Universidade Federal de Goiás
IME - Instituto de Matemática e Estatística
Programa de Pós Graduação em Matemática

Yuri Rafael Leite Pereira

Methods for Vector Optimization: Trust Region and Proximal on Riemannian Manifolds and Newton with Variable Order

Doctoral Thesis

Funded by CAPES

Goiânia, Goiás, Brazil
September 20, 2017

sistema de bibliotecas ufg

PRPG
PRÓ-REITORIA DE PÓS-GRADUAÇÃO

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: [ ] Dissertação [X] Tese

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Yuri Rafael Leite Pereira

Título do trabalho: Methods for Vector Optimization: Trust Region and Proximal on Riemannian Manifolds and Newton with Variable Order

3. Informações de acesso ao documento:

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

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

Yuri Rafael Leite Pereira

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

Ciente e de acordo:

Glaydston Carvalho Bento

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

Data: 21 / 09 / 2017

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

Casos de embargo:
- Solicitação de registro de patente
- Submissão de artigo em revista científica
- Publicação como capítulo de livro
- Publicação da dissertação/tese em livro

[2] A assinatura deve ser escaneada.

Yuri Rafael Leite Pereira

Methods for Vector Optimization: Trust Region and Proximal on Riemannian Manifolds and Newton with Variable Order

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. Glaydston de Carvalho Bento
Coorientador: Prof. Dr. Orizon Pereira Ferreira

Goiânia
September 20, 2017

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

Pereira, Yuri Rafael Leite
Methods for Vector Optimization: [manuscrito] : Trust Region and Proximal on Riemannian Manifolds and Newton with Variable Order / Yuri Rafael Leite Pereira. - 2017.
LXII, 62 f.

Orientador: Prof. Dr. Glaydston de Carvalho Bento; co-orientador Dr. Orizon Pereira Ferreira.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2017.
Bibliografia.

1. Vector Optimization. 2. Riemannian Manifolds. 3. Trust Region. 4. Proximal Point method. 5. Newton method. I. Bento, Glaydston de Carvalho, orient. II. Título.

CDU 51

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

ATA DA REUNIÃO DA BANCA EXAMINADORA DA DEFESA DE TESE DE YURI RAFAEL LEITE PEREIRA – Aos vinte e oito dias do mês de agosto do ano de dois mil e dezessete (28/08/2017), às 10:00 horas, reuniram-se os componentes da Banca Examinadora: Prof. Glaydston de Carvalho Bento - Orientador, Prof. Orizon Pereira Ferreira, Prof. Luis Román Lucambio Pérez, Prof. João Xavier da Cruz Neto e Prof. Paulo Sergio Marques dos Santos, sob a presidência do primeiro, e em sessão pública realizada no auditório do Instituto de Matemática e Estatística, procederem a avaliação da defesa de tese intitulada: **"METHODS FOR VECTOR OPTIMIZATION: TRUST REGION AND PROXIMAL ON RIEMANNIAN MANIFOLDS AND NEWTON WITH VARIABLE ORDER"**, em nível de Doutorado, área de concentração em Otimização, de autoria de Yuri Rafael Leite Pereira, discente do Programa de Pós-Graduação em Matemática da Universidade Federal de Goiás. A sessão foi aberta pelo Presidente da Banca, Prof. Glaydston de Carvalho Bento que fez a apresentação formal dos membros da Banca. A seguir, a palavra foi concedida ao autor da tese que, em 45 minutos procedeu a apresentação de seu trabalho. Terminada a apresentação, cada membro da Banca arguiu o examinando, tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se a avaliação da defesa. Tendo-se em vista o que consta na Resolução nº. 1513 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Matemática e procedidas as correções recomendadas, a tese foi **APROVADA** por unanimidade, considerando-se integralmente cumprido este requisito para fins de obtenção do título de **DOUTOR EM MATEMÁTICA**, na área de concentração em Otimização pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega na secretaria do PPGM da versão definitiva da tese, com as devidas correções supervisionadas e aprovadas pelo orientador. Cumpridas as formalidades de pauta, às 12:00 horas a presidência da mesa encerrou esta sessão de defesa de tese e para constar eu, Ulisses José Gabry, secretário do PPGM, lavrei a presente Ata que, depois de lida e aprovada, será assinada pelos membros da Banca Examinadora em quatro vias de igual teor.

