in the literature. Many of them are extensions of Newton-type, spectral gradient and conjugate gradient methods for solving the unconstrained monotone nonlinear system; see, e.g., [55, 56, 66, 77, 78, 81, 84].

The third goal of this thesis is to present a framework with approximate projections for solving (1.9). More precisely, at each iteration, the framework imposes a safeguard strategy on the search directions. A suitable line search procedure is considered based on [73], which, in particular, provides a hyperplane that strictly separates the current iteration from zeroes of the system of equations. Then, we compute an approximate projection of a point, which belongs to the aforementioned hyperplane, onto the intersection between C and the hyperplane (or onto the constrained set C). Under mild assumptions, we prove that the sequence generated by the proposed framework converges to a solution of (1.9). Some examples of methods which fall into this framework are reported. Essentially the examples are inexact versions of methods based on spectral gradient and quasi-Newton methods for convex-constrained monotone nonlinear equations; see, e.g., [1, 52, 78, 81, 83]. In order to illustrate the robustness and effectiveness of the instances of the framework, we report some preliminary numerical experiments on a set of problems in the form (1.9). Moreover, we also applied the framework for solving the constrained absolute value equation and compare its performance with the inexact Newton method with feasible inexact projections [65].

This thesis is organized as follows. In Chapter 2, we first establish some notations and basic results. A concept of approximate solution to the problem similar to (1.6) and some of its properties are discussed. In Chapter 3, we describe a modified inexact variable metric method and present its global convergence theorem. Moreover, some numerical experiments of the proposed method are presented. In Chapter 4, we propose the Gauss-Newton method with approximate projections (GNM-AP) and present its main local convergence theorem. We also present two applications of the main theorem and establish a global version of the GNM-AP. To illustrate its performance, some numerical experiments are reported. In Chapter 5, a framework with approximate projections for solving monotone nonlinear equations and its

global convergence are discussed. We also present some instances of the latter framework by means of some examples of search directions d_k that satisfy the safeguard conditions. Some preliminary numerical experiments are reported to illustrate its performance. Finally, we conclude this thesis with some remarks in Chapter 6.

We mention that the material of this thesis originated three papers, two of them [39, 40] are published and one is in the final stage of preparation.

Chapter 2

Preliminaries

In this chapter, we introduce some definitions, notations and basic results used throughout this thesis. In particular, we discuss our concept of approximate solution of a quadratic problem and establish some properties, which will be fundamental in the course of this work.

2.1 Notations and basic definitions

The open ball in $\mathbb{R}^n$ with center a and radius r is denoted by $\mathcal{B}(a, r)$. Denote $D^+f(0)$ as the right-hand derivative of a convex function $f : [0, \infty) \to \mathbb{R}$. Let $\mathbb{B}$ be the set of $n \times n$ symmetric positive definite matrices such that

$$\|B\| \leq L \qquad \text{and} \qquad \|B^{-1}\| \leq L, \tag{2.1}$$

where $L > 1$ and $\|.\|$ is a sub-multiplicative matrix norm. Note that $\mathbb{B}$ is a compact set of $\mathbb{R}^{n\times n}$. Consider also the inner product on $\mathbb{R}^n$ defined by $\langle x, z\rangle_B = \langle x, Bz\rangle$, where $B \in \mathbb{B}$ and $\langle ., .\rangle$ denotes the usual inner product. Notice that the corresponding induced norm $\|.\|_B$ is equivalent to the Euclidean norm on $\mathbb{R}^n$, since the following inequalities hold

$$\frac{1}{\|B^{-1}\|}\|x\|^2 \leq \|x\|_B^2 \leq \|B\|\|x\|^2. \tag{2.2}$$

Let be $A \in \mathbb{R}^{m\times n}$ with rank $r \leq \min\{m, n\}$. The Moore-Penrose inverse of A is a matrix $A^\dagger \in \mathbb{R}^{n\times m}$ which satisfies:

$$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^T = AA^\dagger, \quad (A^\dagger A)^T = A^\dagger A.$$

Note that, if $\text{rank}(A) = n$ or A^TA is invertible in $\mathbb{R}^{n\times n}$, then

$$A^\dagger = (A^TA)^{-1}A^T, \qquad A^\dagger A = I = AA^\dagger, \qquad \|A^\dagger\|^2 = \|(A^TA)^{-1}\|. \tag{2.3}$$

2.2 Approximate solutions of a quadratic problem

In this section, we will introduce our concept of inexact solutions of a subproblems of the form in (1.6) and establish some of its useful properties. For a suitable choice of inputs, such subproblems can be interpreted as approximate projections.

Definition 2.2.1 *Given $B \in \mathbb{B}$, $w \in \mathbb{R}^n$, $\varepsilon \geq 0$ and a nonempty closed convex set $C \subset \mathbb{R}^n$, we say that $\tilde{y}_C^B(w)$ is an ε–approximate solution for the problem*

$$\min_{y \in C} \ \frac{1}{2}\langle By, y\rangle - \langle w, y\rangle, \tag{2.4}$$

iff

$$\tilde{y}_C^B(w) \in C \qquad \text{and} \qquad \langle B\tilde{y}_C^B(w) - w, y - \tilde{y}_C^B(w)\rangle \geq -\varepsilon, \quad \forall\, y \in C. \tag{2.5}$$

Remark 2.2.2 Since in (2.4) we are minimizing a strictly convex quadratic function over a convex set, condition (2.5) is a natural condition for an approximate solution. Indeed, the optimality condition for (2.4) is

$$\langle B\bar{y} - w, y - \bar{y}\rangle \geq 0, \quad \forall\, y \in C.$$

Hence, one could define an approximate solution as $\tilde{y} \in C$ such that $\langle B\tilde{y} - w, y - \tilde{y}\rangle \geq -\varepsilon$, for all $y \in C$, which coincides with (2.5). Note that, if $\tilde{y}_C^B(w)$ is a *zero*–approximate solution, then $\tilde{y}_C^B(w)$ is the unique exact solution of (2.4), which we will denote by $y_C^B(w)$.

Note that, if $w := Bx$ with $x \in \mathbb{R}^n$, then problem (2.4) can be rewritten, ignoring constant terms, as

$$\min_{y \in C} \frac{1}{2}\|y - x\|_B^2. \tag{2.6}$$

and (2.5) is equivalent to

$$\langle x - \tilde{y}_C^B(Bx), y - \tilde{y}_C^B(Bx)\rangle_B \leq \varepsilon, \quad \forall\, y \in C. \tag{2.7}$$

In this case, we can say that $\tilde{y}_C^B(Bx)$ is an approximate projection (in the norm $\|\cdot\|_B$) of x onto C. It is easy to prove that the exact projection $y_C^B(\cdot)$ is nonexpansive in the norm $\|\cdot\|_B$, i.e.

$$\|y_C^B(Bx) - y_C^B(B\hat{x})\|_B \leq \|x - \hat{x}\|_B, \quad x, \hat{x} \in \mathbb{R}^n. \tag{2.8}$$

Moreover, for every $B \in \mathbb{B}$, $w := Bx \in \mathbb{R}^n$ and $\varepsilon \geq 0$, the following relationship between y_C^B and $\tilde{y}_C^B$ holds

$$\|\tilde{y}_C^B(Bx) - y_C^B(Bx)\|_B \leq \sqrt{\varepsilon}. \tag{2.9}$$

Indeed, since $y_C^B(w) \in C$ and $\tilde{y}_C^B(w) \in C$, it follows from Definition 2.2.1 that

$$\langle B\tilde{y}_C^B(Bx) - Bx, \tilde{y}_C^B(Bx) - y_C^B(Bx)\rangle \leq \varepsilon, \qquad \langle Bx - By_C^B(Bx), \tilde{y}_C^B(Bx) - y_C^B(Bx)\rangle \leq 0.$$

By adding the last two inequalities, we obtain

$$\|\tilde{y}_C^B(Bx) - y_C^B(Bx)\|_B \leq \sqrt{\varepsilon}. \tag{2.10}$$

In Figure 2.1, an admissible approximation of $y_C^B(Bx)$ with $B \equiv I$ is depicted.

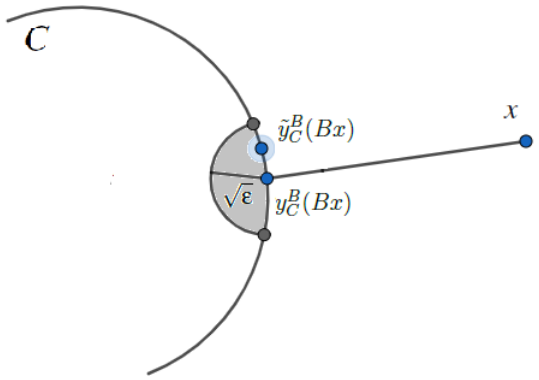

Figure 2.1: ε–approximate projection

We emphasize that criterion (2.5) can be easily checked when, for example, C is bounded and the conditional gradient method [27] is used to solve (2.4). The conditional gradient (CondG) method, also known as Frank-Wolfe method [32], is designed to solve the convex optimization problem $\min_{x\in C} h(x)$, where C is a nonempty compact convex set and h is a differentiable convex function. Given $z_{j-1} \in C$, its j-th step first finds $\bar{z}_j$ as a minimum of the linear function $\langle \nabla h(z_{j-1}), \cdot\rangle$ over C and then set $z_j = (1-\alpha_j)z_{j-1} + \alpha_j \bar{z}_j$ for some $\alpha_j \in [0,1]$. Its major distinguishing feature compared to other first-order algorithms such as the projected gradient (or accelerated gradient) method is that it replaces the usual projection onto C by a linear oracle which computes $\bar{z}_j$ as above. Since, for some relevant cases of C (for example, when C is the spectrahedron; see Subsection 3.2.2), the latter operation is considerably cheaper than the first one, the CondG method is competitive with first-order projection methods and it has recently re-gained attention in different application areas (see, e.g., [33,50]). If we apply the CondG method to (2.4), then $\bar{z}_j$ is a solution of the subproblem

$$\begin{array}{ll} \min & \langle Bz_{j-1} - w, z - z_{j-1}\rangle, \\ \text{s.t.} & z \in C \end{array} \tag{2.11}$$

and, hence, if the CondG iterations are stopped when

$$\langle Bz_{j-1} - w, \bar{z}_j - z_{j-1}\rangle \geq -\varepsilon, \tag{2.12}$$

then condition (2.5) holds with $\tilde{y}_C^B(w) = z_{j-1}$.

We next discuss a way to use the CondG method to obtain an approximate solution for (2.4) when the diameter of C is very large or even when C is unbounded. Note that the exact solution $y_C^B(w)$ of (2.4), with $w := Bx - \nabla f(x)$ and $x \in C$, satisfies

$$\langle B(x - y_C^B(w)) - \nabla f(x), x - y_C^B(w)\rangle \leq 0,$$

which, combined with the Cauchy-Schwarz inequality yields

$$\|x - y_C^B(w)\|_B^2 \leq \langle \nabla f(x), x - y_C^B(w)\rangle \leq \|\nabla f(x)\|\|x - y_C^B(w)\|.$$

It follows from the last inequality and (2.2) that

$$\|x - y_C^B(w)\| \leq \|B^{-1}\|\|\nabla f(x)\|,$$

which implies that the ball $\mathcal{B}\left(x, \|B^{-1}\|\|\nabla f(x)\|\right)$ contains the (unknown) exact solution $y_C^B(w)$ of (2.4). Therefore, one can apply the conditional gradient method to (2.4) with C replaced by $C \cap \mathcal{B}\left(x, \|B^{-1}\|\|\nabla f(x)\|\right)$ in order to obtain a point $\tilde{y}_C^B(w)$ satisfying

$$\tilde{y}_C^B(w) \in C, \quad \langle B(x - \tilde{y}_C^B(w)) - \nabla f(x), y - \tilde{y}_C^B(w)\rangle \leq \varepsilon, \quad \forall\, y \in C \cap \mathcal{B}\left(x, \|B^{-1}\|\|\nabla f(x)\|\right). \tag{2.13}$$

It can be proven, using that the quadratic function in (2.4) is strongly convex and $y_C^B(w) \in C \cap \mathcal{B}\left(x, \|B^{-1}\|\|\nabla f(x)\|\right)$, that if $\varepsilon = 0$ in the last inequality, then $\tilde{y}_C^B(w) = y_C^B(w)$. Therefore, we claim that the results of the algorithms proposed in this work can also be shown if (2.5) is replaced by (2.13).

We also mention that other iterative methods can take place to obtain an ε–approximate solution for (2.4), being enough to solve periodically, or at each iteration j, the linear subproblem (2.11) to test our criterion: unboundness of the linear subproblem implies that the criterion does not hold.

We next establish some useful relationships between exact and inexact solutions of (2.4) when B varies.

Lemma 2.2.3 *Let $B, D \in \mathbb{B}$ and $x \in \mathbb{R}^n$. Then,*

$$\|y_C^B(Bx) - y_C^D(Dx)\|_B \leq \|B^{-1}\|^{1/2}\|(B - D)(y_C^D(Dx) - x)\|, \quad \forall\, x \in \mathbb{R}^n.$$

Proof. Denote $z = y_C^B(Bx)$ and $\hat{z} = y_C^D(Dx)$. Hence, it follows from Definition 2.2.1 that

$$\langle B(z - x), \hat{z} - z\rangle \geq 0, \qquad \langle D(\hat{z} - x), z - \hat{z}\rangle \geq 0. \tag{2.14}$$

Combining the last two inequalities, we obtain

$$\langle B(z - \hat{z}), z - \hat{z}\rangle \leq \langle (D - B)(\hat{z} - x), z - \hat{z}\rangle,$$

which, combined with the Cauchy-Schwarz inequality, yields

$$\|z-\hat{z}\|_B^2 \leq \|(B-D)(\hat{z}-x)\|\|z-\hat{z}\| \leq \|B^{-1}\|^{1/2}\|(B-D)(\hat{z}-x)\|\|z-\hat{z}\|_B.$$

Therefore, the desired inequality now follows from the last one. ■

Lemma 2.2.4 *Let $B, D \in \mathbb{B}$. Then, for every $x, \hat{x} \in \mathbb{R}^n$ and $\varepsilon \geq 0$, we have*

$$\|\tilde{y}_C^B(Bx) - y_C^D(D\hat{x})\|_B \leq \|x-\hat{x}\|_B + \|B^{-1}\|^{1/2}\|(B-D)(y_C^D(D\hat{x})-\hat{x})\| + \sqrt{\varepsilon}.$$

Proof. By Lemma 2.2.3, we obtain

$$\begin{aligned}\|\tilde{y}_C^B(Bx) - y_C^D(D\hat{x})\|_B &\leq \|\tilde{y}_C^B(Bx) - y_C^B(Bx)\|_B + \|y_C^B(Bx) - y_C^B(B\hat{x})\|_B + \|y_C^B(B\hat{x}) - y_C^D(D\hat{x})\|_B \\ &\leq \|\tilde{y}_C^B(Bx) - y_C^B(Bx)\|_B + \|y_C^B(Bx) - y_C^B(B\hat{x})\|_B + \|B^{-1}\|^{1/2}\|(B-D)(y_C^D(D\hat{x})-\hat{x})\|.\end{aligned}$$

Combining last inequality with (2.8) and (2.10), we find

$$\|\tilde{y}_C^B(Bx) - y_C^D(D\hat{x})\|_B \leq \sqrt{\varepsilon} + \|x-\hat{x}\|_B + \|B^{-1}\|^{1/2}\|(B-D)(y_C^D(D\hat{x})-\hat{x})\|,$$

which is equivalent to the desired inequality. ■

Lemma 2.2.5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. The exact solution $y_C^B(Bx - \nabla f(x))$ of (2.4) is a continuous function of $B \in \mathbb{B}$ and $x \in C$.*

Proof. Let $w := Bx - \nabla f(x)$, $\bar{w} := Bz - \nabla f(z)$ and $\hat{w} := Dz - \nabla f(z)$ with $x, z \in C$ and $B, D \in \mathbb{B}$. It follows from Definition 2.2.1 that

$$\begin{aligned}\langle B(y_C^B(w) - x) + \nabla f(x), y_C^B(\bar{w}) - y_C^B(w)\rangle &\geq 0, \\ \langle B(z - y_C^B(\bar{w})) - \nabla f(z), y_C^B(\bar{w}) - y_C^B(w)\rangle &\geq 0.\end{aligned}$$

Summing the above inequalities

$$\langle B(y_C^B(w) - y_C^B(\bar{w})) - B(x-z) + \nabla f(x) - \nabla f(z), y_C^B(\bar{w}) - y_C^B(w)\rangle \geq 0,$$

and after some manipulation

$$-\|y_C^B(\bar{w}) - y_C^B(w)\|_B^2 + \langle B(z-x), y_C^B(\bar{w}) - y_C^B(w)\rangle + \langle \nabla f(x) - \nabla f(z), y_C^B(\bar{w}) - y_C^B(w)\rangle \geq 0.$$

Then,

$$\frac{1}{\|B^{-1}\|}\|y_C^B(\bar{w}) - y_C^B(w)\|^2 \leq (\|B\|\|z-x\| + \|\nabla f(x) - \nabla f(z)\|)\,\|y_C^B(\bar{w}) - y_C^B(w)\|,$$

where in the left-hand side we used (2.2) and on the right-hand side we used the Cauchy-Schwarz inequality and consistency of the matrix norm.

Supposing $y_C^B(\bar{w}) \neq y_C^B(w)$, from the above inequality and Eq. (2.1), we arrive at

$$\|y_C^B(\bar{w}) - y_C^B(w)\| \leq L\left(L\|z - x\| + \|\nabla f(x) - \nabla f(z)\|\right). \tag{2.15}$$

Since inequality (2.15) is also valid when $y_C^B(\bar{w}) = y_C^B(w)$, x, z and B were taken arbitrarily, the inequality is valid for all $x, z \in C$ and $B \in \mathbb{B}$.

Now, again from Definition 2.2.1, we have in particular

$$\begin{aligned}\langle B(y_C^B(\bar{w}) - z) + \nabla f(z), y_C^D(\hat{w}) - y_C^B(\bar{w})\rangle &\geq 0,\\ \langle D(z - y_C^D(\hat{w})) - \nabla f(z), y_C^D(\hat{w}) - y_C^B(\bar{w})\rangle &\geq 0.\end{aligned}$$

Summing the above inequalities yields

$$\langle By_C^B(\bar{w}) - Bz + Dz - Dy_C^D(\hat{w}), y_C^D(\hat{w}) - y_C^B(\bar{w})\rangle \geq 0,$$

or, equivalently (after some manipulation),

$$\langle B(y_C^B(\bar{w}) - y_C^D(\hat{w})) + (D - B)(z - y_C^D(\hat{w})), y_C^D(\hat{w}) - y_C^B(\bar{w})\rangle \geq 0.$$

The above inequality leads to

$$\|y_C^D(\hat{w}) - y_C^B(\bar{w})\|_B^2 \leq \langle (D - B)(z - y_C^D(\hat{w})), y_C^D(\hat{w}) - y_C^B(\bar{w})\rangle.$$

Assuming $y_C^D(\hat{w}) \neq y_C^B(\bar{w})$, invoking (2.2) and the Cauchy-Schwarz inequality, we obtain

$$\|y_C^D(\hat{w}) - y_C^B(\bar{w})\| \leq \|B^{-1}\|\|z - y_C^D(\hat{w})\|\|D - B\| \leq L\|z - y_C^D(\hat{w})\|\|D - B\|, \tag{2.16}$$

where the last inequality follows from (2.1). On the other hand, from Definition 2.2.1, we also have

$$\langle D(y_C^D(\hat{w}) - z) + \nabla f(z), z - y_C^D(\hat{w})\rangle \geq 0,$$

or, equivalently,

$$\frac{1}{\|D^{-1}\|}\|z - y_C^D(\hat{w})\| \leq \|z - y_C^D(\hat{w})\|_D \leq \|\nabla f(z)\| \leq \|\nabla f(x)\| + \|\nabla f(x) - \nabla f(z)\|.$$

Combining the last inequality with (2.16) and Eq. (2.1), we obtain

$$\|y_C^D(\hat{w}) - y_C^B(\bar{w})\| \leq \|B^{-1}\|\|z - y_C^D(\hat{w})\|\|D - B\| \leq L^2\left(\|\nabla f(x)\| + \|\nabla f(x) - \nabla f(z)\|\right)\|D - B\|, \tag{2.17}$$

Since (2.17) also holds when $y_C^D(\hat{w}) = y_C^B(\bar{w})$, and because z, B, D were chosen arbitrarily, we conclude that it is valid for all $z \in C$ and $B, D \in \mathbb{B}$.

Finally, using (2.15), (2.17) and the triangle inequality, we find

$$\begin{aligned}\|y_C^B(w) - y_C^D(\hat{w})\| &\leq \|y_C^B(w) - y_C^B(\bar{w})\| + \|y_C^B(\bar{w}) - y_C^D(\hat{w})\|\\ &\leq L^2\|z - x\| + L(1 + L\|D - B\|)\|\nabla f(x) - \nabla f(z)\| + L^2\|\nabla f(x)\|\|D - B\|,\end{aligned}$$

which, combined with the fact that ∇f is continuous and $\mathbb{B}$ is compact, implies that $y_C^B(Bx - \nabla f(x))$ is continuous as a function of $x \in C$ and $B \in \mathbb{B}$. ■

Lemma 2.2.6 *For every $w, \hat{w} \in \mathbb{R}^n$ and $\varepsilon \geq 0$, we have*

$$\|\tilde{y}_C^I(w) - y_C^I(\hat{w})\|^2 \leq \|w - \hat{w}\|^2 + 2\varepsilon.$$

Proof. Since $\tilde{y}_C^I(w) \in C$ and $y_C^I(\hat{w}) \in C$, it follows from Definition 2.2.1 that

$$\langle \tilde{y}_C^I(w) - w, \tilde{y}_C^I(w) - y_C^I(\hat{w})\rangle \leq \varepsilon, \qquad \langle \hat{w} - y_C^I(\hat{w}), \tilde{y}_C^I(w) - y_C^I(\hat{w})\rangle \leq 0. \tag{2.18}$$

On the other hand, after some simple algebraic manipulations we have

$$\begin{aligned}\|w - \hat{w}\|^2 = \|\tilde{y}_C^I(w) - y_C^I(\hat{w})\|^2 &+ 2\langle w - \tilde{y}_C^I(w) - (\hat{w} - y_C^I(\hat{w})),\ \tilde{y}_C^I(w) - y_C^I(\hat{w})\rangle \\ &+ \|(w - \tilde{y}_C^I(w)) - (\hat{w} - y_C^I(\hat{w}))\|^2,\end{aligned}$$

which implies that

$$\begin{aligned}\|\tilde{y}_C^I(w) - y_C^I(\hat{w})\|^2 \leq \|w - \hat{w}\|^2 &+ 2\langle \tilde{y}_C^I(w) - w,\ \tilde{y}_C^I(w) - y_C^I(\hat{w})\rangle \\ &+ 2\langle \hat{w} - y_C^I(\hat{w}),\ \tilde{y}_C^I(w) - y_C^I(\hat{w})\rangle.\end{aligned}$$

By the last inequality, (2.18) and (2.1), yields

$$\|\tilde{y}_C^I(w) - y_C^I(\hat{w})\|^2 \leq \|w - \hat{w}\|^2 + 2\varepsilon,$$

which is equivalent to the desired inequality. ■

Chapter 3

A Modified inexact variable metric method for convex-constrained optimization problem

In this chapter, we propose a modified inexact variable metric (M-IVM) method for solving convex-constrained optimization problems. The convergence analysis of the proposed method is established under suitable conditions. Some numerical experiments are given in order to illustrate the performance of the new method. The material in this chapter is published in [40].

3.1 The method and its convergence analysis

In this section, we present and study an inexact variable metric method for solving (1.1). Basically, the method differs from the one studied in [4, 14] by using a different inaccuracy criterion for the search direction subproblems.

We are now able to formally describe the inexact method for solving (1.1).

Modified Inexact Variable Metric Method (M-IVM)

Step 0 (Initialization). Given $x_0 \in C$, $B_0 \in \mathbb{B}$, $\tau \in (0,1)$, an integer $M \geq 1$ and $\{\theta_k\} \subset [0,\infty)$. Set $k = 0$.

Step 1 (Inexact search direction). Set $w_k = B_k x_k - \nabla f(x_k)$. Compute $d_k = \tilde{y}_C^{B_k}(w_k) -$

x_k, where $\tilde{y}_C^{B_k}(w_k) \in C$ and

$$\langle B_k(x_k - \tilde{y}_C^{B_k}(w_k)) - \nabla f(x_k), y - \tilde{y}_C^{B_k}(w_k)\rangle \leq \varepsilon_k := \theta_k^2 \|\tilde{y}_C^{B_k}(w_k) - x_k\|_{B_k}^2, \quad \forall\, y \in C, \tag{3.1}$$

i.e., $\tilde{y}_C^{B_k}(w_k)$ is an ε_k–approximate solution of (2.4).

Step 2 (Termination Criterion). If $\|d_k\| = 0$, then **stop**.

Step 3 (Backtracking). Define $f_{max} = \max\{f(x_{k-j});\ 0 \leq j \leq \min\{k, M-1\}\}$. Set $\alpha \leftarrow 1$.

Step 3.1. If

$$f(x_k + \alpha d_k) \leq f_{max} + \tau\alpha\langle \nabla f(x_k), d_k\rangle, \tag{3.2}$$

then $\alpha_k = \alpha$, $x_{k+1} = x_k + \alpha d_k$, and go to Step 4. Otherwise, set $\alpha \leftarrow \alpha/2$ and go to Step 3.1.

Step 4 (Update of the Hessian approximation). Form a matrix $B_{k+1} \in \mathbb{B}$.

end

Remark 3.1.1 Some comments about the M-IVM are in order.

(i) Note that the problem (2.4) can be rewritten here, with $w_k = B_k x_k - \nabla f(x_k)$ and ignoring constant terms, as

$$\min_{y \in C} \frac{1}{2}\|y - (x_k - B_k^{-1}\nabla f(x_k))\|_{B_k}^2, \tag{3.3}$$

and, consequently, (3.1) is equivalent to

$$\langle x_k - B_k^{-1}\nabla f(x_k) - \tilde{y}_C^{B_k}(x_k), y - \tilde{y}_C^{B_k}(x_k)\rangle_{B_k} \leq \varepsilon_k, \quad \forall\, y \in C,$$

where $\tilde{y}_C^{B_k}(x_k) \in C$. We can say that $\tilde{y}_C^{B_k}(x_k)$ is an approximate projection (in the norm $\|\cdot\|_{B_k}$) of an unconstrained quasi-Newton step.

(ii) If $d_k = 0$, then $\tilde{y}_C^{B_k}(w_k) = x_k$. Hence, it follows from (3.1) that

$$\langle \nabla f(x_k), y - x_k\rangle \geq 0, \quad \forall\, y \in C,$$

i.e., $x_k \in C$ is a stationary point of (1.1). Conversely, if x_k is a stationary point of (1.1), then it follows from (3.1) with $y = x_k$ and the optimality condition that $\tilde{y}_C^{B_k}(w_k) = x_k$.

(iii) Notice that Step 1 is well-defined because the exact solution of (2.4) clearly satisfies (3.1). Nevertheless, iterative methods can be used to obtain an approximate solution of (2.4)

such that condition (3.1) holds. If the conditional gradient is employed, for example, the stopping criterion (2.12) now reads

$$\langle B_k(z_{j-1} - x_k) + \nabla f(x_k), \bar{z}_j - z_{j-1}\rangle \geq -\theta_k^2 \|z_{j-1} - x_k\|_{B_k}^2. \tag{3.4}$$

From item (i) in this remark, we observe that if $z_{j-1} = x_k$, then

$$\forall\, z \in C : \langle \nabla f(x_k), z - x_k\rangle = \langle \nabla f(x_k), z - z_{j-1}\rangle \geq \langle \nabla f(x_k), \bar{z}_j - z_{j-1}\rangle \geq 0,$$

showing that z_{j-1} is stationary for the original problem.

(iv) If $\theta_k = 0$ in (3.1), we obtain that $\tilde{y}_C^{B_k}(w_k)$ is the unique exact solution of the problem (2.4) and then, the inexact variable metric method reduces to its exact version. Additionally, if $B_k := \lambda_k I$ for every $k \geq 0$, where λ_k is as in (1.3), the inexact variable metric method corresponds to the inexact SPG method.

(v) As it will be proven later, the search directions generated by M-IVM are descent directions, which will imply that the backtracking process given in Step 3 is well-defined.

(vi) There are different choices for, or ways to build, the matrix B_k. For example, B_k can be the Hessian of function f if it is positive definite or a modification of it in order to guarantee the positive definiteness of the approximation. The approximation B_k can be a specific multiple of the identity matrix such as the spectral choice in [12, 14].

In order to investigate the global convergence of the method, we need to establish some properties of its search directions.

Proposition 3.1.2 *Assume that the sequence $\{\theta_k\}$ satisfies $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0, 1)$. Then, for every $k \geq 0$, we have*

$$\langle d_k, \nabla f(x_k)\rangle \leq -(1 - \bar{\theta}^2) L \|d_k\|^2 \tag{3.5}$$

and

$$\frac{1}{(1+\bar{\theta})L}\|y_C^{B_k}(w_k) - x_k\| \leq \|d_k\| \leq \frac{L}{1-\bar{\theta}^2}\|\nabla f(x_k)\|, \tag{3.6}$$

where $y_C^{B_k}(w_k)$ is the exact solution of the problem (2.4).

Proof. Since $d_k = \tilde{y}_C^{B_k}(w_k) - x_k$, from (3.1) with $y = x_k$, we have

$$\langle \nabla f(x_k), d_k\rangle \leq (\theta_k^2 - 1)\|d_k\|_{B_k}^2, \tag{3.7}$$

which, combined with the fact that $\theta_k \leq \bar{\theta} < 1$ for all $k \geq 0$, (2.1) and (2.2), yields

$$\langle \nabla f(x_k), d_k\rangle \leq -(1 - \bar{\theta}^2) L \|d_k\|^2.$$

Thus, (3.5) is proved. It follows from (3.7) and the Cauchy-Schwarz inequality that

$$(1-\theta_k^2)\|d_k\|_{B_k}^2 \leq -\langle \nabla f(x_k), d_k\rangle \leq \|\nabla f(x_k)\|\|d_k\|.$$

Hence, the second inequality in (3.6) now follows from (2.1), (2.2) and the fact that $\theta_k \leq \bar{\theta} < 1$ for all $k \geq 0$. Now, from (2.2) and the triangle inequality, we obtain

$$\begin{aligned}\|y_C^{B_k}(w_k) - x_k\| &\leq \|B_k^{-1}\|^{1/2}\|y_C^{B_k}(w_k) - x_k\|_{B_k} \\ &\leq \|B_k^{-1}\|^{1/2}\|y_C^{B_k}(w_k) - \tilde{y}_C^{B_k}(w_k)\|_{B_k} + \|B_k^{-1}\|^{1/2}\|\tilde{y}_C^{B_k}(w_k) - x_k\|_{B_k} \\ &\leq \|B_k^{-1}\|^{1/2}[\sqrt{\varepsilon_k} + \|d_k\|_{B_k}],\end{aligned}$$

where the last inequality is due to (2.10) and $d_k = \tilde{y}_C^{B_k}(w_k) - x_k$. Since $\varepsilon_k = \theta_k^2\|d_k\|_{B_k}^2$ (see Step 1 of the M-IVM), it follows from the last inequality that

$$\|y_C^{B_k}(w_k) - x_k\| \leq (1+\theta_k)\|B_k^{-1}\|^{1/2}\|d_k\|_{B_k}.$$

Therefore, the first inequality in (3.6) now follows from (2.1), (2.2) and the fact that $\theta_k \leq \bar{\theta}$ for all $k \geq 0$.

■

We next establish the global convergence of the M-IVM.

Theorem 3.1.3 *Assume that the level set $C_0 := \{x \in C : f(x) \leq f(x_0)\}$ is bounded and the sequence $\{\theta_k\}$ satisfies $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0, 1)$. Then, either the M-IVM stops at some stationary point x_k, or every limit point of the generated sequence is stationary.*

Proof. If the M-IVM stops at a point x_k, then $d_k = 0$. Hence, $\tilde{y}_C^{B_k}(w_k) = x_k$ and it follows from (3.1) that

$$\langle \nabla f(x_k), y - x_k\rangle \geq 0, \quad \forall\, y \in C,$$

i.e., x_k is a stationary point of (1.1). If $d_k \neq 0$, for every $k \geq 0$, it follows from (3.5) that d_k is a descent direction. So, the backtracking process given in Step 3 is well-defined, and, as a consequence, the M-IVM is also well-defined. Our goal is now to show that every limit point of the $\{x_k\}$ is a stationary point of (1.1). Let $l(k)$ be an integer such that $k - \min\{k, M-1\} \leq l(k) \leq k$ and

$$f(x_{l(k)}) = \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}).$$

Using the first part of the proof of the theorem in [44] with $m(k) := \min\{k, M-1\}$ (note that this choice of $m(k)$ satisfies the conditions of the mentioned theorem), it can be shown that $\{f(x_{l(k)})\}$ is monotonically nonincreasing, and from the boundedness of C_0 we have that $\{f(x_{l(k)})\}$ admits a limit for $k \to \infty$. From (3.2), it follows, for $k > M-1$, that

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) + \tau\alpha_{(l(k)-1)}\langle \nabla f(x_{(l(k)-1)}), d_{(l(k)-1)}\rangle. \tag{3.8}$$

Now, since $\alpha_{(l(k)-1)} > 0$ and $\langle \nabla f(x_{(l(k)-1)})), d_{(l(k)-1)} \rangle < 0$, by taking limits in (3.8), it follows that

$$\lim_{k\to\infty} \alpha_{(l(k)-1)} \langle \nabla f(x_{(l(k)-1)}), d_{(l(k)-1)} \rangle = 0.$$

Moreover, from (3.5) and (3.6), we conclude that

$$\lim_{k\to\infty} \alpha_{(l(k)-1)} \| y_C^{B_{(l(k)-1)}}(w_{(l(k)-1)}) - x_{(l(k)-1)} \|^2 = 0,$$

and following the idea in the proof of the theorem of [44], we can write

$$\lim_{k\to\infty} \alpha_k \| y_C^{B_k}(w_k) - x_k \|^2 = 0. \tag{3.9}$$

Let $x_* \in C$ be a limit point of $\{x_k\}$. Relabel $\{x_k\}$ a subsequence converging to x_*. From (3.9), there exists a subsequence of indices $K_1 \subset K$ such that: (i) $\lim_{k\in K_1} \| y_C^{B_k}(w_k) - x_k \| = 0$ or (ii) $\lim_{k\in K_1} \alpha_k = 0$.
(i) By the compactness of $\mathbb{B}$ we can extract a subsequence of indices $K_2 \subset K_1$ such that

$$\lim_{k\in K_2} B_k = B_* \in \mathbb{B}.$$

Hence, since $y_C^{B_k}(w_k) = y_C^{B_k}(B_k x_k - \nabla f(x_k))$, by continuity of $y_C^B(w)$ (see Lemma 2.2.5), we have $\| y_C^{B_*}(w_*) - x_* \| = 0$, or equivalently, $y_C^{B_*}(w_*) = x_*$, where $w_* = B_* x_* - \nabla f(x_*)$. Therefore, the definition $y_C^{B_*}(w_*)$ (see Definition 2.2.1) implies that

$$\langle \nabla f(x_*), y - x_* \rangle \geq 0, \quad \forall\, y \in C,$$

i.e., x_* is a stationary point of (1.1).
(ii) Let α_k be the step chosen in the Step 3.2 such that $\alpha_k = \bar{\alpha}_k/2$, where $\bar{\alpha}_k$ was the last step that failed in (3.2), i.e.

$$f(x_k + \bar{\alpha}_k d_k) > \max_{0\leq j\leq \min\{k,M-1\}} f(x_{k-j}) + \tau \bar{\alpha}_k \langle \nabla f(x_k), d_k \rangle \geq f(x_k) + \tau \bar{\alpha}_k \langle \nabla f(x_k), d_k \rangle. \tag{3.10}$$

Now define $s_k = \bar{\alpha}_k d_k$. By the mean value theorem, there exists $\mu_k \in [0, 1]$ such that the relation in (3.10) can be written as

$$\langle \nabla f(x_k + \mu_k s_k), s_k \rangle = f(x_k + s_k) - f(x_k) > \tau \langle \nabla f(x_k), s_k \rangle. \tag{3.11}$$

On the other hand, as $\{x_k\}$ is bounded and f has continuous derivatives, we have, by (3.6), that $\{d_k\}$ is bounded. Thus, since $s_k = 2\alpha_k d_k$, and $\lim_{k\in K_1} \alpha_k = 0$, we obtain that s_k goes to zero as $k \in K_1$ goes to infinity. So, from (3.11), we have

$$\langle \nabla f(x_k + \mu_k s_k), \frac{s_k}{\|s_k\|} \rangle > \tau \langle \nabla f(x_k), \frac{s_k}{\|s_k\|} \rangle.$$

By taking limit in the last inequality as $k \in K_3$ goes to infinity, where $K_3 \subset K_1$ is such that $\lim_{k\in K_3}\{s_k/\|s_k\|\}$ converges to s, we obtain $(1-\tau)\langle \nabla f(x_*), s\rangle \geq 0$. Since $(1-\tau) > 0$, we have

$$\langle \nabla f(x_*), s\rangle \geq 0. \tag{3.12}$$

Now, as d_k is a descent direction for f at x_k (see (3.5)) and $s_k = \bar{\alpha}_k d_k$, we find

$$\langle \nabla f(x_k), \frac{s_k}{\|s_k\|}\rangle < 0.$$

Hence, $\langle \nabla f(x_*), s\rangle \leq 0$, which, combined with (3.12), implies that $\langle \nabla f(x_*), s\rangle = 0$. Using (3.5), (3.6) and the definition of s_k, we have

$$\langle \nabla f(x_k), \frac{s_k}{\|s_k\|}\rangle \leq -(1-\bar{\theta}^2)L\|d_k\| \leq -(1-\bar{\theta})\|y_C^{B_k}(w_k) - x_k\|.$$

By the compactness of $\mathbb{B}$ we can extract a subsequence of indices $K_4 \subset K_3$ such that $\lim_{k\in K_4} B_k = B_* \in \mathbb{B}$. Therefore, by taking limit in the last inequality as $k \in K_4$ goes to infinity, we have

$$0 = \langle \nabla f(x_*), s\rangle \leq -(1-\bar{\theta})\|y_C^{B_*}(w_*) - x_*\|.$$

Since $\bar{\theta} < 1$, we obtain $y_C^{B_*}(w_*) = x_*$, which, from the definition $y_C^{B_*}(w_*)$ (see Definition 2.2.1), implies that x_* is a stationary point of (1.1). ■

3.2 Numerical experiments

We split the numerical experiments in two sets. First, in Subsection 3.2.1, where polyhedral feasible sets $C = \{x \in \mathbb{R}^n \ : \ Ax \leq b\}$ are considered, we aim to evaluate the impact of the new inexactness criterion (3.1) of M-IVM in comparison with the criterion (1.4) used in the Inexact SPG (ISPG) of [4]. Then, in Subsection 3.2.2, we consider a feasible set for which the use of inexact variable metric methods is quite appealing (because the cost of an exact solution of (2.4) is prohibitive) and we show that M-IVM achieves good results with respect to its exact counterpart and an off-the-shelf solver.

All experiments were carried out in Matlab R2018b, in a laptop running Mac OS X 10.13.6, with 8GB of RAM and 1.8 Ghz Intel Core i5 processor.

We implemented M-IVM with the following parameters: $\tau = 10^{-4}$, $\lambda_{\min} = 10^{-10}$, $\lambda_{\max} = 10^{10}$ and $\theta_k = \bar{\theta} = 0.9995$.

3.2.1 Polyhedral feasible set

In order to put in perspective the new inexactness criterion (3.1) with the previously proposed criterion (1.4), we consider a subset of the linearly constrained problems from the CUTEr collection [43] used in [4, Table 2] and compare the results of ISPG with those obtained by M-IVM.

Our implementation of ISPG consists in a modification of M-IVM where Step 1 is replaced by the dual approach of [4] to inexactly solve the quadratic subproblems according to the criterion (1.4) (see [4, Algorithm 5.1] for details). This dual approach demands an iterative method to minimize a non-negative constrained convex quadratic. For this task, we have used the MINQ8 solver [49] which implements an active-set method combining coordinate searches and subspace minimization steps. Algorithm 5.1 of [4] was embedded in MINQ8 to verify criterion (1.4) with the same parameter values as in [4, Section 8.5], namely, $\eta = 0.8\beta$ and $\beta = 0.85$ ($\beta \in (0, 1)$ multiplies the maximum allowed step-size to keep the iterates interior enough).

For both methods, the tolerance in the stopping criterion $\|d_k\| < \epsilon$ was set to $\epsilon = 10^{-6}$. For this set of experiments, we considered the variant of M-IVM with $M = 10$, $B_k = \lambda_k I$, with λ_k as in (1.3) (and $\lambda_0 = 1$).

Since the feasible set $C = \{x \in \mathbb{R}^n \ : \ Ax \leq b\}$, with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, is described by linear inequality constraints, the M-IVM subproblems (2.4) are in fact (strictly convex) quadratic programming problems. Bound constraints were treated as ordinary inequality constraints. The subproblems (2.4) were solved by using a variant of the Frank-Wolfe algorithm known as Away-Step Conditional Gradient (ASCG) [45], whose subproblems (see Eq. (2.11)) were solved by the revised simplex method using Bland's rule to avoid cycling [64, Section 13.3].

In order to handle problems with unbounded C, we have included an additional constraint corresponding to the ball of (2.13) in infinity norm, so that the subproblems (2.11) are well-defined.

Table 3.1 presents the number of variables n, number of original inequality constraints m, the number of outer (OUTIT) and inner (INNIT) iterations required by each method, the CPU time in seconds and the objective value $f(x_k)$ at the last iterate. From these figures, we observe that M-IVM requires less outer iterations than ISPG, at the cost of more inner iterations for certain problems.

Concerning the computational cost of each inner iteration, in ISPG it is the cost of an iteration of the non-negative constrained convex quadratic solver, whereas in each inner iteration of M-IVM, which uses the ASCG, a linear programming problem has to be solved[1].

[1]For MINQ8 solver, it seems that the main cost per iteration is $O(|I|^3)$ for solving a linear system of size

			ISPG				M-IVM			
Problem	n	m	OUTIT	INNIT	time	$f(x_k)$	OUTIT	INNIT	time	$f(x_k)$
HS24	2	3	9	22	0.24	-1.000	7	14	0.13	-1.000
HS35	3	1	16	19	0.22	0.1111	12	37	0.19	0.1111
HS35I	3	1	16	19	0.21	0.1111	12	37	0.17	0.1111
HS36	3	1	8	43	0.32	-3300.	1	4	0.07	-3300.
HS37	3	2	11	24	0.21	-3456	11	58	0.29	-3456
HS44NEW	4	6	15	79	0.31	-15	3	3	0.09	-15
HS76	4	3	12	28	0.23	-4.6818	8	101	0.29	-4.6818
HS76I	4	3	12	28	0.29	-4.6818	8	101	0.28	-4.6818
SIPOW1	2	2000	11	23	3.51	-1.0000	2	4	0.14	-1
SIPOW1M	2	2000	10	23	2.89	-1.0000	2	4	0.14	-1.0000
SIPOW2	2	2000	9	22	1.55	-1.0000	2	4	0.15	-1
SIPOW2M	2	2000	9	20	1.37	-1.0000	2	4	0.15	-1.0000
SIPOW3	4	2000	11	309	4.71	0.5347	2	7	0.76	0.5346
SIPOW4	4	2000	10	397	4.48	0.2724	2	5	1.14	0.2724

Table 3.1: Comparison of ISPG and M-IVM on CUTEr problems

Although this may suggest that the inner iteration of the latter is more expensive, CPU times in Table 3.1 reveal that this is not always the case. Furthermore, in terms of time, M-IVM is quite competitive with ISPG.

The results for some problems deserve a separate explanation. We remark that the problems SIPOW (see [69] for details) are in fact linear programming problems that would be solved in a single (outer) iteration by M-IVM if $\lambda_0 = 0$. Nevertheless, we also observe a better performance of M-IVM for problems where the solution is an extreme point of the feasible polyhedron, as in HS24, HS36 and HS44NEW. In this latter case, what happened is that the solution of the linear programming problem in some iteration of ASCG coincided with the optimal solution of the original problem.

$|I| \leq m$, where I is the set of inactive constraints. On the other hand, each iteration of the revised simplex costs $O(m^2)$. If the revised simplex takes q iterations, then $O(qm^2)$ is the cost for the inner iteration of M-IVM, which is comparable with the cost of a MINQ8 iteration when q and $|I|$ are close to m.

3.2.2 Least squares on the spectrahedron

In this subsection, we consider the least squares problem over the spectrahedron:

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \quad & \frac{1}{2}\|AX - Z\|_F^2 \\ \text{s.t.} \quad & \operatorname{tr}(X) = 1 \\ & X \succeq 0, \end{aligned} \tag{3.13}$$

where $A, Z \in \mathbb{R}^{m \times n}$, with $m > n$, $\mathbb{S}^n$ denotes the vector space of symmetric matrices of order n equipped with the trace inner product $\langle X, Y \rangle = \operatorname{tr}(XY)$ and induced norm $\|X\|_F^2 = \langle X, X \rangle$, and $X \succeq 0$ means that X is positive semidefinite.

Problem (3.13) is related to important applications in many areas. For example, in nonlinear optimization, it can be used to estimate positive definite approximations for the inverse Hessian in quasi-Newton methods, whereas in structural analysis it can be used to estimate the compliance matrix of an elastic structure (see [80] for details).

Clearly, the feasible set $C = \{X \in \mathbb{S}^n \ : \ \operatorname{tr}(X) = 1, \ X \succeq 0\}$ is convex and compact whereas the objective function of (3.13) is strictly convex, provided $\operatorname{rank}(A) = n$.

We remark that, for this feasible set, the computation of the exact orthogonal projection[2] of a point $Y \in \mathbb{S}^n$ onto C requires the full eigendecomposition of Y which is prohibitive for large values of n (for details, see [37, 47]). Since the projection problem is equivalent to (3.3) when B is a positive multiple of the identity, and (3.3) in its turn is equivalent to (2.4) for any positive definite B, we expect that the cost of solving (2.4) *exactly* becomes also prohibitive for large dimensions. Therefore, it seems reasonable to consider inexact variable metric methods in this case.

Since C is neither polyhedral nor a finite intersection of easy convex sets, the approaches in [4, 14] are not directly applicable.

On the other hand, if an ε-approximate solution of (2.4) is allowed (in the sense of (2.5)), one could employ, for example, the Frank-Wolfe algorithm [32] whose iteration cost is dictated by an extreme eigenpair computation when C is the spectrahedron (see [38] and references therein). If only a few Frank-Wolfe iterations are required to achieve (2.5), then overall savings, in terms of computational effort, may be considerable when running variants of M-IVM.

To numerically investigate this claim, we consider random instances of problem (3.13) and compare the performance of variants of M-IVM with SPG using exact projections [37] and an interior point method [75].

[2] with respect to the Frobenius norm.

The first group of problems consists of dense small problems with $n < m \leq 1000$. The matrices A were randomly generated with entries sampled from a uniform distribution in the interval $[0, 1]$. Then, given a positive integer q, we build a symmetric matrix $\tilde{X}$ with q eigenvalues equal to $1/q$, one equal to -1, and all others equal to zero. Finally, we set $Z = A\tilde{X}$. In general, this procedure results in nonzero residue problems. Note that the construction of the Z results that the solution of the unconstrained problem is outside the feasible set C.

For this group of problems, we consider two variants of M-IVM, namely, "Inexact Newton" where $B_k = A^T A$ and "Inexact SPG" where $B_k = \lambda_k I$, with λ_k as in (1.3), and compare them with the off-the-shelf solver QSDP [75] which implements an interior point method for convex quadratic semidefinite programming problems.

Since the classic Frank-Wolfe is known for its slow $O(1/\varepsilon)$ convergence [35], we consider a variant of the conditional gradient proposed in [2], and further enhanced and specialized to the spectrahedron in [24], that we shall call FW-p. FW-p exploits an estimate of the solution with rank p and at each iteration computes p eigenpairs (rather than one eigenpair in the classic FW). It achieves $O(\kappa \log(1/\varepsilon))$ convergence rate, where κ is the condition number[3] of the subproblem (2.4). This scheme fits well the Inexact SPG because $B_k = \lambda_k I$ implies in $\kappa = 1$. Preliminary experiments revealed that it also works fine with Inexact Newton as long as B_k is not ill-conditioned.

In both cases, the strategy for "guessing" the solution rank p_* is paramount for achieving faster convergence. Since the references [2, 24] do not provide a strategy with theoretical guarantees, here we also use a heuristic to update the rank estimate p: we start with $p = 1$ and increase the value of the rank estimate to $p + \delta$ whenever the decrease in the subproblem objective function is not substantial[4]. For the small-dense problems in Table 3.2, $\delta = p$ and for the large-sparse problems in Tables 3.3 and 3.4, $\delta = 1$. We also remark that the value of p is decreased to $p - r$ when r of the p kept eigenvalues are close to zero.

The tolerance in the stopping criterion $\|d_k\| < \epsilon$ in the variants of M-IVM was set to $\epsilon = 10^{-4}$ for the problems in Table 3.2 and $\varepsilon = 10^{-3}$ for the problems of Tables 3.3 and 3.4. The tolerance for the duality gap in QSDP was set to 10^{-3}.

Concerning the parameter M of the nonmonotone line search, we observed in preliminary numerical experiments that the full-step ($\alpha_k = 1$) was always accepted in the "Inexact Newton", so we kept $M = 1$ for this variant. For the "Inexact SPG", we did not observe a pronounced improvement for $M = 5$ or $M = 10$ for this test set, thus we decided to go on with the monotone line search ($M = 1$).

[3] Assuming that the convex function q is α-strongly convex and L-smooth, the corresponding condition number is given by $\kappa = L/\alpha$. See [2] for details.

[4] the reduction in the objective should be at least one percent of its value in the previous iterate.

For each problem, we consider 3 starting points given by $X_0(\gamma) = (1-\gamma)(1/n)I + \gamma\hat{X}$, where $\hat{X} = e_1 e_1^T$ (e_1 is the first canonical vector) and $\gamma \in \{0, 0.5, 0.99\}$.

Table 3.2 brings the number of iterations, running time in seconds, and the achieved objective value $f(X_k)$. The smallest running time for each problem is highlighted in bold. From these results, we observe that the variants of M-IVM provide a non-negligible speed-up with respect to QSDP in the majority of the problems.

In the second group of problems, we consider sparse matrices with dimensions $m > n \geq 1000$. The matrix A was build using the command `sprand(m,n,1e-4)` from Matlab, and $\tilde{X} = QDQ^T$, where Q is the product of a few Givens rotation matrices and D is a diagonal matrix with q entries equal to one and all others equal to zero. This ensures that $Z = A\tilde{X}$ is also sparse.

For this second test set, the interior point solver QSDP was left out of comparison due to excessively high running times. We replace it by a version of SPG where the projection is computed "exactly" as in [37]. This version is referred in Table 3.3 and 3.4 as "Exact SPG".

From Tables 3.3 and 3.4, we observe that the Inexact SPG surpassed SPG with exact projections in the majority of problems. Inexact Newton also shows a good performance and becomes faster than Exact SPG as n and m increase.

				QSDP			Inexact Newton			Inexact SPG		
n	m	q	γ	it	time	$f(X_k)$	it	time	$f(X_k)$	it	time	$f(X_k)$
10	100	4	0	9	1.24	9.13	3	**0.24**	9.13	19	0.36	9.13
			0.5	9	0.67	9.13	4	**0.11**	9.13	23	0.27	9.13
			0.99	9	0.55	9.13	4	**0.09**	9.13	22	0.19	9.13
50	200	4	0	16	2.54	16.51	4	**0.61**	16.51	50	1.04	16.51
			0.5	16	1.98	16.51	4	**0.33**	16.51	41	0.71	16.51
			0.99	16	1.54	16.51	4	**0.55**	16.51	50	0.74	16.51
50	200	10	0	14	1.54	16.70	4	**0.41**	16.70	29	0.73	16.70
			0.5	14	1.43	16.70	4	**0.41**	16.70	49	0.92	16.70
			0.99	14	1.30	16.70	5	**0.47**	16.70	54	0.88	16.70
100	400	5	0	18	4.11	29.07	3	**0.81**	29.04	50	1.32	29.04
			0.5	18	3.23	29.07	4	**0.59**	29.04	73	1.28	29.04
			0.99	18	3.07	29.07	4	**1.05**	29.04	72	1.56	29.04
200	800	5	0	23	11.04	53.02	4	6.54	52.98	86	**3.82**	52.99
			0.5	23	11.31	53.02	4	6.31	52.98	79	**3.03**	52.99
			0.99	23	11.38	53.02	4	6.05	52.98	79	**2.97**	52.98
200	800	20	0	20	10.53	52.60	4	17.46	52.58	53	**2.79**	52.59
			0.5	20	8.84	52.60	4	14.79	52.58	75	**4.91**	52.59
			0.99	20	8.83	52.60	5	18.05	52.58	91	**4.02**	52.60
400	1000	5	0	28	74.01	65.31	5	57.81	65.25	96	**12.08**	65.28
			0.5	28	72.10	65.31	5	41.92	65.25	102	**12.02**	65.35
			0.99	28	73.05	65.31	5	44.79	65.25	125	**13.35**	62.28

Table 3.2: Numerical results for dense small problems.

				Exact SPG			Inexact Newton			Inexact SPG		
n	m	q	γ	it	time	$f(X_k)$	it	time	$f(X_k)$	it	time	$f(X_k)$
1000	2000	10	0.0	5	**2.03**	0.0794	4	5.96	0.0794	11	2.56	0.0794
			0.5	10	3.28	0.0794	3	4.26	0.0794	9	**2.25**	0.0794
			1.0	7	**2.39**	0.0794	4	6.21	0.0794	11	2.49	0.0794
1000	2000	20	0.0	6	**2.34**	0.2311	4	5.85	0.2311	7	2.38	0.2311
			0.5	7	2.48	0.2311	4	5.37	0.2311	7	**2.31**	0.2311
			1.0	8	**2.73**	0.2311	4	7.41	0.2311	10	3.38	0.2311
2000	4000	10	0.0	5	12.33	0.2304	3	8.22	0.2304	5	**6.47**	0.2304
			0.5	6	14.83	0.2304	3	7.94	0.2304	6	**7.65**	0.2304
			1.0	5	12.20	0.2304	3	**8.06**	0.2304	9	9.06	0.2304
2000	4000	20	0.0	5	12.22	0.9442	4	11.11	0.9442	5	**7.94**	0.9442
			0.5	5	12.04	0.9442	4	10.67	0.9442	7	**8.33**	0.9442
			1.0	6	20.92	0.9442	4	10.97	0.9442	6	**7.21**	0.9442
3000	6000	10	0.0	5	37.87	0.1377	3	19.14	0.1377	6	**18.71**	0.1377
			0.5	5	37.82	0.1377	3	**18.04**	0.1377	6	19.11	0.1377
			1.0	7	49.08	0.1377	3	18.45	0.1377	5	**15.96**	0.1377

Table 3.3: Numerical results for sparse medium-scale problems.

				Exact SPG			Inexact Newton			Inexact SPG		
n	m	q	γ	it	time	$f(X_k)$	it	time	$f(X_k)$	it	time	$f(X_k)$
3000	6000	20	0.0	3	24.95	1.0832	4	24.68	1.0832	4	**15.26**	1.0832
			0.5	4	30.90	1.0832	4	24.52	1.0832	6	**16.98**	1.0833
			1.0	4	30.38	1.0832	4	24.55	1.0832	5	**17.44**	1.0832
4000	8000	10	0.0	5	85.50	0.6192	3	34.06	0.6192	6	**33.98**	0.6192
			0.5	4	109	0.6192	3	32.46	0.6192	7	**32.22**	0.6192
			1.0	5	112	0.6192	3	32.69	0.6192	6	**29.55**	0.6192
4000	8000	20	0.0	5	117	2.9650	5	56.92	2.9650	5	**35.69**	2.9650
			0.5	5	116	2.9650	5	56.27	2.9650	10	**54.45**	2.9650
			1.0	6	144	2.9650	5	58.97	2.9650	7	**40.63**	2.9650
5000	10000	10	0.0	6	258	0.9923	4	93.54	0.9923	8	**64.68**	0.9923
			0.5	7	296	0.9923	4	90.46	0.9923	8	**69.61**	0.9923
			1.0	7	300	0.9923	4	90.51	0.9923	9	**77.77**	0.9923
5000	10000	20	0.0	3	107	2.9388	4	89.92	2.9388	7	**80.02**	2.9388
			0.5	4	138	2.9388	4	91.48	2.9388	6	**86.28**	2.9388
			1.0	4	139	2.9388	4	91.86	2.9388	8	**83.77**	2.9388

Table 3.4: Numerical results for sparse medium-scale problems.

Chapter 4

Gauss-Newton methods with approximate projections for convex-constrained nonlinear least squares problems

In this Chapter, we present Gauss-Newton methods with approximate projections for solving convex-constrained nonlinear least squares problems. We first propose a local method and discuss its convergence theorem as well as results on the rate. Our analysis is done by using a majorant condition which covers two large families of nonlinear functions, namely, one satisfying a Lipschitz condition and another one satisfying a Smale condition. We then propose a global Gauss-Newton method with approximate projections for solving nonlinear least squares problems. Our global version combines the local method with the nonmonotone line search based on [44]. Some numerical experiments of proposed methods are discussed. The results of this chapter are published in [39].

4.1 The method and its local convergence

This section describes and investigates a Gauss-Newton method with approximate projections (GNM-AP) for solving (1.8). Basically, the method consists of computing an approximate projection of the unconstrained Gauss-Newton step onto the feasible set C. The main local convergence theorem of the method and results on its rate are established, and its proof is postponed to Subsection 4.1.1. As a result of our analysis, made using a majorant condition, we covers two applications of such condition: one satisfying a Lipschitz

condition and another one satisfying a Smale condition. We also present, in this section, two examples in which all conditions of the convergence theorem hold. The convergence results for these special cases are established in this section. The GNM-AP is formally described as follows.

GNM-AP

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

Step 1 (Projected Gauss-Newton step). Define $B_k = F'(x_k)^T F'(x_k)$ and compute $w_k = B_k x_k - F'(x_k)^T F(x_k) \in \mathbb{R}^n$. Compute $x_{k+1} \in C$ such that

$$\langle w_k - B_k x_{k+1}, y - x_{k+1} \rangle \leq \varepsilon_k := \theta_k^2 \|x_{k+1} - x_k\|_{B_k}^2, \quad \forall\, y \in C, \tag{4.1}$$

i.e., x_{k+1} is an ε_k–approximate solution of (2.4), with $B = B_k$ and $w = w_k$.

Step 2 (Termination criterion and update). If $x_{k+1} = x_k$, then **stop**; Otherwise, set $k \leftarrow k + 1$ and go to Step 1.

end

Remark 4.1.1 Some comments about the GNM-AP are in order.

(i) Note that, if $B_k \in \mathbb{B}$, then (4.1) is equivalent to

$$\langle x_k - B_k^{-1} F'(x_k)^T F(x_k) - x_{k+1}, y - x_{k+1} \rangle_{B_k} \leq \varepsilon_k, \quad \forall\, y \in C,$$

therefore, we can say that x_{k+1} is an approximate projection (in the norm $\|\cdot\|_{B_k}$) of an Gauss-Newton step $y_k := x_k - B_k^{-1} F'(x_k)^T F(x_k)$. Since the Gauss-Newton step y_k may be infeasible for the constraint set C, it is necessary to compute an ε_k–approximate projection of it onto C. As already mentioned, such an approximate projection can be efficiently computed, for example, by the conditional gradient method

(ii) In Step 2, if $x_{k+1} = x_k$, it follows from Step 1 and definition of w_k that

$$0 \geq \langle w_k - B_k x_{k+1}, y - x_{k+1} \rangle = \langle -F'(x_k)^T F(x_k), y - x_k \rangle_{B_k}$$

for all $y \in C$, i.e. x_k is a stationary point of (1.1).

(iii) The characterization of x_{k+1} as an approximate projection of the unconstrained Gauss-Newton step with respect to the norm $\|\cdot\|_{B_k}$ is essential in order to establish the local convergence of the method as well as its fast convergence rate.

In order to analyze GNM-AP, we suppose that the following assumptions hold:

(A1) The point x_* satisfies the first-order necessary condition for (1.8), i.e.

$$\langle F'(x_*)^T F(x_*), x - x_* \rangle \geq 0, \quad \forall\, x \in C,$$

and $F'(x_*)$ is injective;

(A2) The sequence $\{\theta_k\}$ satisfies $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0, 1)$.

For simplicity, let us consider the following constants

$$c := \|F(x_*)\|,\ \beta := \|F'(x_*)^\dagger\|,\ \kappa := \beta \left\|F'(x_*)\right\|,\ \delta := \sup\left\{t \in [0, R) : \mathcal{B}(x_*, t) \subset \mathbb{U}\right\}, \tag{4.2}$$

where $R > 0$ is a given scalar.

We first state a local convergence theorem for GNM-AP under a majorant condition. For technical reasons and for the convenience of the reader, the proof of the next theorem will be given in the next subsection.

Theorem 4.1.2 *Suppose that there exists a continuously differentiable function $f : [0,\ R) \to \mathbb{R}$ such that*

$$\beta \left\|F'(x) - F'(x_* + \tau(x - x_*))\right\| \leq f'\left(\sigma(x)\right) - f'\left(\tau\sigma(x)\right), \tag{4.3}$$

where $x \in \mathcal{B}(x_*, \delta)$, $\tau \in [0, 1]$ and $\sigma(x) := \|x - x_*\|$, and

h1) *$f(0) = 0$ and $f'(0) = -1$;*

h2) *f' is convex and strictly increasing;*

h3) *$c\beta((1 + \sqrt{2})\kappa + 1)D^+ f'(0) + \kappa\bar{\theta} < 1 - \bar{\theta}$.*

Let be given positive constants $\nu := \sup\left\{t \in [0, R) : f'(t) < 0\right\}$,

$$\rho := \sup\Bigg\{t \in (0, \nu) : \frac{[f'(t) + 1 + \kappa]\left[(1 - \bar{\theta})tf'(t) - f(t) + c\beta(1 + \sqrt{2})(f'(t) + 1)\right] + c\beta\left[f'(t) + 1\right]}{(1 - \bar{\theta})t[f'(t)]^2} < 1\Bigg\}, \tag{4.4}$$

$$r := \min\left\{\rho,\ \delta\right\}.$$

Then GNM-AP with starting point $x_0 \in C \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$ is well-defined, the generated $\{x_k\}$ is contained in $\mathcal{B}(x_*, r) \cap C$, converges to x_* and satisfies

$$\|x_{k+1} - x_*\| < \|x_k - x_*\| \tag{4.5}$$

and

$$\|x_{k+1}-x_*\| \leq \frac{[f'(\sigma(x_0))+1+\kappa]\,[\sigma(x_0)f'(\sigma(x_0))-f(\sigma(x_0))]}{(1-\theta_k)[\sigma(x_0)f'(\sigma(x_0))]^2}\|x_k-x_*\|^2 + \frac{\left[(1+\sqrt{2})c\beta\,[f'(\sigma(x_0))+1]-\theta_k\sigma(x_0)f'(\sigma(x_0))\right]\,[f'(\sigma(x_0))+1+\kappa]}{(1-\theta_k)\sigma(x_0)\,[f'(\sigma(x_0))]^2}\|x_k-x_*\| + \frac{c\beta\,[f'(\sigma(x_0))+1]}{(1-\theta_k)\sigma(x_0)\,[f'(\sigma(x_0))]^2}\|x_k-x_*\|, \quad (4.6)$$

for all $k=0,1,\ldots$.

Remark 4.1.3 **(i)** Since $\|x_k-x_*\|<\sigma(x_0)=\|x_0-x_*\|$ (see (4.5)), it follows from (4.6) and **(A2)** that

$$\|x_{k+1}-x_*\| \leq \left[\frac{[f'(\sigma(x_0))+1+\kappa]\left[(1-\bar{\theta})\sigma(x_0)f'(\sigma(x_0))-f(\sigma(x_0))+c\beta(1+\sqrt{2})(f'(\sigma(x_0))+1)\right]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2} + \frac{c\beta\,[f'(\sigma(x_0))+1]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2}\right]\|x_k-x_*\|$$

which, combined with (4.4) and the fact that $x_0\in C\cap\mathcal{B}(x_*,r)\backslash\{x_*\}$, implies that GNM-AP is linearly convergent to x_*.

(ii) Note that, if $c=0$ and $\limsup_{k\to+\infty}\theta_k=0$, then (4.6) implies that GNM-AP converges quadratically to x_*.

(iii) If the scalar $\bar{\theta}$ in **(A2)** is equal to zero (in particular, $\theta_k=0$ for all $k\geq 0$), then iterative x_{k+1} in Step 1 of GNM-AP corresponds to the exact solution of (2.4), with $B=B_k$ and $w=w_k$. In this case, Theorem 4.1.2 is similar to [29, Theorem 7], which is related to the Gauss-Newton method for solving unconstrained nonlinear least squares problems.

Before specializing Theorem 4.1.2 for two important classes of functions, we present an example in which all conditions of Theorem 4.1.2 hold. The following result, which gives a simpler condition to check that condition (4.3) whenever the functions under consideration are twice continuously differentiable, is needed.

Lemma 4.1.4 *Let* $x_*\in\mathbb{U}$ *and* $R>0$ *be given, and assume that* F *is twice continuously differentiable on* $\mathbb{U}$. *If there exists a function* $f:[0,R)\to\mathbb{R}$ *twice continuously differentiable and satisfying*

$$\beta\|F''(x)\| \leqslant f''(\|x-x_*\|), \qquad x\in\mathcal{B}(x_*,R),$$

then F *and* f *satisfy* (4.3).

Proof. The proof follows the same pattern as outlined in [29, Lemma 22]. ■

Example 4.1.5 Consider the constrained nonlinear least squares problem (1.1) with $C = \mathbb{R}^3_+$ and

$$F(x) = \frac{9}{50}\left(\|x\|^{5/3}x - 64(3,2,\sqrt{3})\right).$$

Note that $x_* = 2(3,2,\sqrt{3})$ is a stationary point of (1.1) in this case. Let us apply Theorem 4.1.2 for this instance. First, from (4.2), we have $c = 0$, $\beta = (25/1152)\sqrt{137}$, $\kappa = (48/25)\beta\sqrt{82}$. Moreover, since the second derivative of F is given by

$$F''(x)(v,v) = \frac{9}{50}\left[-\frac{5}{9}\|x\|^{-7/3}\langle x,v\rangle^2 x + \frac{5}{3}\|x\|^{-1/3}\|v\|^2 x + \frac{10}{3}\|x\|^{-1/3}\langle x,v\rangle v\right],$$

for every $x, v \in \mathbb{R}^3$ and $x \neq 0$, and $F''(0) = 0$, we obtain

$$\|F''(x)\| \leq \|x\|^{2/3}, \quad x \in \mathbb{R}^3,$$

or, equivalently,

$$\beta\|F''(x)\| \leqslant f''(\|x - x_*\|), \qquad x \in \mathbb{R}^3,$$

where $f : [0,\infty) \to \mathbb{R}$ is given by

$$f(t) = \frac{9}{40}\beta t^{8/3} - t.$$

Hence, it follows from Lemma 4.1.4 that F and f satisfy (4.3). In particular, as $f(0) = 0$, $f'(t) = (3\beta/5)t^{5/3} - 1$, $f'(0) = -1$ and $f''(t) = \beta t^{2/3} > 0$, we obtain that f satisfies **h1** and **h2**. Therefore, if $\bar{\theta} < [1/(1+\kappa)] \approx 0.2$ (i.e., **h3** holds), it follows from Theorem 4.1.2 that GNM-AP with starting point $x_0 \in \mathbb{R}^3_+ \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$, where

$$r = \left[\frac{5(15\kappa + 48 - 24\bar{\theta}(1+\kappa)) - 5\sqrt{(24\bar{\theta}(1+\kappa) - 15\kappa - 48)^2 + 864(\bar{\theta}(1+\kappa) - 1)}}{54\beta}\right]^{3/5}, \tag{4.7}$$

is well-defined, the generated $\{x_k\}$ is contained in $\mathcal{B}(x_*, r)\cap\mathbb{R}^3_+$, converges to x_* and satisfies

$$\|x_{k+1} - x_*\| \leq \frac{5}{8(1-\bar{\theta})}\left[\frac{9\beta^2\sigma(x_0)^{7/3} + 15\beta\kappa\sigma(x_0)^{2/3}}{9\beta^2\sigma(x_0)^{10/3} + 30\beta\sigma(x_0)^{5/3} + 5}\right]\|x_k - x_*\|^2, \quad k = 0, 1, \ldots.$$

Note that, if $\bar{\theta} = 0.1$, then the radius of convergence r in (4.7) is approximately equal to 1.

We next specialize Theorem 4.1.2 for two important classes of functions. In the first one, F' satisfies a Lipschitz-like condition [41, 42, 48] and, in the second one, F is an analytic function satisfying a Smale condition [71, 72].

Corollary 4.1.6 *Suppose that there exists a $\mathcal{L} > 0$ such that*

$$\lambda = \frac{[(1+\sqrt{2})\kappa + 1]\, c\, \beta\mathcal{L} + \kappa\bar{\theta}}{(1-\bar{\theta})} < 1, \quad \beta\left\|F'(x) - F'(x_* + \tau(x - x_*))\right\| \leq \mathcal{L}(1-\tau)\sigma(x), \tag{4.8}$$

where $x \in \mathcal{B}(x_*, \delta)$, $\tau \in [0,1]$ *and* $\sigma(x) = \|x - x_*\|$. *Let be given the positive constant*

$$r := \min\left\{ \frac{\mu - \sqrt{\mu^2 - 8(1-\lambda)(1-\bar{\theta})}}{2\mathcal{L}},\ \delta \right\}.$$

where $\mu := 4 + \kappa - 2\bar{\theta}(1+\kappa) + 2c(1+\sqrt{2})\beta\mathcal{L}$. *Then GNM-AP with starting point* $x_0 \in C \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$ *is well-defined, the generated* $\{x_k\}$ *is contained in* $\mathcal{B}(x_*, r) \cap C$, *converges to* x_* *and satisfies*

$$\|x_{k+1} - x_*\| < \|x_k - x_*\| \tag{4.9}$$

and

$$\|x_{k+1} - x_*\| \leq \frac{\kappa\mathcal{L} + \mathcal{L}^2\sigma(x_0)}{2(1-\theta_k)[1-\mathcal{L}\sigma(x_0)]^2}\|x_k - x_*\|^2 + \frac{\theta_k(\mathcal{L}\sigma(x_0) + k)}{(1-\theta_k)[1-\mathcal{L}\sigma(x_0)]}\|x_k - x_*\|$$
$$+ \frac{[(1+\sqrt{2})\kappa + 1]c\beta\mathcal{L} + c(1+\sqrt{2})\beta\mathcal{L}^2\sigma(x_0)}{(1-\theta_k)[1-\mathcal{L}\sigma(x_0)]^2}\|x_k - x_*\|, \quad \forall\ k = 0, 1, \ldots.$$

Proof. It is immediate to prove that F, x_* and $f : [0, \delta) \to \mathbb{R}$ defined by $f(t) = \mathcal{L}t^2/2 - t$, satisfy inequality (4.3), conditions **h1** and **h2**. Since $[(1+\sqrt{2})\kappa + 1]\, c\, \beta\mathcal{L} + \kappa\bar{\theta} < 1 - \bar{\theta}$, the condition **h3** also holds. In this case, it is easy to see that the constants ν and ρ as defined in Theorem 4.1.2, satisfy

$$0 < \rho = \frac{\mu - \sqrt{\mu^2 - 8(1-\bar{\theta})(1-\lambda)}}{2\mathcal{L}} \leq \nu = 1/\mathcal{L}.$$

As a consequence, $0 < r = \min\{\delta, \rho\}$. Therefore, as F, r, f and x_* satisfy all of the hypotheses of Theorem 4.1.2, taking $x_0 \in C \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$ the statements of the corollary follow from Theorem 4.1.2. ■

We next specialize Theorem 4.1.2 for the class of analytic functions satisfying a Smale condition.

Corollary 4.1.7 *Suppose that*

$$\gamma := \sup_{n>1} \beta \left\| \frac{F^{(n)}(x_*)}{n!} \right\|^{1/(n-1)} < +\infty \quad \textit{and} \quad 2\gamma c\beta((1+\sqrt{2})\kappa + 1) + \kappa\bar{\theta} < 1 - \bar{\theta}.$$

Let constants $a = \gamma c\beta$, $b = (1+\sqrt{2})\gamma c\beta$,

$$\bar{\rho} := \inf\left\{ s \in (\sqrt{2}/2, 1) : p(s) := \zeta s^4 + \eta s^3 + \iota s^2 + (b-1)s + b < 0 \right\} \tag{4.10}$$

where $\zeta := -4 + (\kappa+1)2\bar{\theta}$, $\eta := 1 - \kappa + a + b(\kappa - 1)$, *and* $\iota := 3 + \kappa - (\kappa+1)\bar{\theta} + a + b(\kappa - 1)$, *and*

$$r := \min\left\{(1-\bar{\rho})/\gamma,\ \delta\right\}.$$

Then GNM-AP with starting point $x_0 \in C \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$ is well-defined, the generated $\{x_k\}$ is contained in $\mathcal{B}(x_*, r) \cap C$, converges to x_* and satisfies

$$\|x_{k+1} - x_*\| < \|x_k - x_*\| \tag{4.11}$$

and

$$\begin{aligned}\|x_{k+1} - x_*\| \leq & \frac{\gamma\left[1 + (\kappa - 1)(1 - \gamma\sigma(x_0))^2\right]}{(1 - \theta_k)\left[1 - 2(1 - \gamma\sigma(x_0))^2\right]^2}\|x_k - x_*\|^2 \\ + & \left[\frac{\left[(1 + \sqrt{2})c\beta(1 - (1 - \gamma\sigma(x_0))^2) - \theta_k\sigma(x_0)(1 - 2(1 - \gamma\sigma(x_0))^2)\right]\left[1 + (\kappa - 1)(1 - \gamma\sigma(x_0))^2\right]}{(1 - \theta_k)\sigma(x_0)\left[1 - 2(1 - \gamma\sigma(x_0))^2\right]^2}\right. \\ & \left. + \frac{c\beta\left[1 - (1 - \gamma\sigma(x_0))^2\right](1 - \gamma\sigma(x_0))^2}{(1 - \theta_k)\sigma(x_0)\left[1 - 2(1 - \gamma\sigma(x_0))^2\right]^2}\right]\|x_k - x_*\|, \end{aligned} \tag{4.12}$$

for all $k = 0, 1, \ldots$.

Proof. Consider the real function $f : [0, 1/\gamma) \to \mathbb{R}$ defined by

$$f(t) = \frac{t}{1 - \gamma t} - 2t.$$

It is straightforward to show that f is analytic and that

$$f(0) = 0, f'(t) = 1/(1 - \gamma t)^2 - 2, f'(0) = -1, f''(t) = (2\gamma)/(1 - \gamma t)^3, f^n(0) = n!\,\gamma^{n-1},$$

for $n \geq 2$. It follows from the last equalities that f satisfies **h1** and **h2**. Since

$$2\gamma c\beta((1 + \sqrt{2})\kappa + 1) + \kappa\bar{\theta} < 1 - \bar{\theta},$$

condition **h3** also holds. Now, note that

$$\beta\|F''(x)\| \leqslant f''(\|x - x_*\|),$$

for all $x \in \mathcal{B}(x_*, 1/\gamma) \cap C$, the proof of this fact follows the same pattern as outlined in [29, Lemma 21]. As $f''(t) = (2\gamma)/(1 - \gamma t)^3$, we conclude, from Lemma 4.1.4, that F and f satisfy (4.3) with $R = 1/\gamma$. In this case,

$$\nu = (2 - \sqrt{2})/2\gamma < 1/\gamma.$$

Now, we will obtain the constant ρ as defined in Theorem 4.1.2. For simplicity, consider the following change of variable $s = 1 - \gamma t$, which implies that $t = (1 - s)/\gamma$. Moreover, if t satisfies $0 < t < \nu = (2 - \sqrt{2})2\gamma$, then $\sqrt{2}/2 < s < 1$. Hence, to determine the constant ρ as defined in Theorem 4.1.2 is equivalent to determine the constant s such that

$$\bar{\rho} := \inf\left\{s \in (\sqrt{2}/2, 1) : p(s) := \zeta s^4 + \eta s^3 + \iota s^2 + (b - 1)s + b < 0\right\},$$

where $a = \gamma c\beta$, $b = (1+\sqrt{2})\gamma c\beta$, $\zeta := -4 + (\kappa+1)2\bar{\theta}$, $\eta := 1 - \kappa + a + b(\kappa - 1)$, and $\iota := 3 + \kappa - (\kappa+1)\bar{\theta} + a + b(\kappa-1)$. Thus, taking into account the change of variable, we have $\rho = (1-\bar{\rho})/\gamma$ and

$$r = \min\left\{(1-\bar{\rho})/\gamma,\, \delta\right\}.$$

Thus, as F, r, f and x_* satisfy all hypothesis of Theorem 4.1.2, taking $x_0 \in C \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$, the statements of the corollary follow from Theorem 4.1.2. ■

We end this section by presenting a numerical example, adapted from Dedieu and Shub [22], in which all conditions of Corollary 4.1.6 hold.

Example 4.1.8 Let $F : \mathbb{R} \to \mathbb{R}^2$ such that $F(x) = (x, x^2 + a)^T$, where $a \in \mathbb{R}$ is given, and consider

$$\min_{x\in[-2,2]} \|F(x)\|^2 = x^4 + (2a+1)x^2 + a^2. \tag{4.13}$$

Note that $x_* = 0$ is a stationary point of (4.13). Let us apply Corollary 4.1.6 for this instance. First, from (4.2), we have $c = |a|$, $\beta = 1$, $\kappa = 1$. Moreover, since $\beta\|F'(x) - F'(\tau x)\| = (1-\tau)2|x|$, for all $x \in \mathbb{R}$ and $\tau \in [0,1]$, we obtain the Lipschitz-Like constant $\mathcal{L}$ is 2. Therefore, if $2[(2+\sqrt{2})|a| + \bar{\theta}] < 1$ (i.e., the first inequality in (4.8) holds), it follows from Corollary 4.1.6 that GNM-AP with starting point $x_0 \in [-2,2] \cap \mathcal{B}(x_*, r)\backslash\{x_*\}$, where

$$r = \frac{5 - 4\bar{\theta} + 4|a|(1+\sqrt{2}) - \sqrt{\left[5 - 4\bar{\theta} + 4|a|(1+\sqrt{2})\right]^2 - 8(1 - 2\bar{\theta} - 2(2+\sqrt{2})|a|)}}{4}, \tag{4.14}$$

is well-defined, the generated $\{x_k\}$ is contained in $\mathcal{B}(x_*, r) \cap [-2,2]$, converges to x_* and satisfies

$$\begin{aligned}\|x_{k+1} - x_*\| \leq{}& \frac{1 + 2\sigma(x_0)}{(1-\bar{\theta})[1-2\sigma(x_0)]^2}\|x_k - x_*\|^2 + \frac{\bar{\theta}(1+2\sigma(x_0))}{(1-\bar{\theta})[1-2\sigma(x_0)]}\|x_k - x_*\| \\ &+ \frac{2|a|(2+\sqrt{2}+2(1+\sqrt{2})\sigma(x_0))}{(1-\bar{\theta})[1-2\sigma(x_0)]^2}\|x_k - x_*\|, \quad \forall\, k = 0, 1, \ldots.\end{aligned}$$

Note that, if $a = 0$ and $\bar{\theta} = 0.1$, then the radius of convergence r in (4.14) is approximately equal to 0.2.

4.1.1 Proof of Theorem 4.1.2

Our goal in this subsection is to prove Theorem 4.1.2. To this end, we first present some auxiliary results, which establish positiveness of the constants δ, ν and ρ, as well as some useful relationships between the majorant function and the nonlinear function F.

First of all, note that constant δ in (4.2) is positive, because $\mathbb{U}$ is an open set and $x_* \in \mathbb{U}$.

Proposition 4.1.9 *The constant ν as in Theorem 4.1.2 is positive and $f'(t) < 0$ for all $t \in (0, \nu)$. Furthermore, the following functions defined on the interval $(0,\, \nu)$*

$$t \mapsto -\frac{1}{f'(t)}, \quad t \mapsto -\frac{[f'(t)+1+\kappa]}{f'(t)}, \quad t \mapsto \frac{[tf'(t)-f(t)]}{t^2}, \quad t \mapsto \frac{f'(t)+1}{t}, \tag{4.15}$$

are positive and increasing.

Proof. First, as f' is continuous in $(0, R)$ and $f'(0) = -1$, there exists $\epsilon > 0$ such that $f'(t) < 0$ for all $t \in (0, \epsilon)$. Hence, $\nu > 0$. Moreover, using **h2** and the definition of ν, it follows that $f'(t) < 0$ for all $t \in (0, \nu)$. Note now that the first two functions in (4.15) are positive and increasing due to the facts that $-1 < f'(t) < 0$, for all $t \in [0, \nu)$, and f' is strictly increasing. Finally, for the proofs that the last two functions in (4.15) are positive and increasing, see items **ii** and **iii** of [29, Proposition 10]. ■

We next prove, in particular, that constant ρ in (4.4) is positive.

Proposition 4.1.10 *The constant ρ is positive and there holds*

$$\frac{[f'(t)+1+\kappa]\left[(1-\bar\theta)tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2} < 1, \tag{4.16}$$

for all $t \in (0,\, \rho)$.

Proof. Using **h1** and some algebraic manipulation, we obtain

$$\frac{tf'(t)-f(t)}{t} = \left[f'(t) - \frac{f(t)-f(0)}{t-0}\right], \qquad \frac{f'(t)+1}{t} = \frac{f'(t)-f'(0)}{t-0},$$

which, combined with the fact that $f'(0) = -1$, yields

$$\lim_{t\to 0}[tf'(t)-f(t)]/t = 0, \qquad \lim_{t\to 0}[f'(t)+1]/t = D^+f'(0), \tag{4.17}$$

where the existence of the right derivative $D^+f'(0)$ is guaranteed due to the fact that f' is convex. Note now that equation (4.16) is equivalent to

$$\begin{aligned}&\frac{[f'(t)+1+\kappa]\,[tf'(t)-f(t)]}{(1-\bar\theta)[tf'(t)]^2}t - \frac{\bar\theta\,[f'(t)+1+\kappa]}{(1-\bar\theta)f'(t)} + \frac{c\beta(1+\sqrt{2})\,[f'(t)+1+\kappa]\,(f'(t)+1)}{(1-\bar\theta)t[f'(t)]^2}\\ &+\frac{c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2}.\end{aligned} \tag{4.18}$$

Hence, using $f'(0) = -1$, it follows from (4.18) and (4.17) that

$$\begin{aligned}&\lim_{t\to 0}\left[\frac{[f'(t)+1+\kappa]\left[(1-\bar\theta)tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta\,[f'(t)+1]}{(1-\bar\theta)t[f'(t)]^2}\right]\\ &\quad= \frac{\kappa\bar\theta + c\beta(1+\sqrt{2})\kappa D^+f'(0) + c\beta D^+f'(0)}{(1-\bar\theta)} = \frac{c\beta\left[(1+\sqrt{2})\kappa+1\right]D^+f'(0)+\kappa\bar\theta}{(1-\bar\theta)}.\end{aligned}$$

Therefore, since **h3** implies that $[c\beta[(1+\sqrt{2})\kappa+1]D^+f'(0)+\kappa\bar{\theta}]/[(1-\bar{\theta})] < 1$, we conclude that there exists an $\varepsilon > 0$ such that

$$\frac{[f'(t)+1+\kappa]\left[(1-\bar{\theta})tf'(t)-f(t)+c\beta(1+\sqrt{2})(f'(t)+1)\right]+c\beta\left[f'(t)+1\right]}{(1-\bar{\theta})t[f'(t)]^2} < 1,$$

for all $t \in (0,\varepsilon)$. So, $\varepsilon \leq \rho$, which proves the first statement.

Again, since (4.16) is equivalent to (4.18), the proof of the last part of proposition trivially follows from definition of ρ and last part of Proposition 4.1.9. ■

The next two lemmas present some useful relationships between operator F and majorant function f.

Lemma 4.1.11 *Let $x \in \mathbb{U}$. If $\sigma(x) < \min\{\nu,\delta\}$, then following statements hold:*

- **i)** $\beta\|F(x_*) - [F(x)+F'(x)(x_*-x)]\| \leq f(0) - [f(\sigma(x))+f'(\sigma(x))(0-\sigma(x))] := e_f(\sigma(x),0)$;
- **ii)** $B(x) = F'(x)^TF'(x)$ *is invertible and*

$$\|F'(x)^\dagger\| \leq \frac{-\beta}{f'(\sigma(x))}, \qquad \|F'(x)^\dagger - F'(x_*)^\dagger\| < \frac{-\sqrt{2}\beta[f'(\sigma(x))+1]}{f'(\sigma(x))}.$$

In particular, $B(x) = F'(x)^TF'(x)$ is invertible in $\mathcal{B}(x_*,r)$.

Proof. The proof follows the pattern of the proofs of Lemmas 13 and 14 in [29] (see also Lemma 7 in [42]). ■

Lemma 4.1.12 *Let $x \in \mathbb{U}$. If $\sigma(x) < \min\{\nu,\delta\}$, then the following inequalities hold:*

- **i)** $\|B(x)\|^{1/2} \leq [f'(\sigma(x))+1+\kappa]/\beta$;
- **ii)** $\|B(x)^{-1}\|^{1/2} \leq -\beta/[f'(\sigma(x))]$;
- **iii)** $\beta\|(B(x)-B(x_*))F'(x_*)^\dagger\| \leq (f'(\sigma(x))+2+\kappa)(f'(\sigma(x))+1)$,

*where $B(x)$ is defined as in Lemma 4.1.11(***ii***).*

Proof. (i) Using inequality in (4.3) and definition of κ in (4.2), we have

$$\beta\|F'(x)\| \leq \beta\|F'(x)-F'(x_*)\| + \beta\|F'(x_*)\| \leq f'(\sigma(x))+1+\kappa. \tag{4.19}$$

As $\|B(x)\|^{1/2} = \|F'(x)^TF'(x)\|^{1/2} = \|F'(x)\|$, the desired inequality follows.

(ii) To show item **ii**, use the definition of B, the last equality in (2.3) and Lemma 4.1.11(ii).

(iii) Note that the definition of $B(x)$, some algebraic manipulations and (2.3) gives

$$\begin{aligned}
&\beta\|(B(x)-B(x_*))F'(x_*)^\dagger\| \\
&= \beta\|F'(x)^T(F'(x)-F'(x_*))F'(x_*)^\dagger + (F'(x)-F'(x_*))^T\| \\
&\leq (\|F'(x)\|\|F'(x_*)^\dagger\|+1)\beta\|F'(x)-F'(x_*)\|,
\end{aligned}$$

which, combined with definition of β in (4.2) and inequalities in (4.3) and (4.19), yields the desired inequality. ■

Lemma 4.1.11 implies that B is invertible for any $x \in \mathcal{B}(x_*, r)$ and hence $F'(x)^\dagger$ and $\tilde{y}_C^B(w)$, characterized as an approximate projection of a step $y = x - F'(x)^\dagger F(x)$, are well-defined in this region. Therefore, since the starting point $x_0 \in C \cap \mathcal{B}(x_*, r)$, we have x_1 is well-defined, but we do not show that $x_1 \in C \cap \mathcal{B}(x_*, r)$ and, therefore, if the next iteration x_2 will be well-defined. In the next lemma, we ensure that sequence $\{\|x_k - x_*\|\}$ is strictly decreasing and, hence, that $\{x_k\}$ is well-defined and contained in $C \cap \mathcal{B}(x_*, r)$.

Lemma 4.1.13 *Let $x_k \in C \cap \mathcal{B}(x_*, r)$. Then, for every $k \geq 0$,*

$$\begin{aligned}
\|x_{k+1}-x_*\| \leq{}& \frac{[f'(\sigma(x_k))+1+\kappa]\,[\sigma(x_k)f'(\sigma(x_k))-f(\sigma(x_k))]}{(1-\theta_k)\,[\sigma(x_k)f'(\sigma(x_k))]^2}\|x_k-x_*\|^2 \\
&+ \frac{[f'(\sigma(x_k))+1+\kappa]\left[(1+\sqrt{2})c\beta\,[f'(\sigma(x_k))+1]-\theta_k\sigma(x_k)f'(\sigma(x_k))\right]}{(1-\theta_k)\sigma(x_k)\,[f'(\sigma(x_k))]^2}\|x_k-x_*\| \\
&+ \frac{c\beta\,[f'(\sigma(x_k))+1]}{(1-\theta_k)\sigma(x_k)\,[f'(\sigma(x_k))]^2}\|x_k-x_*\|. \quad (4.20)
\end{aligned}$$

As a consequence,

$$\|x_{k+1}-x_*\| < \|x_k-x_*\|. \quad (4.21)$$

Proof. Since x_* is a stationary point of (1.1) (see **(A1)**), we trivially have

$$y_C^{B_*}(w_*) = x_*.$$

Hence, it follows from Lemma 2.2.4 with $B = B_k$, $x = x_k - F'(x_k)^\dagger F(x_k)$, $\hat{x} = x_* - F'(x_*)^\dagger F(x_*)$ and $\varepsilon = \theta_k^2\|x_k - x_{k+1}\|_{B_k}^2$ that

$$\begin{aligned}
&\|\tilde{y}_C^{B_k}(w_k) - x_*\|_{B_k} \leq \|B_k^{-1}\|^{1/2}\|(B_*-B_k)(F'(x_*)^\dagger F(x_*))\| \\
&+ \|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\|_{B_k} + \theta_k\|x_k - x_{k+1}\|_{B_k}.
\end{aligned}$$

For simplicity, the notation defines the following terms:

$$A(x_k, x_*) = \|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\|_{B_k} \quad (4.22)$$

and

$$\bar{A}(x_k, x_*) = \|B_k^{-1}\|^{1/2}\|(B_k - B_*)F'(x_*)^\dagger\|\|F(x_*)\|. \tag{4.23}$$

So, from the three latter inequalities, we obtain

$$\|x_{k+1} - x_*\|_{B_k} \leq A(x_k, x^*) + \bar{A}(x_k, x^*) + \theta_k\|x_k - x_{k+1}\|_{B_k}.$$

Hence, since $\|x_k - x_{k+1}\|_{B_k} \leq \|x_{k+1} - x_*\|_{B_k} + \|B_k\|^{1/2}\|x_k - x_*\|$, we obtain

$$(1-\theta_k)\|x_{k+1} - x_*\|_{B_k} \leq A(x_k, x^*) + \bar{A}(x_k, x^*) + \theta_k\|B_k\|^{1/2}\|x_k - x_*\|.$$

Since $\theta_k < 1$, for all $k \geq 0$, (see **(A2)**), the last inequality and (2.2) imply that

$$\begin{aligned}\|x_{k+1} - x_*\| \leq & \frac{\|B_k^{-1}\|^{1/2}}{(1-\theta_k)}A(x_k, x^*) + \frac{\|B_k^{-1}\|^{1/2}}{(1-\theta_k)}\bar{A}(x_k, x^*) \\ & + \frac{\theta_k\left[\|B_k^{-1}\|\|B_k\|\right]^{1/2}}{(1-\theta_k)}\|x_k - x_*\|. \end{aligned} \tag{4.24}$$

Now, we will obtain upper bounds of $A(x_k, x^*)$ and $\bar{A}(x_k, x^*)$. First, some algebraic manipulations and the second equality in (2.3) yield

$$\begin{aligned}&\|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\| \\ =&\|F'(x_k)^\dagger[F'(x_k)(x_k - x_*) - F(x_k) + F(x_*)] + (F'(x_*)^\dagger - F'(x_k)^\dagger)F(x_*)\| \\ \leq&\|F'(x_k)^\dagger\|\|F(x_*) - [F(x_k) + F'(x_k)(x_* - x_k)]\| + \|F'(x_*)^\dagger - F'(x_k)^\dagger\|\|F(x_*)\|.\end{aligned}$$

Combining last inequality, Lemma 4.1.11 and definition of c in (4.2), we have

$$\|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\| = \frac{e_f(\sigma(x_k), 0)}{-f'(\sigma(x_k))} + \frac{\sqrt{2}c\beta[f'(\sigma(x_k)) + 1]}{-f'(\sigma(x_k))},$$

which, combined with (4.22), the fact that $\|\cdot\|_{B_k} \leq \|B_k\|^{1/2}\|\cdot\|$ and Lemma 4.1.12(i), yields

$$\begin{aligned}A(x_k, x^*) &\leq \|B_k\|^{1/2}\|x_k - F'(x_k)^\dagger F(x_k) - x_* + F'(x_*)^\dagger F(x_*)\| \\ &\leq \frac{(f'(\sigma(x_k)) + 1 + \kappa)}{-\beta f'(\sigma(x_k))}\Big(e_f(\sigma(x_k), 0) + \sqrt{2}c\beta[f'(\sigma(x_k)) + 1]\Big).\end{aligned} \tag{4.25}$$

On the other hand, from definition in (4.23) and Lemma 4.1.12(ii)–(iii), we have

$$\bar{A}(x_k, x_*) \leq \frac{c}{-f'(\sigma(x_k))}(f'(\sigma(x_k)) + 2 + \kappa)(f'(\sigma(x_k)) + 1). \tag{4.26}$$

Hence, using (4.24)–(4.26) and Lemma 4.1.12(i)–(ii), we obtain

$$\begin{aligned}\|x_{k+1} - x_*\| \leq & \frac{[f'(\sigma(x_k)) + 1 + \kappa]\, e_f(\sigma(x_k), 0) + (1+\sqrt{2})c\beta\,[f'(\sigma(x_k)) + 1]^2}{(1-\theta_k)\,[f'(\sigma(x_k))]^2} \\ & + \frac{c\beta\left[(1+\sqrt{2})\kappa + 1\right][f'(\sigma(x_k)) + 1]}{(1-\theta_k)\,[f'(\sigma(x_k))]^2} - \frac{\theta_k(f'(\sigma(x_k)) + 1 + \kappa)}{(1-\theta_k)f'(\sigma(x_k))}\sigma(x_k),\end{aligned}$$

which, combined with definition of $e_f(\sigma(x_k), 0)$ in Lemmas 4.1.11(i) and **h1**, proves (4.20).

Now, using $\theta_k < \bar{\theta}$, for all $k \geq 0$, (see **(A2)**), we obtain that the right-hand side of (4.20) is equivalent to

$$\left[\frac{[f'(\sigma(x_k)) + 1 + \kappa]\left[(1-\bar{\theta})\sigma(x_k)f'(\sigma(x_k)) - f(\sigma(x_k)) + c\beta(1+\sqrt{2})(f'(\sigma(x_k))+1)\right]}{(1-\bar{\theta})\sigma(x_k)\left[f'(\sigma(x_k))\right]^2} + \frac{c\beta\left[f'(\sigma(x_k))+1\right]}{(1-\bar{\theta})\sigma(x_k)\left[f'(\sigma(x_k))\right]^2}\right]\sigma(x_k).$$

Therefore, as $x_k \in C \cap \mathcal{B}(x_*, r)/\{x_*\}$, it follows from Proposition 4.1.10 with $t = \sigma(x_k)$ that the quantity in the bracket above is less than one and hence (4.21) follows. ■

Proof of Theorem 4.1.2: Since $x_0 \in C \cap \mathcal{B}(x_*, r)/\{x_*\}$, combining Lemma 4.1.11(ii), inequality (4.21) and an induction argument, we have that (4.5) holds and $\{x_k\}$ is well-defined and remains in $C \cap \mathcal{B}(x_*, r)$. Our goal is now to show that $\{x_k\}$ converges to x_*. Using the second part of Lemma 4.1.13, we find

$$\sigma(x_k) = \|x_k - x_*\| < \|x_0 - x_*\| = \sigma(x_0), \qquad k = 1, 2 \ldots. \tag{4.27}$$

Hence, by combining (4.20) with last part of Proposition 4.1.9, we obtain

$$\begin{aligned}\|x_{k+1} - x_*\| \leq{}& \frac{[f'(\sigma(x_0)) + 1 + \kappa]\left[\sigma(x_0)f'(\sigma(x_0)) - f(\sigma(x_0))\right]}{(1-\theta_k)[\sigma(x_0)f'(\sigma(x_0))]^2}\|x_k - x_*\|^2 \\ &+ \frac{(1+\sqrt{2})c\beta\left[f'(\sigma(x_0)) + 1 + \kappa\right]\left[f'(\sigma(x_0))+1\right] + c\beta\left[f'(\sigma(x_0))+1\right]}{(1-\theta_k)\sigma(x_0)\left[f'(\sigma(x_0))\right]^2}\|x_k - x_*\| \\ &- \frac{\theta_k(f'(\sigma(x_0)) + 1 + \kappa)}{(1-\theta_k)f'(\sigma(x_0))}\|x_k - x_*\|, \quad k = 0, 1, \ldots,\end{aligned}$$

which is equivalent to (4.6). Combining last inequality with (4.27) and **(A2)**, we obtain

$$\begin{aligned}&\|x_{k+1} - x_*\| \leq \\ &\left[\frac{[f'(\sigma(x_0)) + 1 + \kappa]\left[(1-\bar{\theta})\sigma(x_0)f'(\sigma(x_0)) - f(\sigma(x_0)) + c\beta(1+\sqrt{2})(f'(\sigma(x_0))+1)\right]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2}\right. \\ &\left. + \frac{c\beta\left[f'(\sigma(x_0))+1\right]}{(1-\bar{\theta})\sigma(x_0)[f'(\sigma(x_0))]^2}\right]\|x_k - x_*\|,\end{aligned}$$

for all $k = 0, 1, \ldots$. Hence, applying Proposition 4.1.10 with $t = \sigma(x_0)$, we conclude that $\{\|x_k - x_*\|\}$ converges to zero. So, $\{x_k\}$ converges to x_*. ■

4.2 Globalized method

We now present a globalized version of GNM-AP. The globalization strategy used here is based on the nonmonotone line search in [44]. Since the Gauss-Newton step can not be

well-defined in some regions, our global method uses, in these cases, the projected gradient step.

The method is formally described as follows.

Global GNM-AP (G-GNM-AP)

Step 0 (Initialization). Let $x_0 \in C$, $\tau \in (0,1)$, an integer $M \geq 1$ and $\{\theta_k\} \subset [0,\infty)$ be given, and set $k = 0$.

Step 1 (projected Gauss-Newton or projected gradient step). If $F'(x_k)^T F'(x_k)$ is non-singular, then $B_k = F'(x_k)^T F'(x_k)$. Otherwise, $B_k = I_n$. Compute $w_k = B_k x_k - F'(x_k)^T F(x_k) \in \mathbb{R}^n$ and $\tilde{y}_C^{B_k}(w_k) \in C$ such that

$$\langle w_k - B_k \tilde{y}_C^{B_k}(w_k), y - \tilde{y}_C^{B_k}(w_k)\rangle \leq \varepsilon_k := \theta_k^2 \|\tilde{y}_C^{B_k}(w_k) - x_k\|_{B_k}^2, \quad \forall\, y \in C, \tag{4.28}$$

i.e., $\tilde{y}_C^{B_k}(w_k)$ is an ε_k–approximate solution of (2.4).

Step 2 (Backtracking). Define $d_k = \tilde{y}_C^{B_k}(w_k) - x_k$ and $f_{max} = \max\{f(x_{k-j});\ 0 \leq j \leq \min\{k, M-1\}\}$. Set $\alpha \leftarrow 1$.

Step 2.1 Set $x_+ = x_k + \alpha d_k$.

Step 2.2 If

$$f(x_+) \leq f_{max} + \tau\alpha\langle F'(x_k)^T F(x_k), d_k\rangle, \tag{4.29}$$

then $\alpha_k = \alpha$, $x_{k+1} = x_+$, and go to Step 3. Otherwise, set $\alpha \leftarrow \alpha/2$ and go to Step 2.2.

Step 3 (Termination criterion and update). If $x_{k+1} = x_k$, then **stop**; otherwise, set $k \leftarrow k+1$ and go to Step 1.

end

The following theorem, which is also an extension to the constrained case of [44, Theorem 1], summarizes the convergence properties of the G-GNM-AP method.

Theorem 4.2.1 *Assume that $B_k \in \mathbb{B}$. Furthermore, assume that level set $C_0 := \{x \in C : f(x) \leq f(x_0)\}$ is bounded and sequence $\{\theta_k\}$ satisfies $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0,1)$. Then, either G-GNM-AP stops at some stationary point x_k or every limit point of the generated sequence is stationary.*

Proof. By definitions of d_k and w_k, and the inequality in (4.28), we have

$$\langle -d_k - B_k^{-1}F'(x_k)^T F(x_k), y - x_k - d_k\rangle_{B_k} \leq \theta_k^2 \|d_k\|_{B_k}^2, \quad \forall\, k \geq 0. \tag{4.30}$$

If G-GNM-AP stops, then $x_{k+1} = x_k$, which in turn implies that $d_k = 0$. Hence, it follows from (4.30) that

$$\langle -B_k^{-1}F'(x_k)^T F(x_k), y - x_k\rangle_{B_k} \leq 0, \quad \forall\, y \in C,$$

or, equivalently,

$$\langle F'(x_k)^T F(x_k), y - x_k\rangle \geq 0, \quad \forall\, y \in C,$$

i.e., x_k is a stationary point of (1.1). Now, under the assumption $B_k \in \mathbb{B}$, for all $k \geq 0$, $C_0 := \{x \in C : f(x) \leq f(x_0)\}$ is bounded and that $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, where $\bar{\theta} \in [0, 1)$, we conclude, from Theorem 3.1.3, that every limit point of $\{x_k\}$ is a stationary point of (1.8).

■

4.3 Numerical experiments

This section summarizes the results of the numerical experiments we carried out in order to verify the effectiveness of GNM-AP and G-GNM-AP methods. The algorithms were tested on some box- and polyhedral-constrained nonlinear least squares problems. We took $\theta_k = 1/3$, for every k, in both algorithms. Moreover, the inexactness criterion (4.1) (and (4.28)) was computed by the conditional gradient method, which stopped when either the stopping criterion given in Step 1 was satisfied or a maximum of 300 iterations were performed (in this case we did not stop the outer procedure). In order to avoid an excessive number of inner iterations, input ε_k was replaced by $\max\{\theta_k^2\|x_{k+1} - x_k\|_{B_k}^2, 10^{-2}\}$. Linear optimization subproblems in the conditional gradient method (see (2.11)) were solved via the MATLAB command `linprog`. Other initialization parameters of G-GNM-AP method were set $\tau = 10^{-4}$ and $M = 10$. Nonmonotone parameter $M = 10$ was the best from $\{1, 5, 10, 15, 20, 25\}$ for a preliminary small number of problems.

For a comparison purpose, we also run the proximal Gauss-Newton (Prox-GN) method of [70], applied to (1.8), which corresponds to our GNM-AP method with exact projections (i.e., $\theta_k = 0$ for every k). In the latter method, exact projections were computed by the MATLAB command `quadprog`. In the box-constrained case, we also compare the performance of G-GNM-AP method with the inexact Gauss-Newton trust-region method (ITREBO) of [68]. ITREBO is an algorithm designed for solving nonlinear least-squares problems with simple bounds where, at each iteration, a trust-region subproblem is approximately solved by the Conjugate Gradient method. For GNM-AP, G-GNM-AP and Prox-GN methods, we used the same termination condition $\|x_{k+1} - x_k\|_{B_k} < 10^{-4}$, whereas in ITREBO we used $\|P_C(x_k -$

$\nabla f(x_k)) - x_k\| < 10^{-4}$. For all algorithms, a failure was declared if the number of iterations was greater than 300 or no progress was detected. The computational results were obtained using MATLAB R2016a on a 2.4GHz Intel(R) i5 with 8GB of RAM and Windows 10 ultimate system.

4.3.1 Nonlinear least squares problems with box constraints

In this subsection, our aim is to illustrate the behavior of the algorithms to solve 23 problems of the form (1.8) with $C = \{x \in \mathbb{R}^n;\ c \leq x \leq d\}$, where $c, d \in \mathbb{R}^n$; see Table 4.1. The first four problems were taken from [70]. The others are originally unconstrained problems for which box constrains were added.

We firstly chose 10 initial points of the form $x_0(\gamma) = c + (\gamma/11)(d - c)$ for $\gamma = 1, 2, \ldots, 10$. We report in Figure 4.1 the numerical results of GNM-AP, G-GNM-AP and Prox-GN methods for solving the 23 problems using performance profiles [26]. We adopted the CPU time as performance measurement. It is worth pointing out that 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 the performance profile, efficiency and robustness can be accessed on the left and right extremes of the graphic, respectively. We consider that a method is the most efficient if its runtime does not exceed in 5% the CPU time of the fastest one.

From Figure 4.1, we see that GNM-AP method was more robust and efficient in terms of time than Prox-GN method. This fact illustrates the advantages of allowing inexactness in the calculation of projections. On the other hand, we also see, as expected, that G-GNM-AP method was more robust than the local methods. Its robustness rate was approximately 95%, whereas for GNM-AP (resp. Prox-GN) the robustness rate was approximately 85% (resp. 71%).

Since our schemes and ITREBO use different stopping criteria, in order to provide a fair comparison, we report in Table 4.2 the performance of G-GNM-AP and ITREBO with three initial point of the form $x_0(\gamma) = c + 0.25\gamma(d - c)$, where $\gamma > 0$, for solving the 23 box-constrained nonlinear least squares problems aforementioned. As can be seen, G-GNM-AP and ITREBO successfully ended 60 and 51 times, respectively, on a total of 69 runs. Moreover, G-GNM-AP (resp. ITREBO) was faster in 31 (resp. 14) cases. Therefore, we can say that our global scheme outperformed ITREBO for the instances considered.

Problem	Function($F(x)$) and source	n	m	Box
Pb 1	Rosenbrock [70, Problem 1]	2	2	As [70]
Pb 2	Osborne1 [70, Problem 3]	5	33	As [70]
Pb 3	Osborne2 [70, Problem 4]	11	65	As [70]
Pb 4	Twoeq6 [70, Problem 5]	2	2	As [70]
Pb 5	Freudenstein [63, Problem 2]	2	2	$[1, 5]$
Pb 6	Powell badly scaled [63, Problem 3]	2	2	$[0, 9.106]$
Pb 7	Brown badly scaled [63, Problem 4]	2	3	$[0, 10^6]$
Pb 8	Beale [63, Problem 5]	2	3	$[0, 3]$
Pb 9	Jennrich and Sampson [63, Problem 6]	2	10	$[-2, 1]$
Pb 10	Bard [63, Problem 8]	3	15	$[-10, 1]$
Pb 11	Gaussian [63, Problem 9]	3	15	$[-1, 1.02]$
Pb 12	Box three-dimensional [63, Problem 12]	3	100	$[0, 10]$
Pb 13	Powell singular [63, Problem 13]	4	4	$[-3, 3]$
Pb 14	Biggs EXP6 [63, Problem 18]	6	10	$[-1, 10]$
Pb 15	Penalty I [63, Problem 23]	4	5	$[-10, 1]$
Pb 16	Penalty I [63, Problem 23]	10	11	$[-10, 1]$
Pb 17	Variably dimensioned [63, Problem 25]	100	102	$[-1, 2]$
Pb 18	Variably dimensioned [63, Problem 25]	450	452	$[-1, 2]$
Pb 19	Trigonometric [63, Problem 26]	6	6	$[-2, 3]$
Pb 20	Broyden tridiagonal [63, Problem 30]	10	10	$[-2, 2]$
Pb 21	Broyden tridiagonal [63, Problem 30]	1000	1000	$[-2, 2]$
Pb 22	Example 6.1.10 [31, Chap. 6]	1	2	$[-10, 20]$
Pb 23	Example 10.2.4 [23, Chap. 10]	1	3	$[-2, 1]$

Table 4.1: Test problems

4.3.2 Nonlinear least squares problems with polyhedral constraints

In this subsection, we are interested in solving 23 test problems of the form (1.1) with $C = \{x \in \mathbb{R}^n;\ c \leq x \leq d,\ Ax \leq b\}$, where $c, d \in \mathbb{R}^n,\ b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$. Our test problems are the nonlinear least squares problems with box constraints of Subsection 4.3.1,

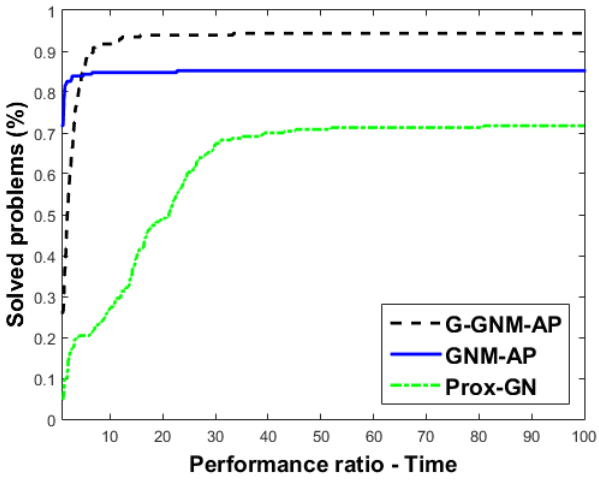

Figure 4.1: Performance of G-GNM-AP, GNM-AP and Prox-GN methods

for which randomly generated constraints $Ax \leq b$ were added. In this application, we considered 5 different initial points belonging to the feasible set C.

As in Subsection 4.3.1, we reported in Figure 4.2 numerical comparisons of the obtained results using performance profiles. Illustrating again the advantages of allowing inexactness in the calculation of projections, we observe, from Figure 4.2, that GNM-AP was more robust and efficient in terms of saving time than Prox-GN method. Moreover, G-GNM-AP was more robust than GNM-AP, which, on the other hand, was more robust than Prox-GN method.

Finally, we conclude that the proposed schemes seems to be promising tools for solving box- and polyhedral-constrained nonlinear least squares problems.

		G-GNM-AP	ITREBO			G-GNM-AP	ITREBO
Pb	γ	it/time/Fnorm	it/time/Fnorm	Pb	γ	it/time/Fnorm	it/time/Fnorm
	1	273/5.7e+0/1.5e-1	*		1	11/1.1e-2/2.7e-5	8/1.5e-2/3.8e-4
Pb 1	2	6/4.1e-3/1.3e-1	*	Pb 13	2.5	10/7.9e-3/3.5e-5	7/1.1e-2/6.0e-4
	3	5/3.9e-3/1.3e-1	*		3	11/1.2e-2/4.6e-5	8/1.4e-2/3.8e-4
	1	12/9.1e-2/9.0e-2	*		1	186/7.8e-1/4.4e-1	7/1.5e-2/5.4e-1
Pb 2	2	13/1.0e-1/8.8e-2	*	Pb 14	2	195/6.4e-1/4.4e-1	*
	3	12/9.6e-2/9.0e-2	*		3	31/1.0e-1/4.2e-1	*
	1	7/2.9e-2/6.8e-1	*		1	9/6.8e-3/7.9e-3	9/1.2e-2/7.9e-3
Pb 3	2	8/3.2e-2/6.8e-1	*	Pb 15	2	8/3.6e-3/7.9e-3	8/9.7e-3/7.9e-3
	3	11/4.6e-2/6.8e-1	*		3	7/3.2e-3/7.9e-3	6/8.6e-3/7.9e-3
	1	11/6.0e-3/7.1e-5	*		1	9/6.1e-3/1.1e-2	11/1.4e-2/1.1e-2
Pb 4	2	12/6.5e-3/7.1e-5	*	Pb 16	2	9/5.2e-3/1.1e-2	9/1.3e-2/1.1e-2
	3	16/1.1e-2/1.0e-5	*		3	7/4.4e-3/1.1e-2	7/7.2e-3/1.1e-2
	1	6/3.5e-3/3.5e-10	7/9.0e-3/1.2e-7		1	17/4.0e-2/9.1e-6	18/4.7e-2/4.8e-9
Pb 5	2	6/2.8e-3/2.6e-10	5/6.8e-3/5.3e-8	Pb 17	2	16/3.5e-2/7.7e-8	16/4.2e-2/8.6e-12
	3	2/1.9e-3/0.0e+0	3/5.0e-3/1.8e-7		3	15/3.3e-2/7.7e-8	14/3.4e-2/4.6e-7
	1	11/8.6e-3/9.8e-1	11/1.2e-2/9.8e-1		1	30/5.7e-1/9.9e-6	23/3.8e-1/6.1e-7
Pb 6	2	12/8.4e-3/9.8e-1	14/1.2e-2/9.8e-1	Pb 18	2	63/1.2e+0/9.5e-5	21/3.6e-1/8.7e-10
	3	12/8.9e-3/9.8e-1	15/1.2e-2/9.8e-1		3	17/3.5e-1/9.9e-6	19/3.1e-1/4.9e-8
	1	18/3.3e-2/0.0e+0	36/2.8e-2/0.0e+0		1	7/3.7e-3/5.3e-8	6/7.3e-3/2.2e-7
Pb 7	2	19/3.2e-2/0.0e+0	35/2.9e-2/2.2e+0	Pb 19	2	*	14/1.6e-2/1.6e-2
	3	20/3.7e-2/0.0e+0	37/2.8e-2/1.6e+0		3	*	17/1.6e-2/1.6e-2
	1	5/3.6e-3/6.0e-5	7/8.7e-3/2.4e-7		1	4/3.6e-3/9.1e-5	4/6.8e-3/3.4e-8
Pb 8	2	6/3.5e-3/6.0e-5	9/1.0e-2/7.2e-7	Pb 20	2	5/2.4e-2/4.5e-5	7/1.5e-2/1.3e-7
	3	11/7.3e-3/6.4e-5	10/1.0e-2/7.8e-8		3	*	26/2.8e-2/1.1e+0
	1	*	10/1.4e-2/1.1e+1		1	5/4.9e-1/1.0e-9	6/4.1e-1/6.9e-6
Pb 9	2	35/5.1e-1/1.1e+1	7/1.4e-2/1.1e+1	Pb 21	2	*	189/1.7e+1/9.6e+0
	3	*	5/1.3e-2/1.1e+1		3	139/1.4e+1/1.2e+0	16/1.2e+0/1.0e+0
	1	*	*		1	6/2.3e-2/1.4e+0	6/5.8e-2/1.4e+0
Pb 10	2	*	*	Pb 22	2	7/2.4e-3/1.4e+0	7/9.4e-3/1.4e+0
	3	*	*		3	8/4.8e-3/1.4e+0	9/1.5e-2/1.4e+0
	1	9/2.1e-2/1.0e-1	51/6.7e-2/1.0e-1		1	7/8.6e-3/8.7e-6	7/1.0e-2/7.6e-8
Pb 11	2	7/6.7e-3/1.0e-1	5/1.0e-2/1.0e-1	Pb 23	2	6/4.7e-3/1.9e-7	5/1.3e-2/7.6e-8
	3	4/4.1e-3/1.0e-1	3/8.4e-3/1.0e-1		3	5/2.9e-3/1.9e-7	5/6.9e-3/7.6e-8
	1	2/1.2e-2/1.7e-15	*				
Pb 12	2.5	4/3.1e-2/6.0e-6	4/3.9e-2/1.3e-4				
	3	8/3.9e-2/4.2e-7	5/3.4e-2/5.8e-12				

Table 4.2: Performance of G-GNM-AP and ITREBO

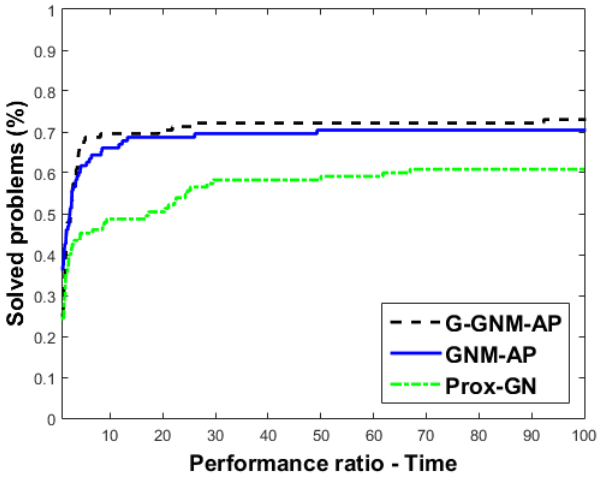

Figure 4.2: Performance of G-GNM-AP, GNM-AP and Prox-GN methods

Chapter 5

A framework with approximate projections for convex-constrained monotone nonlinear equations and its special cases

In this chapter, we propose a framework, which is obtained by combining a safeguard strategy on the search directions with a notion of approximate projections, for solving convex-constrained monotone nonlinear systems of equations. The global convergence of our framework is obtained under appropriate assumptions and some examples of methods which fall into this framework are presented.

5.1 The framework and its convergence analysis

This section describes a framework for solving (1.9) and presents its global convergence analysis.

Formally, the framework is described as follows.

Framework 1. Framework with approximate projections for convex-constrained monotone equations

---

Step 0. Let $x_0 \in C$, η_1, $\eta_2 > 0$, $\gamma, \sigma \in (0,1)$, $\bar{\mu} \in [0,1)$ and $\{\mu_k\} \subset [0, \bar{\mu}]$ be given, and set $k = 0$.

Step 1. If $\|F(x_k)\| = 0$, then **stop**.

Step 2. Compute the direction d_k in $\mathbb{R}^n$ such that

$$F(x_k)^T d_k \leq -\eta_1 \|F(x_k)\|^2, \tag{5.1}$$

$$\|d_k\| \leq \eta_2 \|F(x_k)\|. \tag{5.2}$$

Step 3. Find $z_k = x_k + \alpha_k d_k$, where $\alpha_k = \gamma^{m_k}$ with m_k being the smallest non-negative integer m such that

$$-\langle F(x_k + \gamma^m d_k),\ d_k \rangle \geq \sigma \gamma^m \|d_k\|^2. \tag{5.3}$$

Step 4. Define $\xi_k := \left(\langle F(z_k),\ x_k - z_k \rangle\right) / \|F(z_k)\|^2$, $w_k := x_k - \xi_k F(z_k)$ and $\varepsilon_k := \mu_k^2 \|\xi_k F(z_k)\|^2$. Set

$$x_{k+1} := \tilde{y}^I_{C \cap H_k}(w_k), \tag{5.4}$$

where $H_k := \{x \in \mathbb{R}^n;\ \langle F(z_k), x - z_k \rangle \leq 0\}$.

Step 5. Set $k \leftarrow k + 1$ and go to Step 1.

end

Remark 5.1.1 Some comments about Framework 1 are in order.

(i) If F is the gradient of some function $f : \mathbb{R}^n \to \mathbb{R}$, then condition (5.1) implies that d_k is a sufficient descent direction for f at x_k. In its turn, condition (5.2) essentially says that the length of $d(x_k)$ should be proportional to the length of $F(x_k)$. The way to obtain d_k satisfying (5.1) and (5.2) will depend on the particular instance of the framework; see Section 5.2 for some examples.

(ii) Note that condition (5.1) implies that there exists a non-negative number m_k satisfying (5.3), for all $k \geq 1$. Indeed, suppose that there exists $k_0 \geq 1$ such that (5.3) is not satisfied for any non-negative integer m, i.e.,

$$-\langle F(x_{k_0} + \gamma^m d_{k_0}),\ d_{k_0} \rangle < \sigma \gamma^m \|d_{k_0}\|^2,\ \forall\ m \geq 1.$$

Let $m \to \infty$ and by continuity of F, we have

$$-\langle F(x_{k_0}),\ d_{k_0} \rangle \leq 0. \tag{5.5}$$

On the other hand, by (5.1), we obtain

$$-\langle F(x_{k_0}),\ d_{k_0} \rangle \geq \delta \|F(x_{k_0})\|^2 > 0,$$

which contradicts (5.5). Therefore, the line search procedure in Step 3 is well-defined.

(iii) In Step 4, note that w_k is the projection of x_k in H_k (which has a closed-form) and x_{k+1} is an ε_k–approximate solution of the problem (2.4) with $B = I$, $w := w_k$ and feasible set $C \cap H_k$. Another choice of x_{k+1} in (5.4) would be

$$x_{k+1} := \tilde{y}_C^I(w_k). \tag{5.6}$$

For this choice, we mention that the Lemma 5.1.2 and Theorem 5.1.3 also holds.

iv) It will follow form (5.8) and (5.9) that the hyperplane H_k strictly separates the current iteration from zeroes of the system of equations (1.9).

In the course of this section, we will assume that the solution set of (1.9), denoted by S^*, is nonempty. In order to investigate the global convergence of Framework 1, the following properties of the sequences $\{x_k\}$ and $\{z_k\}$ will be needed.

Lemma 5.1.2 *The sequences $\{x_k\}$ and $\{z_k\}$ generated by Framework 1 are both bounded. Furthermore, it holds that*

$$\lim_{k\to\infty} \|x_k - z_k\| = 0. \tag{5.7}$$

Proof. From Step 3, we have

$$\langle F(z_k),\ x_k - z_k\rangle = -\alpha_k \langle F(z_k),\ d_k\rangle \geq \sigma\alpha_k^2 \|d_k\|^2 = \sigma\|x_k - z_k\|^2. \tag{5.8}$$

Note that $\|x_k - z_k\| > 0$, for all $k \geq 0$. Otherwise, since (5.1) and the Cauchy-Schwartz inequality imply that $\eta_1\|F(x_k)\| \leq \|d_k\|$, we would have $F(x_k) = 0$. Let $x_* \in S^*$ be given. By the monotonicity of F and the fact that $F(x_*) = 0$, we obtain

$$\langle F(z_k),\ x_* - z_k\rangle \leq 0. \tag{5.9}$$

Hence, $x_* \in H_k$ (see the definition of H_k in Step 4). Since $x_{k+1} = \tilde{y}^I_{C\cap H_k}(w_k)$, it follows from the fact that $x^* \in C \cap H_k$ and Lemma 2.2.6 with $x = w_k$ and $\hat{x} = x_*$ that

$$\begin{aligned} \|x_{k+1} - x_*\|^2 &= \|\tilde{y}^I_{C\cap H_k}(w_k) - y^I_{C\cap H_k}(x_*)\|^2 \leq \|w_k - x_*\|^2 + 2\varepsilon_k \\ &= \|x_k - x_*\|^2 - 2\xi_k\langle F(z_k),\ x_k - x_*\rangle + \xi_k^2\|F(z_k)\|^2 + \mu_k^2\xi_k^2\|F(z_k)\|^2. \end{aligned} \tag{5.10}$$

where we used that $\varepsilon_k^2 = (\mu_k^2\|\xi_k F(z_k)\|^2)/2$ in the last equality. It is easy to see that (5.10) also holds when $x_{k+1} = \tilde{y}_C^I(w_k)$. By the monotonicity of the mapping F and the fact that $x_* \in S^*$, we get

$$\begin{aligned} \langle F(z_k),\ x_k - z_k\rangle &= \langle F(x_*),\ z_k - x_*\rangle + \langle F(z_k),\ x_k - z_k\rangle \\ &\leq \langle F(z_k),\ z_k - x_*\rangle + \langle F(z_k),\ x_k - z_k\rangle \\ &= \langle F(z_k),\ x_k - x_*\rangle. \end{aligned} \tag{5.11}$$

By combining (5.10) and (5.11), we find

$$\begin{aligned}\|x_{k+1}-x_*\|^2 &\leq \|x_k-x_*\|^2 - 2\xi_k\langle F(z_k),\ x_k-z_k\rangle + \xi_k^2\|F(z_k)\|^2 + \mu_k^2\xi_k^2\|F(z_k)\|^2 \\ &\leq \|x_k-x_*\|^2 + (\mu_k^2-1)\frac{\langle F(z_k),\ x_k-z_k\rangle^2}{\|F(z_k)\|^2} \\ &\leq \|x_k-x_*\|^2 + (\bar{\mu}^2-1)\sigma^2\frac{\|x_k-z_k\|^4}{\|F(z_k)\|^2},\end{aligned} \tag{5.12}$$

where the second inequality follows from the definition of ξ_k and the last inequality is due to the fact that $\mu_k \leq \bar{\mu}$ and (5.8). By (5.12) and the fact that $\bar{\mu} < 1$, we have

$$\|x_{k+1}-x_*\|^2 \leq \|x_k-x_*\|^2, \quad k \geq 0, \tag{5.13}$$

which implies that the sequence $\{x_k\}$ is bounded. It follows from the Cauchy-Schwartz inequality, the monotonicity of F and (5.8) that

$$\|F(x_k)\| \geq \frac{\langle F(x_k),\ x_k-z_k\rangle}{\|x_k-z_k\|} \geq \frac{\langle F(z_k),\ x_k-z_k\rangle}{\|x_k-z_k\|} \geq \sigma\|x_k-z_k\|.$$

Therefore, by the continuity of F and the boundedness of $\{x_k\}$, we have that $\{z_k\}$ is also bounded. Since $\{z_k\}$ is bounded and F is continuous on $\mathbb{R}^n$, there exists a constant $M > 0$ such that $\|F(z_k)\| \leq M$ for all $k \geq 0$, which combined with (5.12), yields

$$\frac{(1-\bar{\mu}^2)\sigma^2}{M^2}\sum_{k=0}^{\infty}\|x_k-z_k\|^4 \leq \sum_{k=0}^{\infty}\left(\|x_k-x_*\|^2 - \|x_{k+1}-x_*\|^2\right) < \infty,$$

which implies $\lim_{k\to\infty}\|x_k-z_k\| = 0$.

■

We are now ready to establish the global convergence of Framework 1.

Theorem 5.1.3 *The sequence $\{x_k\}$ generated by Framework 1 converges to a solution of (1.9).*

Proof. Since $z_k = x_k + \alpha_k d_k$, from Lemma 5.1.2, it holds that

$$\lim_{k\to\infty}\alpha_k\|d_k\| = \lim_{k\to\infty}\|x_k-z_k\| = 0. \tag{5.14}$$

We also have, from Lemma 5.1.2, that $\{x_k\}$ is bounded and therefore $\{F(x_k)\}$ is bounded as well. Thus, it follows from the second inequality in (5.2) that $\{d_k\}$ is bounded. Consider now two different cases: (i) $\liminf_{k\to\infty}\|d_k\| = 0$ or (ii) $\liminf_{k\to\infty}\|d_k\| > 0$.
Case (i). Note that (5.1) and the Cauchy-Schwartz inequality imply that $\eta_1\|F(x_k)\| \leq \|d_k\|$. Hence, since $\liminf_{k\to\infty}\|d_k\| = 0$, it follows that $\liminf_{k\to\infty}\|F(x_k)\| = 0$. Since F

is continuous, we have the sequence $\{x_k\}$ has some cluster point $\bar{x}$ such that $F(\bar{x}) = 0$. Replacing x_* by $\bar{x}$ in (5.13), we obtain

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2,$$

which implies that $\{\|x_k - \bar{x}\|\}$ converges. Therefore, we can conclude that the whole sequence $\{x_k\}$ converges to $\bar{x}$, a solution of (1.9).
Case (ii). Since $\liminf_{k\to\infty} \|d_k\| > 0$, it follows from (5.14) that there exists a subsequence of indices $K \subset \mathbb{N}$ such that $\lim_{k\to\infty} \alpha_k = 0$, where $k \in K$. By (5.3), we have

$$-\langle F(x_k + \gamma^{m_k - 1} d_k),\ d_k\rangle < \sigma\gamma^{m_k-1}\|d_k\|^2.$$

Since $\{x_k\}$ and $\{d_k\}$ are bounded, we can choose a subsequence $K_1 \subset K$ such that $\{(x_k, d_k)\} \xrightarrow{K_1} (\bar{x}, \bar{d})$. Hence, using the continuity of F and taking the limit in the last inequality as $k \to \infty$ with $k \in K_1$, we have

$$-\langle F(\bar{x}),\ \bar{d}\rangle \leq 0. \tag{5.15}$$

On the other hand, by taking the limit in (5.1) as $k \to \infty$ with $k \in K_1$, we obtain

$$-\langle F(\bar{x}),\ \bar{d}\rangle \geq \delta\|F(\bar{x})\|^2 > 0,$$

where the last inequality is due to the inequality in (5.2) and the fact that $\liminf_{k\to\infty} \|d_k\| > 0$. Thus, the last inequality contradicts (5.15). Hence, $\liminf_{k\to\infty} \|d_k\| = 0$. Therefore, using a similar argument as in the first case, we conclude that the whole sequence $\{x_k\}$ converges to a solution of (1.9). This completes the proof. ■

5.2 Some instances of the framework

This section presents some examples of search directions d_k that satisfy the safeguard conditions (5.1) and (5.2) and as a consequence some instances of Framework 1. These instances of methods allow inexact projections onto $C \cap H_k$, which can be advantageous when the exact projections are difficult (where the projection cannot be easily performed).

Let us begin by presenting inexact versions of two well-known methods.

1) Steepest descent-based method with approximate projections (SDM-AP). This method corresponds to Framework 1 with the direction d_k in the Step 2 defined by $d_k = -F(x_k)$, for every $k \geq 0$. It is easy to see that this choice of d_k satisfies the conditions (5.1) and (5.2) with $\eta_1 = 1$ and $\eta_2 \geq 1$. Therefore, from Theorem 5.1.3, it holds that the sequence $\{x_k\}$ generated by SDM-AP converges to a solution of (1.9).

2) Newton's method with approximate projections (NM-AP). Assume that F is continuously differentiable. By taking d_k in the Step 2 of Framework 1 as $d_k = -B(x_k)^{-1}F(x_k)$ for every $k \geq 0$, where $B(x_k)$ is a positive definite matrix, we obtain a variant of Newton's method proposed in [78] with approximate projections. Note that $B(x_k)$ may be the Jacobian of F at x_k or an approximation of it. Assuming that there exist constants $0 < a \leq b$ such that $aI \prec B(x_k) \prec bI$, for every k, then d_k satisfies (5.1) and (5.2) with $\eta_1 = 1/b$ and $\eta_2 = 1/a$. Indeed, since $B_k d_k = -F(x_k)$, we obtain

$$\langle d_k, F(x_k)\rangle = \langle -B_k^{-1}F(x_k), F(x_k)\rangle = -\|F(x_k)\|^2_{B_k^{-1}} \leq -\left(\frac{1}{b}\right)\|F(x_k)\|^2$$

and

$$a\|d_k\|^2 \leq \|d_k\|^2_{B_k} = \langle B_k d_k,\ d_k\rangle = -\langle F(x_k),\ d_k\rangle \leq \|F(x_k)\|\|d_k\|,$$

which proves the statement. Therefore, since this method can be seen as an instance of Framework 1, we trivially have, from Theorem 5.1.3, that the sequence $\{x_k\}$ generated by it converges to a solution of (1.9).

We next present two examples of methods, in the spirit of the method in example 2, for the nonsmooth case. Here, we define $F : \mathbb{R}^n \to \mathbb{R}^n$ as τ-strongly monotone if there is a constant $\tau > 0$ such that $\langle x - y, F(x) - F(y)\rangle \geq \tau\|x - y\|^2$, for all $x, y \in \mathbb{R}^n$. Moreover, F is defined as $\mathcal{L}$-Lipschitz continuous if there is a constant $\mathcal{L} > 0$ such that $\|F(x) - F(y)\| \leq \mathcal{L}\|x - y\|$, for all $x, y \in \mathbb{R}^n$.

3) Spectral gradient-like methods with approximate projections (SGM-AP). Consider $d_k = -\lambda_k F(x_k)$ for every $k \geq 0$, where λ_k is the spectral coefficient which is related to the Barzilai-Borwein choice of the step-size [8]. Let us first discuss some existing choices of λ_k.

3.1) In [52], λ_k is defined by

$$\lambda_k = \frac{\langle s_k, s_k\rangle}{\langle s_k, u_k\rangle}, \tag{5.16}$$

where $s_k := x_k - x_{k-1}$ and $u_k := F(x_k) - F(x_{k-1})$. Under the assumption that F is τ-strongly monotone and $\mathcal{L}$-Lipschitz continuous, we have that $d_k = -\lambda_k F(x_k)$ satisfies (5.1) and (5.2) with $\eta_1 = 1/\mathcal{L}$ and $\eta_2 = 1/\tau$. Indeed, using that F is τ-strongly monotone, we have

$$\langle s_k, u_k\rangle = \langle x_k - x_{k-1}, F(x_k) - F(x_{k-1})\rangle \geq \tau\langle x_k - x_{k-1}, x_k - x_{k-1}\rangle = \tau\langle s_k, s_k\rangle > 0,$$

for some $\tau > 0$, and therefore, $\lambda_k \leq 1/\tau$. Now, using the Cauchy-Schwarz inequality and that F is $\mathcal{L}$-Lipschitz continuous, we obtain

$$\langle s_k, u_k\rangle = \langle s_k, F(x_k) - F(x_{k-1})\rangle \leq \|F(x_k) - F(x_{k-1})\|\|s_k\| \leq \mathcal{L}\langle s_k, s_k\rangle,$$

which implies $1/\mathcal{L} \leq \lambda_k$. Thus, $1/\mathcal{L} \leq \lambda_k \leq 1/\tau$ and, as a consequence, $d_k = -\lambda_k F(x_k)$ satisfies (5.1) and (5.2) with $\eta_1 = 1/\mathcal{L}$ and $\eta_2 = 1/\tau$.

3.2) In [81, 82], the coefficient λ_k is as in (5.16) with $s_k := x_k - x_{k-1}$ and $u_k := F(x_k) - F(x_{k-1}) + rs_k$, where $r > 0$ is a given scalar. Using that F is monotone, we have

$$\begin{aligned}\langle s_k, u_k\rangle &= \langle s_k, F(x_k) - F(x_{k-1}) + rs_k\rangle \\ &= \langle x_k - x_{k-1}, F(x_k) - F(x_{k-1})\rangle + r\langle s_k, s_k\rangle \\ &\geq r\langle s_k, s_k\rangle > 0,\end{aligned}$$

which implies that $\lambda_k \leq 1/r$. Now, by assuming that F is $\mathcal{L}$-Lipschitz continuous, we obtain

$$\begin{aligned}\langle s_k, u_k\rangle &= \langle s_k, F(x_k) - F(x_{k-1}) + rs_k\rangle \\ &= \langle x_k - x_{k-1}, F(x_k) - F(x_{k-1})\rangle + r\langle s_k, s_k\rangle \\ &\leq (\mathcal{L} + r)\langle s_k, s_k\rangle,\end{aligned}$$

which yields $1/(\mathcal{L} + r) \leq \lambda_k$. Therefore, as $1/(\mathcal{L} + r) \leq \lambda_k \leq 1/r$, we can conclude, from the fact that $d_k = -\lambda_k F(x_k)$, that d_k satisfies the conditions (5.1) and (5.2) with $\eta_1 = 1/(\mathcal{L} + r)$ and $\eta_2 = 1/r$.

3.3) In the works [1, 62] the coefficient λ_k is a convex combination of the default spectral coefficient in [8] and the positive spectral coefficient in [21]. More specifically, λ_k defined by

$$\lambda_k = (1 - t)\theta_k^* + t\theta_k^{**},$$

where $t \in [0, 1]$, $\theta_k^* = \|s_k\|^2/\langle u_k, s_k\rangle$, $\theta_k^{**} = \|s_k\|/\|u_k\|$, $s_k := x_k - x_{k-1}$, $u_k := F(x_k) - F(x_{k-1}) + rs_k$ and $r > 0$. In [1, Lemma 2], it was shown that if F is $\mathcal{L}$-Lipschitz continuous, then $d_k = -\lambda_k F(x_k)$ satisfies (5.1) and (5.2) with $\eta_1 = \max\{1, 1/(\mathcal{L} + r)\}$ and $\eta_2 = \min\{1, 1/r\}$.

Since the search directions in examples 3.1, 3.2 and 3.3 satisfy (5.1) and (5.2) for specific values of η_1 and η_2, we can conclude, from Theorem 5.1.3, that the SGM-AP (i.e., Framework 1 with the above three choice of search directions) converges to a solution of (1.9).

4) Limited memory BFGS method with approximate projections (L-BFGS-AP). Consider the L-BFGS direction d_k proposed in [83] obtained by solving the system $B_k d_k = -F(x_k)$, where the sequence $\{B_k\}$ is given by $B_0 = I$ and B_{k+1} is computed by the following modified L-BFGS update process: let $m > 0$ be given and set $\tilde{m} = \min\{k + 1, m\}$ and $B_k^{(0)} = B_0 = I$. Choose a set of increasing integers $L_k = \{j_0, \ldots, j_{\tilde{m}-1}\} \subset \{0, \ldots, k\}$. Update B_{k+1} using the pairs $\{y_{j_l}, s_{j_l}\}_{l=0}^{\tilde{m}-1}$, i.e., for $l = 0, \ldots, \tilde{m} - 1$,

$$B_{k+1} := B_{k+1}^{(l+1)} = \begin{cases} B_k^{(l)} - \frac{B_k^{(l)} s_{j_l} s_{j_l}^T B_k^{(l)}}{s_{j_l}^T B_k^{(l)} s_{j_l}} + \frac{y_{j_l} y_{j_l}^T}{y_{j_l}^T s_{j_l}}, & if\ \frac{y_{j_l}^T s_{j_l}}{\|s_{j_l}\|^2} \geq \varepsilon, \\ B_k^{(l)}, & \text{otherwise,} \end{cases}$$

where $s_k := x_{k+1} - x_k$ and $y_k := F(x_{k+1}) - F(x_k)$. If d_k in the Step 2 of Framework 1 is defined as above, we obtain an L-BFGS method with approximate projections. Under the assumption that F is $\mathcal{L}$-Lipschitz continuous, it was proven in [83] that B_k and B_k^{-1} are bounded for all $k \geq 0$, i.e., $\{B_k\} \subset \mathbb{B}$. Since $B_k d_k = -F(x_k)$ and using (2.2), we obtain

$$\langle d_k, F(x_k)\rangle = \langle -B_k^{-1}F(x_k), F(x_k)\rangle = -\|F(x_k)\|^2_{B_k^{-1}} \leq -\left(\frac{1}{\|B_k\|}\right)\|F(x_k)\|^2,$$

which yields $\langle d_k, F(x_k)\rangle \leq -1/L\|F(x_k)\|^2$. Now, from Cauchy-Schwarz inequality, we have

$$\|d_k\|^2_{B_k} = \langle B_k d_k,\ d_k\rangle = -\langle F(x_k),\ d_k\rangle \leq \|F(x_k)\|\|d_k\|,$$

which, combined with (2.2) and $\|B_k^{-1}\| \leq L$, yields $(1/L)\|d_k\| \leq \|F(x_k)\|$. Thus, d_k satisfies (5.1) and (5.2) with $\eta_1 = 1/L$ and $\eta_2 = L$. Therefore, we conclude that, from Theorem 5.1.3, the sequence $\{x_k\}$ generated by L-BFGS-AP (i.e., Framework 1 with the above choice of search direction) converges to a solution of (1.9).

We end this section by proposing a new convergent method for solving (1.9), which is an instance of Framework 1. This method is inspired by [74, Algorithm 2.1] for solving variational inequalities. In the context that the projection operator is computationally expensive, the latter algorithm was devised in order to minimize the total number of performed projection operations. Let us now present our extension of [74, Algorithm 2.1] to the convex-constrained monotone nonlinear equations context.

5) Modified Newton-like method with approximate projections (MNM-AP). Consider the direction d_k defined as follows: let $\eta > 0$, $\bar{\theta} \in [0, \eta)$ and $\{\theta_k\} \subset [0, \bar{\theta}]$ be given. Let $B_k \subset \mathbb{B}$, and set $w_k^1 := B_k x_k - F(x_k)$ and $\varepsilon_k^1 := \theta_k^2\|F(x_k)\|^2$. Compute s_k^1 in $\mathbb{R}^n$ such that

$$s_k^1 = \tilde{y}_C^{B_k}(w_k^1) - x_k, \tag{5.17}$$

where $\tilde{y}_C^{B_k}(w_k^1)$ is an ε_k^1–approximate solution of the problem (2.4). If $\eta\|F(x_k)\| \leq \|s_k^1\|_{B_k}$, then $d_k := s_k^1$. Otherwise, compute s_k^2 in $\mathbb{R}^n$ such that

$$F(x_k) + B_k s_k^2 = 0, \tag{5.18}$$

and set $d_k := s_k^2$. Note that the matrix B_k can be taken as those in the examples 3 and 4. We will now prove that d_k described above satisfies (5.1) and (5.2), for all $k \geq 0$. If $\eta\|F(x_k)\| \leq \|s_k^1\|_{B_k}$, then $d_k = \tilde{y}_C^{B_k}(w_k^1) - x_k$. By (2.5) with $B = B_k$, $w = w_k^1$ and $y = x_k$, we have

$$\begin{aligned}\theta_k^2\|F(x_k)\|^2 \geq& \langle B_k(x_k - \tilde{y}_C^{B_k}(w_k^1)) - F(x_k), x_k - \tilde{y}_C^{B_k}(w_k^1)\rangle \\ =& \|\tilde{y}_C^{B_k}(w_k^1) - x_k\|^2_{B_k} - \langle F(x_k),\ x_k - \tilde{y}_C^{B_k}(w_k^1)\rangle,\ \forall\ k \geq 0,\end{aligned}$$

which, combined with the definition of d_k, yields

$$-\langle F(x_k), d_k\rangle + \theta_k^2\|F(x_k)\|^2 \geq \|d_k\|^2_{B_k} \geq \|F(x_k)\|^2\eta^2,\ \forall\ k \geq 0, \tag{5.19}$$

or, equivalently,

$$-\langle F(x_k), d_k\rangle \geq \|F(x_k)\|^2(\eta^2 - \theta_k^2),\ \forall\ k \geq 0.$$

Therefore, since $\theta_k \leq \bar{\theta}$ for all $k \geq 0$ and $\bar{\theta} \in [0, \eta)$, we have

$$\langle F(x_k), d_k\rangle \leq -(\eta^2 - \bar{\theta}^2)\|F(x_k)\|^2.$$

Hence, (5.1) holds with $\eta_1 = (\eta^2 - \bar{\theta}^2)$. From (5.19) and using the Cauchy-Schwarz inequality, we have

$$\|d_k\|_{B_k}^2 \leq \theta_k^2\|F(x_k)\|^2 - \langle B_k^{-1}F(x_k), d_k\rangle_{B_k} \leq \theta_k^2\|F(x_k)\|^2 + \|B_k^{-1}F(x_k)\|_{B_k}\|d_k\|_{B_k},$$

which, combined with some algebraic manipulations, yields

$$\|d_k\|_{B_k}^2 \leq \theta_k^2\|F(x_k)\|^2 + \frac{\|B_k^{-1}F(x_k)\|_{B_k}^2}{2} + \frac{\|d_k\|_{B_k}^2}{2}.$$

Using the definition of scalar product $\langle\cdot,\cdot\rangle_B = \langle\cdot, B\cdot\rangle$ and (2.2), we obtain

$$\frac{\|d_k\|_{B_k}^2}{2} \leq \theta_k^2\|F(x_k)\|^2 + \frac{\|F(x_k)\|_{B_k^{-1}}^2}{2} \leq \theta_k^2\|F(x_k)\|^2 + \frac{\|F(x_k)\|^2\|B_k^{-1}\|}{2} = \left(\theta_k^2 + \frac{\|B_k^{-1}\|}{2}\right)\|F(x_k)\|^2,$$

which implies that

$$\|d_k\|_{B_k}^2 \leq \left(2\theta_k^2 + \|B_k^{-1}\|\right)\|F(x_k)\|^2.$$

Therefore, by (2.2), $\|B_k^{-1}\| \leq L$ and $\theta_k \leq \bar{\theta}$ for all $k \geq 0$, we have

$$\|d_k\|^2 \leq L\left(2\bar{\theta}^2 + L\right)\|F(x_k)\|^2,$$

and hence (5.2) holds with $\eta_2 = \sqrt{L\left(2\bar{\theta}^2 + L\right)}$. On the other hand, if $d_k := s_k^2$, then the proof is similar to the one in example 4. Therefore, we conclude that, from Theorem 5.1.3, the sequence $\{x_k\}$ generated by the MNM-AP (i.e., Framework 1 with the above choice of search direction) converges to a solution of (1.9).

5.3 Numerical experiments

This section summarizes the numerical experiments carried out to verify the efficiency of the instances of Framework 1. Numerical experiments are divided into two subsections. In Subsection 5.3.1, the methods are tested for a group of convex-constrained monotone nonlinear equations, whereas, in Subsection 5.3.2, they are tested for solving the system of constrained absolute value equations (CAVE). The computational results are obtained using MATLAB R2018a on a 2.4GHz Intel(R) i5 with 8GB of RAM and Windows 10 ultimate system.

5.3.1 Monotone nonlinear equations with polyhedral constraints

In this subsection, our aim is to illustrate the behavior of the methods to solve 52 monotone nonlinear equations with polyhedral constraints; see Tables 5.1 and 5.2. Some of these problems are originally unconstrained for which constrains were added. In Pb11, the matrix $A \in \mathbb{R}^{10\times n}$ of Table 5.2 is randomly generated so that a solution of the problem 11 belongs to the feasible set.

Problem	Ref.	n
Pb 1	[81, Problem 1]	1000/5000/10000
Pb 2	[81, Problem 2]	1000/5000/10000
Pb 3	[78, Problem 2]	5/5/5
Pb 4	[78, Problem 3]	10/10/10
Pb 5	[78, Problem 4]	4
Pb 6	[1, Problem 1]	1000/5000/10000
Pb 7	[1, Problem 2]	1000/5000/10000
Pb 8	[1, Problem 3]	1000/5000/10000
Pb 9	[1, Problem 5]	1000/5000/10000
Pb 10	[1, Problem 6]	1000/5000/10000
Pb 11	[51, Problem 1]	1000/5000/10000
Pb 12	[51, Problem 4]	1000/5000/10000
Pb 13	[51, Problem 7]	1000/5000/10000
Pb 14	[51, Problem 8]	1000/5000/10000
Pb 15	[51, Problem 9]	1000/5000/10000
Pb 16	[55, Problem 2]	1000/5000/10000
Pb 17	[55, Problem 3]	1000/5000/10000
Pb 18	[55, Problem 7]	1000/5000/10000

Table 5.1: Test problems

The tolerance in the stopping criterion $\|F(x_k)\| < \epsilon$ was set to $\epsilon = 10^{-6}$. If the stopping criterion is not satisfied, the method stops when a maximum of 500 iterations has been performed. In this first group of test problems, it is taken $\sigma = 10^{-4}$, $\gamma = 1/2$ and $\mu_k = \bar{\mu} = 0.25$, for every k, in all algorithms. Moreover, the ε_k–approximate solution in (5.4)

Problem	Set C
Pb 1	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 2	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 3	$[0, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 4	$[0, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 5	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq 3$
Pb 6	$[-1, 2]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 7	$[-1, 2]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 8	$[0, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 9	$[-1, 7]$ and $\sum_{i=1}^{n} x_i \leq 1.1 \cdot n$
Pb 10	$[0, e]$ and $\sum_{i=1}^{n} x_i \leq e \cdot n$
Pb 11	$[-1, 2]$; $Ax \leq b$, where $A \in \mathbb{R}^{10\times n}$ and $b = (n, \ldots, n) \in \mathbb{R}^{10}$
Pb 12	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq 20 \cdot n$
Pb 13	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 14	$[-n, 1]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 15	$[-n, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 16	$[-n, 1]$ and $\sum_{i=1}^{n} x_i \leq 1$
Pb 17	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq n$
Pb 18	$[-1, n]$ and $\sum_{i=1}^{n} x_i \leq n$

Table 5.2: Polyhedral feasible sets

was computed by the conditional gradient method, which stopped when either the stopping criterion is satisfied or a maximum of 300 iterations is performed. In order to avoid an excessive number of inner iterations, input ε_k was replaced by $\max\{\mu_k^2 \|\xi_k F(z_k)\|^2, 10^{-2}\}$. Linear optimization subproblems in the conditional gradient method (see (2.11)) were solved via the MATLAB command `linprog`. We denote by SGM-AP1, SGM-AP2 and SGM-AP3, the method SGM-AP, with the coefficient λ_k given in examples 3.1, 3.2 and 3.3, respectively. In SGM-AP2, we set $r = 0.01$, whereas, in SGM-AP3, we set $t = 1/(exp(k+1)^{k+1})$ and $r = 1/(k+1)^2$. In the L-BFGS-AP, we used $m = 1$. Finally, we set $\eta = 0.5$, $\theta_k = \bar{\theta} = 0.25$ in the MNM-AP.

We consider 4 different starting points (following the suggestions where the problems

were proposed) for each problem of Table 5.1: For problem 1, $x_1 = (0.1, \ldots, 0.1)$, $x_2 = (1, \ldots, 1)$, $x_3 = ((n-1)/n, 0.1, \ldots, 0.1, (n-1)/n)$ and $x_4 = (-1, \ldots, -1)$. For problem 2, $x_1 = (0.1, \ldots, 0.1)$, $x_2 = (1, \ldots, 1)$, $x_3 = (0, \ldots, 0)$ and $x_4 = (-1, \ldots, -1)$. For problem 3, $x_1 = (10, 0, \ldots, 0)$, $x_2 = (9, 0, \ldots, 0)$, $x_3 = (3, 0, 3, 0, 3)$ and $x_4 = (0, 2, 2, 2, 2)$. For problem 5, $x_1 = (0, \ldots, 0)$, $x_2 = (3, 0, 0, 0)$, $x_3 = (1, 1, 1, 0)$ and $x_4 = (0, 1, 1, 1)$. For problems 14 and 16, $x_1 = (-1, \ldots, -1)$, $x_2 = (-0.1, \ldots, -0.1)$, $x_3 = (-1/2, -1/2^2, \ldots, -1/2^n)$ and $x_4 = (-1, -1/2, \ldots, -1/n)$. For problem 17, $x_1 = ((n-1)/n, 0.1, \ldots, 0.1, (n-1)/n)$, $x_2 = (0.1, \ldots, 0.1)$, $x_3 = (1/2, 1/2^2, \ldots, 1/2^n)$ and $x_4 = (1, 1/2, \ldots, 1/n)$. For problems 4, 6 to 13, 15 and 18, $x_1 = (1, \ldots, 1)$, $x_2 = (0.1, \ldots, 0.1)$, $x_3 = (1/2, 1/2^2, \ldots, 1/2^n)$ and $x_4 = (1, 1/2, \ldots, 1/n)$. Figures 5.2 and 5.1 report the numerical results of SDM-AP, SGM-AP1, SGM-AP2, SGM-AP3, L-BFGS-AP and MNM-AP for solving the 52 problems using performance profiles [26]. We adopted the CPU time as performance measurement. Recall that in the performance profile, efficiency and robustness can be accessed on the left and right extremes of the graphic, respectively. We consider that a method is the most efficient if its runtime does not exceed in 5% the CPU time of the fastest one.

From Figures 5.1 and 5.2, we can see that all the variations of the SGM-AP achieved better performance (in terms of efficient and robust) compared to L-BFGS-AP and MNM-AP. In the group of SGM-AP variants, SGM-AP1 and SGM-AP2 were better than the others.

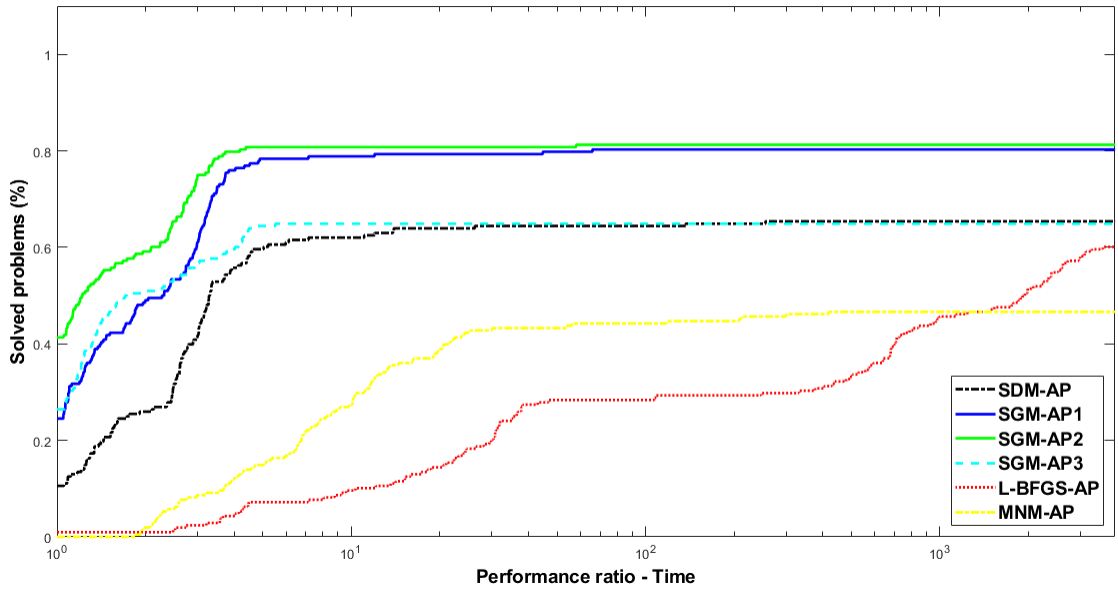

Figure 5.1: Performance of SDM-AP, SGM-AP1, SGM-AP2, SGM-AP3, L-BFGS-AP and MNM-AP with x_{k+1} as in (5.4)

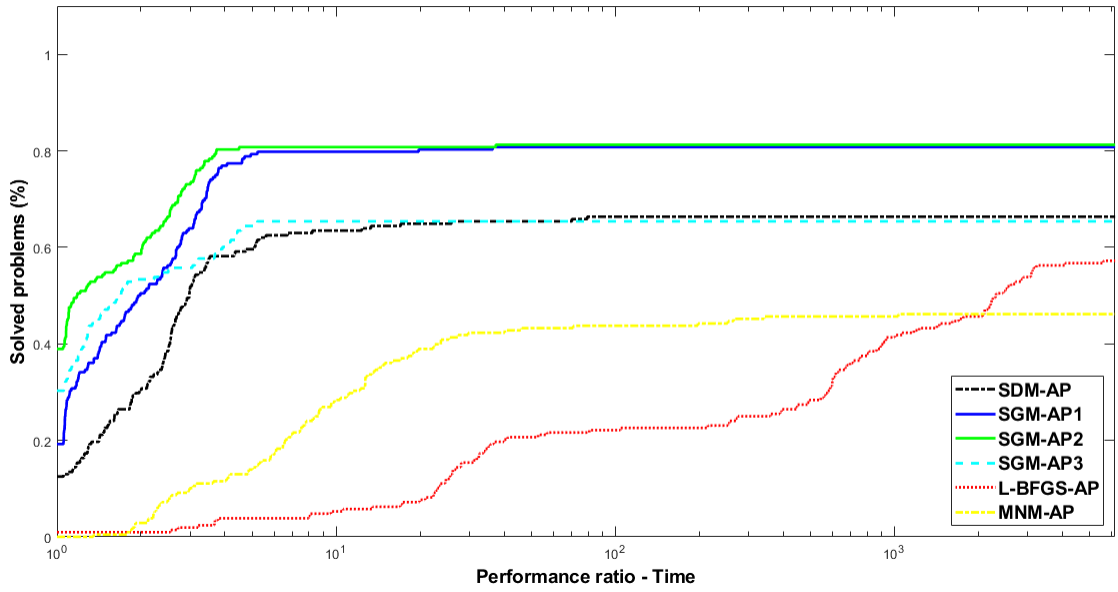

Figure 5.2: Performance of SDM-AP, SGM-AP1, SGM-AP2, SGM-AP3, L-BFGS-AP and MNM-AP with x_{k+1} as in (5.6)

5.3.2 Absolute value equations with polyhedral constraints

In this subsection, we consider the problem of finding a solution of the CAVE problem:

$$\text{find } x \in C \text{ such that } Ax - |x| = b, \tag{5.20}$$

where $C := \{x \in \mathbb{R}^n; \sum_{i=1}^n x_i \leq d, x_i \geq -1, i = 1, \ldots, n\}$, $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n \equiv \mathbb{R}^{n\times 1}$, and $|x|$ denotes the vector whose i-th component is equal to $|x_i|$. The problem (5.20) draws attention for its simple formulation when compared to its equivalent linear complementarity problem (LCP) (see [18,19,59]) which in turn includes linear programs, quadratic programs, bimatrix games and other problems. Hence, interesting algorithms relating to Newton-type methods to solve (5.20) have been developed; see, for example, [20,58] and [65] for the unconstrained and constrained case, respectively.

Under the assumption that $\|A^{-1}\| \leq 1$, it was proven in [59, Proposition 4] that the problem (5.20), with $C = \mathbb{R}^n$, is uniquely solvable for any b. Now, if A is symmetric positive definite, then $F(x) = Ax - |x| - b$ is monotone. In fact, for all $x, y \in \mathbb{R}^n$, we have

$$\begin{aligned}\langle F(x) - F(y), x - y\rangle &= \langle Ax - |x| - Ay + |y|, x - y\rangle = \|x - y\|_A^2 + \langle |y| - |x|, x - y\rangle \\ &\geq \|x - y\|^2 \frac{1}{\|A^{-1}\|} + \langle |y| - |x|, x - y\rangle \geq \|x - y\|^2 + \langle |y| - |x|, x - y\rangle.\end{aligned} \tag{5.21}$$

where in the second equality we use that $\langle \cdot, \cdot\rangle_B = \langle \cdot, B\cdot\rangle$, (2.2) and $\|A^{-1}\| \leq 1$. Now, note

that $|x|$ can be written as $|x| = P_{\mathbb{R}^n_+}(x) + P_{\mathbb{R}^n_+}(-x)$. So, from (5.21), the Cauchy-Schwarz inequality and the fact that $P_{\mathbb{R}^n_+}(\cdot)$ is monotone and nonexpansive, we obtain

$$\begin{aligned}\langle F(x) - F(y), x - y\rangle &\geq \|x - y\|^2 + \langle P_{\mathbb{R}^n_+}(y) + P_{\mathbb{R}^n_+}(-y) - P_{\mathbb{R}^n_+}(x) - P_{\mathbb{R}^n_+}(-x), x - y\rangle \\ &= \|x - y\|^2 - \langle P_{\mathbb{R}^n_+}(x) - P_{\mathbb{R}^n_+}(y), x - y\rangle + \langle P_{\mathbb{R}^n_+}(-y) - P_{\mathbb{R}^n_+}(-x), x - y)\rangle \\ &\geq \|x - y\|^2 - \|P_{\mathbb{R}^n_+}(x) - P_{\mathbb{R}^n_+}(y)\|\|x - y\| \\ &\geq \|x - y\|^2 - \|x - y\|^2 = 0,\end{aligned}$$

which proves the statement. In our implementation, we used the Matlab routine `sprandsym` to construct matrix A randomly, which generates a symmetric positive definite sparse matrix with predefined dimension, density and singular values. For this process, the density of matrix A was set to 0.003 and the vector of singular values was randomly generated from a uniform distribution on $(0, 1)$. In this case, as the vector of singular values (rc) is a vector of length n, then A has eigenvalues rc. Thus, if rc is a positive (non-negative) vector then A is a positive (non-negative) definite matrix. We chose a random solution x_* from a uniform distribution on $(0.1, 10)$ 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_*. The initial points were defined as $x_0 = (0, \ldots, 0, d/2, 0, \ldots, 0, d/2, 0, \ldots, 0) \in \mathbb{R}^n$, where the two positions of $d/2$ were generated randomly on the set $\{1, 2, \ldots, n\}$.

For the CAVE problem, we consider only the SGM-AP2 since it was the best method in our first class of experiment described in Subsection 5.3.1. For a comparative purpose, we also run the inexact Newton method with feasible inexact projections (INM-InexP) of [65]. INM-InexP is an algorithm designed for solving smooth and nonsmooth equations subject to a set of constraints. We rescale the vector of singular values to ensure that the condition $\|A^{-1}\| \leq 1/3 < 1$ is fulfilled and consequently ensure the good definition of INM-InexP (see [58, Theorem 2] for more details). In INM-InexP, we set $\theta = \bar{\theta} = \bar{\mu} = 0.25$ and the other parameters were set as in [65]. For both algorithms, a failure was declared if the number of iterations was greater than 500. The procedure to obtain inexact projections used in the implementation of INM-InexP was also the CondG method and the procedure stopped when either the condition as in [65, Algorithm 1] was satisfied or a maximum of 10 iterations were performed. For our algorithms, the procedure stopped when either the stopping criterion, i.e., $\langle w_k - x_{k+1}, y - x_{k+1}\rangle \leq \varepsilon_k := \mu_k^2 \|\xi_k F(z_k)\|^2$, for all $y \in C \cap H_k (or\ C)$, was satisfied or a maximum of 10 iterations were performed.

As in Subsection 5.3.1, Figure 5.3 reports numerical results of algorithms using performance profiles. We generated 50 CAVEs of dimensions 1000, 5000 and 10000 and for each of them we test the algorithm for 5 different initial points. We see, from Figure 5.3, that the SGM-AP2-C (with x_{k+1} as in (5.6)) was the most robust whereas INM-InexP was more efficient in terms of time saving than SGM-AP2-C and SGM-AP2-CH (with x_{k+1} as in

(5.4)).

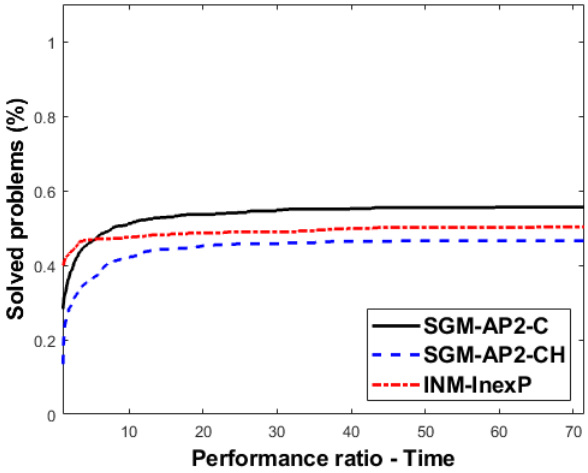

Figure 5.3: Performance of SGM-AP2-C, SGM-AP2-CH and INM-InexP

Chapter 6

Final remarks

In this thesis, we proposed and analyzed some methods to solve constrained optimization problems and constrained monotone nonlinear systems of equations

In Chapter 3, we proposed an modified inexact variable metric method (M-IVM), with a new inexactness criterion for its subproblems, for solving convex-constrained optimization problems. When necessary, such inexact solutions of the subproblems can be obtained by using suitable iterative algorithms; for example, the conditional gradient method (Frank-Wolfe) [27, 32] and its variants. Under mild assumptions, we proved that any accumulation point of the sequence generated by the proposed method is a stationary point of (1.1). Preliminary numerical experiments showed that the new algorithm works well and compares favorably with a previous IVM on linearly constrained problems, and with its exact version and the interior point method in [75] for semidefinite least squares problems.

In Chapter 4, we proposed Gauss-Newton methods with approximate projections (GNM-AP) for solving constrained nonlinear least squares problems. For the local method, we were able to show, under a majorant condition, that the generated sequence converges locally linearly. In zero-residual problems, quadratic convergence rate can be achieved with a stronger condition on the inexactness of the projections. As special cases of the majorant condition, convergence results for the method with F' satisfying a Lipschitz-like condition and F being an analytic function satisfying a Smale condition were also discussed. For the global method, under suitable conditions, the global convergence of the algorithm to a stationary point of the problem was established. The numerical experiments showed that the new algorithms work quite well and compare favorably with the proximal Gauss-Newton method in [70] (which corresponds to an exact version of our GNM-AP) and the inexact Gauss-Newton trust-region method in [68] for simple bounds.

In Chapter 5, we proposed a framework with approximate projections for constrained monotone equations. Under mild assumptions, we proved that the sequence generated by

the proposed framework converges to a solution of (1.9). Some examples of methods which fall into this framework were presented. Preliminary numerical experiments showed that some methods, which fall into the framework, performed well to solve constrained monotone nonlinear equations, and they are competitive in terms of robustness with the Inexact Newton method with feasible inexact projections in [65] for solving absolute value equations with polyhedral constraints.

Bibliography

[1] A. B. Abubakar, P. Kumam, and H. Mohammad. A note on the spectral gradient projection method for nonlinear monotone equations with applications. *Comp. Appl. Math.*, 39:1 – 35, 2020.

[2] Z. Allen-Zhu, E. Hazan, W. Hu, and Y. Li. Linear convergence of a Frank-Wolfe type algorithm over trace-norm balls. In *Adv. Neural Inf. Processing Syst.*, volume 30, pages 6191 – 6200. Curran Associates, Inc., 2017.

[3] N. M. Alsaifi and P. Englezos. Prediction of multiphase equilibrium using the PC-SAFT equation of state and simultaneous testing of phase stability. *Fluid Phase Equilib.*, 302(1):169 – 178, 2011.

[4] R. Andreani, E. G. Birgin, J. M. Martínez, and J. Yuan. Spectral projected gradient and variable metric methods for optimization with linear inequalities. *IMA J. Numer. Anal.*, 25:221 – 252, 2005.

[5] R. Andreani, A. Friedlander, M. P. Mello, and S. A. Santos. Box-constrained minimization reformulations of complementarity problems in second-order cones. *J. Glob. Optim.*, 40(4):505 – 527, 2008.

[6] M. Andretta, E. G. Birgin, and J. M. Martínez. Partial spectral projected gradient method with active-set strategy for linearly constrained optimization. *Numer. Algor.*, 53:23 – 52, 2009.

[7] I. K. Argyros and A. A. Magreñán. Local convergence analysis of proximal Gauss-Newton method for penalized nonlinear least squares problems. *Appl. Math. Comput.*, 241:401 – 408, 2014.

[8] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8(1):141 – 148, 1988.

[9] A. Ben-Israel. A modified Newton-Raphson method for the solution of systems of equations. *Isr. J. Math.*, 3(2):94 – 98, 1965.

[10] L. Bencini, R. Fantacci, and L. Maccari. Analytical model for performance analysis of ieee 802.11 dcf mechanism in multi-radio wireless networks. In *2010 IEEE International Conference on Communications*, pages 1 – 5. IEEE, 2010.

[11] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, second edition, 1999.

[12] E. G. Birgin, J. M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, 10:1196 – 1211, 2000.

[13] E. G. Birgin, J. M. Martínez, and M. Raydan. Algorithm 813: SPG - software for convex-constrained optimization. *ACM Trans. Math. Softw.*, 27(3):340 – 349, 2001.

[14] E. G. Birgin, J. M. Martínez, and M. Raydan. Inexact spectral projected gradient methods on convex sets. *IMA J. Numer. Anal.*, 23:539 – 559, 2003.

[15] E. G. Birgin, J. M. Martínez, and M. Raydan. Spectral projected gradient methods: Review and perspectives. *J. Stat. Softw.*, 60(3):1 – 21, 2014.

[16] J. P. Boyle and R. L. Dykstra. A method for finding projections onto the intersection of convex sets in Hilbert spaces. In R. Dykstra, T. Robertson, and F. T. Wright, editors, *Advances in Order Restricted Statistical Inference*, volume 37, pages 28 – 47, New York, NY, 1986. Springer New York.

[17] D. Cores, R. Escalante, M. González-Lima, and O. Jimenez. On the use of the Spectral Projected Gradient method for Support Vector Machines. *J. Comput. Appl. Math.*, 28:327 – 364, 2009.

[18] R. W. Cottle and G. B. Dantzig. Complementary pivot theory of mathematical programming. *Linear Algebra Appl.*, 1(1):103 – 125, 1968.

[19] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Society for Industrial and Applied Mathematics, 2009.

[20] J. Y. B. 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.

[21] Y.-H. Dai, M. Al-Baali, and X. Yang. A positive Barzilai-Borwein-like stepsize and an extension for symmetric linear systems. In *Numer. Anal. Optim.*, volume 134, pages 59 – 75. Springer, 2015.

[22] J.-P. Dedieu and M. Shub. Newton's method for overdetermined systems of equations. *Math. Comput.*, 69(231):1099 – 1115, 2000.

[23] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1996.

[24] L. Ding, Y. Fei, Q. Xu, and C. Yang. Spectral Frank-Wolfe algorithm: Strict complementarity and linear convergence. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.

[25] S. P. Dirkse and M. C. Ferris. MCPLIB: a collection of nonlinear mixed complementarity problems. *Optim. Meth. Softw.*, 5(4):319 – 345, 1995.

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

[27] J. C. Dunn. Convergence rates for conditional gradient sequences generated by implicit step length rules. *SIAM J. Control Optim.*, 18(5):473 – 487, 1980.

[28] M. E. El-Hawary. *Optimal Power Flow: Solution Techniques, Requirements and Challenges: IEEE Tutorial Course*. Piscataway: IEEE, 1996.

[29] O. P. Ferreira, M. L. N. Gonçalves, and P. R. Oliveira. Local convergence analysis of the Gauss-Newton method under a majorant condition. *J. Complex.*, 27(1):111 – 125, 2011.

[30] O. P. Ferreira, M. L. N. Gonçalves, and P. R. Oliveira. Convergence of the Gauss-Newton method for convex composite optimization under a majorant condition. *SIAM J. Optim.*, 23(3):1757 – 1783, 2013.

[31] R. Fletcher. *Practical methods of optimization*. A Wiley-Interscience Publication. John Wiley & Sons, Ltd., Chichester, second edition, 1987.

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

[33] R. M. Freund and P. Grigas. New analysis and results for the Frank-Wolfe method. *Mathematical Programming*, 155:199 – 230, 2016.

[34] H. Fu, H. Liu, B. Han, Y. Yang, and Y. Hu. A proximal iteratively regularized Gauss-Newton method for nonlinear inverse problems. *J. Inverse Ill-Posed Probl.*, 25(3):341 – 356, 2017.

[35] D. Garber. Faster projection-free convex optimization over the spectrahedron. In *Adv. Neural Inf. Processing Syst.*, pages 874 – 882, 2016.

[36] M. A. Gomes-Ruggiero, J. M. Martínez, and S. A. Santos. Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. *J. Sci. Comput.*, 31(3):1628 – 1652, 2009.

[37] D. Gonçalves, M. Gomes-Ruggiero, and C. Lavor. A projected gradient method for optimization over density matrices. *Optim. Meth. Softw.*, 31(2):328 – 341, 2016.

[38] D. S. Gonçalves, M. L. N. Gonçalves, and F. R. Oliveira. An inexact projected LM type algorithm for solving convex constrained nonlinear equations. *J. Comput. Appl. Math.*, 391:113421, 2021.

[39] M. L. N. Gonçalves and T. C. Menezes. Gauss-Newton methods with approximate projections for solving constrained nonlinear least squares problems. *J. Complex.*, 58:101459, 2020.

[40] D. S. Gonçalves, M. L. N. Gonçalves, and T. C. Menezes. Inexact variable metric method for convex-constrained optimization problems. *Optim.*, pages 1 – 19, 2021.

[41] M. L. N. Gonçalves. Local convergence of the Gauss-Newton method for injective-overdetermined systems of equations under a majorant condition. *Comput. Math. Appl.*, 66(4):490 – 499, 2013.

[42] M. L. N. Gonçalves. Inexact Gauss-Newton like methods for injective-overdetermined systems of equations under a majorant condition. *Numer. Algor.*, 72(2):377 – 392, 2016.

[43] N. I. M. Gould, D. Orban, and P. L. Toint. Cuter, a constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Softw.*, 29:373 – 394, 2003.

[44] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, 23(4):707 – 716, 1986.

[45] J. Guélat and P. Marcotte. Some comments on Wolfe's 'away step'. *Math. Program.*, 35:110 – 119, 1986.

[46] J. Guerrero, M. Raydan, and M. Rojas. A hybrid-optimization method for large-scale non-negative full regularization in image restoration. *Inver. Probl. Sci. En.*, 21:741 – 766, 2011.

[47] N. J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.*, 103:103 – 118, 1988.

[48] Z. Huang. The convergence ball of Newton's method and the uniqueness ball of equations under Hölder-type continuous derivatives. *Comput. Math. Appl.*, 47(2):247 – 251, 2004.

[49] W. Huyer and A. Neumaier. MINQ8: general definite and bound constrained indefinite quadratic programming. *Comput. Optim. Appl.*, 69:351 – 381, 2018.

[50] M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, pages 427 – 435, 2013.

[51] W. La Cruz. A spectral algorithm for large-scale systems of nonlinear monotone equations. *Numer. Algor.*, 76(4):1109 – 1130, 2017.

[52] W. La Cruz and M. Raydan. Nonmonotone spectral methods for large-scale nonlinear systems. *Optim. Meth. Softw.*, 18:583 – 599, 2003.

[53] Z. F. Li, M. R. Osborne, and T. Prvan. Adaptive algorithm for constrained least-squares problems. *J. Optim. Theory Appl.*, 114(2):423 – 441, 2002.

[54] J. Liu and Y. Duan. Two spectral gradient projection methods for constrained equations and their linear convergence rate. *J. Inequal. Appl.*, 2015(1):1 – 13, 2015.

[55] J. Liu and Y. Feng. A derivative-free iterative method for nonlinear monotone equations with convex constraints. *Numer. Algor.*, 82(1):245 – 262, 2019.

[56] J. Liu and S. Li. A projection method for convex constrained monotone nonlinear equations with applications. *Comput. Math. Appl.*, 70(10):2442 – 2453, 2015.

[57] N. Mahdavi-Amiri and R. H. Bartels. Constrained nonlinear least squares: An exact penalty approach with projected structured quasi-Newton updates. *ACM Trans. Math. Softw.*, 15(3):220 – 242, 1989.

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

[59] O. L. Mangasarian and R. R. Meyer. Absolute value equations. *Linear Algebra Appl.*, 419(2-3):359 – 367, 2006.

[60] J. M. Martínez, E. A. Pilotta, and M. Raydan. Spectral gradient methods for linearly constrained optimization. *J. Optim. Theory Appl.*, 125:629 – 651, 2005.

[61] K. Meintjes and A. P. Morgan. A methodology for solving chemical equilibrium systems. *Appl. Math. Comput.*, 22(4):333 – 361, 1987.

[62] H. Mohammad. A positive spectral gradient-like method for large-scale nonlinear monotone equations. *Bull. Comput. Appl. Math.*, 5(1):99 – 115, 2017.

[63] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. Testing unconstrained optimization software. *ACM Trans. Math. Softw.*, 7(1):17 – 41, 1981.

[64] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, NY, 2nd edition, 2006.

[65] F. R. Oliveira and O. P. Ferreira. Inexact Newton method with feasible inexact projections for solving constrained smooth and nonsmooth equations. *Appl. Numer. Math.*, 156:63 – 76, 2020.

[66] Y. Ou and Y. Liu. Supermemory gradient methods for monotone nonlinear equations with convex constraints. *Comp. Appl. Math.*, 36(1):259 – 279, 2017.

[67] P. M. Pardalos and M. G. C. Resende, editors. *Handbook of applied optimization*. Oxford University Press, Oxford, 2002.

[68] M. Porcelli. On the convergence of an inexact Gauss-Newton trust-region method for nonlinear least-squares problems with simple bounds. *Optim. Lett.*, 7(3):447 – 465, 2013.

[69] M. J. D. Powell. Log barrier methods for semi-infinite programming calculations. In E. A. Lipitakis, editor, *Advances on Computer Mathematics and its Applications*, pages 1 – 21, 1993.

[70] S. Salzo and S. Villa. Convergence analysis of a proximal Gauss-Newton method. *Comput. Optim. Appl.*, 53(2):557 – 589, 2012.

[71] W. Shen and C. Li. Smale's α-theory for inexact Newton methods under the γ-condition. *J. Math. Anal. Appl.*, 369(1):29 – 42, 2010.

[72] S. Smale. Newton's method estimates from data at one point. In *The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985)*, pages 185 – 196. Springer, New York, 1986.

[73] M. V. Solodov and B. F. Svaiter. *A Globally Convergent Inexact Newton Method for Systems of Monotone Equations*, pages 355 – 369. Springer US, Boston, MA, 1999.

[74] M. V. Solodov and B. F. Svaiter. A new projection method for variational inequality problems. *SIAM J. Control Optim.*, 37(3):765 – 776, 1999.

[75] K.-C. Toh. An inexact primal-dual path following algorithm for convex quadratic SDP. *Math. Program.*, 112(1):221 – 254, 2008.

[76] C. Wang, Q. Liu, and X. Yang. Convergence properties of nonmonotone spectral projected gradient methods. *J. Comput. Appl. Math.*, 182:51 – 66, 2005.

[77] C. Wang and Y. Wang. A superlinearly convergent projection method for constrained systems of nonlinear equations. *J. Glob. Optim.*, 44(2):283 – 296, 2009.

[78] W. Y. Wang, C. and C. Xu. A projection method for a system of nonlinear monotone equations with convex constraints. *Math. Meth. Oper. Res.*, 66(1):33 – 46, 2007.

[79] A. J. Wood and B. F. Wollenberg. Power generation operation and control, published by John Wiley and Sons. *New York, January*, 1996.

[80] K. G. Woodgate. Least-squares solution of $F = PG$ over positive semidefinite symmetric P. *Linear Algebra Appl.*, 245:171 – 190, 1996.

[81] Z. Yu, J. Lin, J. Sun, Y. Xiao, L. Liu, and Z. Li. Spectral gradient projection method for monotone nonlinear equations with convex constraints. *Appl. Numer. Math.*, 59(10):2416 – 2423, 2009.

[82] L. Zhang and W. Zhou. Spectral gradient projection method for solving nonlinear monotone equations. *J. Comput. Appl. Math.*, 196(2):478 – 484, 2006.

[83] W. Zhou and D. Li. Limited memory BFGS method for nonlinear monotone equations. *J. Comput. Math.*, 25(1):89 – 96, 2007.

[84] W.-J. Zhou and D.-H. Li. A globally convergent BFGS method for nonlinear monotone equations without any merit functions. *Math. Comput.*, 77(264):2231 – 2240, 2008.