Glaydston Carvalho Bento
Prof. Dr. Glaydston de Carvalho Bento
Presidente - IME/UFG

Prof. Dr. Orizon Pereira Ferreira
Membro – IME/UFG

Prof. Dr. Luis Román Lucambio Pérez
Membro – IME/UFG

Prof. Dr. Paulo Sergio Marques dos Santos
Membro – DMCCN/UFPI

João Xavier da C. Neto
Prof. Dr. João Xavier da Cruz Neto
Membro – DMCCN/UFPI

YURI RAFAEL LEITE PEREIRA

METHODS FOR VECTOR OPTIMIZATION: TRUST REGION AND PROXIMAL ON RIEMANNIAN MANIFOLDS AND NEWTON WITH VARIABLE ORDER

Tese defendida no Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Doutor em Matemática, aprovada em 28 de agosto de 2017, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Glaydston de Carvalho Bento
Instituto de Matemática e Estatística - UFG
Presidente da Banca

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

Prof. Dr. Luis Román Lucambio Pérez
Instituto de Matemática e Estatística - UFG

Prof. Dr. João Xavier da Cruz Neto
Centro de Ciências da Natureza - Departamento de Matemática - UFPI

Prof. Dr. Paulo Sergio Marques dos Santos
Centro de Ciências da Natureza - Departamento de Matemática - UFPI

Dedicado a:

À minha Família.

Agradecimentos

Primeiramente gostaria de agradecer à CAPES pelo apoio financeiro, pois sem o auxílio seria inviável a realização do doutorado. Ao meu orientador Prof. Dr. Glaydston de Carvalho Bento, pelos ensinamentos e pela sua grande paciência e compreensão. Ao meu coorientador Prof. Dr. Orizon Pereira Ferreira, pelos ensinamentos e no auxílio na construção dos papers. À Prof(a). Dr(a). Gemayqzel Bouza Allende por ter me orientado no meu doutarado sanduíche durante o período que estive em Havana-Cuba na Universidade de Havana. Aos professores João Xavier da Cruz Neto, Paulo Sérgio Marques dos Santos e Luís Román Lucambio Pérez por aceitarem participar da minha banca de defesa e pelas valiosas sugestões. Aos professores do IME-UFG pelo apoio, em especial aos que integram o grupo de Otimização, e a todos os funcionários do IME-UFG. Aos meus amigos do Piauí que sempre me ajudaram desde o começ do doutorado. Ao Oniel, Letícia e a Prof(a). Aymée que cuidaram de mim na minha estadia em Havana. E agradeço a todos que me ajudaram direta e(ou) indiretamente para a conclusão deste trabalho.

Abstract

In this work, we will analyze three types of method to solve vector optimization problems in different types of context. First, we will present the trust region method for multiobjective optimization in the Riemannian context, which retrieves the classical trust region method for minimizing scalar functions. Under mild assumptions, we will show that each accumulation point of the generated sequences by the method, if any, is Pareto critical. Next, the proximal point method for vector optimization and its inexact version will be extended from Euclidean space to the Riemannian context. Under suitable assumptions on the objective function, the well-definedness of the methods will be established. Besides, the convergence of any generated sequence, to a weak efficient point, will be obtained. The last method to be investigated is the Newton method to solve vector optimization problem with respect to variable ordering structure. Variable ordering structures are set-valued map with cone values that to each element associates an ordering. In this analyze we will prove the convergence of the sequence generated by the algorithm of Newton method and, moreover, we also will obtain the rate of convergence under variable ordering structures satisfying mild hypothesis.

Keywords: Vector Optimization, Multiobjective Optimization, Trust Region, Proximal Point method, Newton method, Riemannian Manifolds, Variable order.

Resumo

Neste trabalho, analisaremos três tipos de métodos para resolver problemas de otimização vetorial em diferentes tipos contextos. Primeiro, apresentaremos o método de Região de Confiança para resolver problemas multiobjetivo no contexto Riemanniano, o qual recupera o método de Região de Confiança clássico para minimizar funções escalares. Sob determinadas suposições, mostraremos que cada ponto de acumulação das sequências geradas pelo método, caso existam, é Pareto crítico. Em seguida, o método do ponto proximal para otimização vetorial e sua versão inexata serão generalizados do espaço Euclidiano para o contexto Riemanniano. Sob suposies adequadas sobre a função objetiva, as boas definições dos métodos serão estabelecidos. Além disso, a convergência de qualquer sequência gerada pelos algoritmos, para um ponto fracamente eficiente, é obtida. O último método a ser investigado é o método de Newton para resolver problemas de otimização vetorial com respeito a estruturas de ordem variável. Estruturas de ordem variável são aplicações ponto-conjunto cujas imagens são cones que induzem uma ordem. Nesta análise, provaremos a convergência da sequência gerada pelo algoritmo do método de Newton e, além disso, também obteremos a taxa de convergência sob estruturas de ordem variável satisfazendo hipóteses adequadas.

Palavras-chave : Otimização Vetorial, Otimização Multiobjetivo, Região de Cofiança , Método do ponto proximal, Variedades Riemannianas, Ordem Variável.

Contents

Chapter 1

Introduction

In mathematical optimization one of the main objectives is minimizing (or maximizing) an objective function over some feasible set. When the objective function is scalar-valued, a feasible element is said to be a optimal solution if attain the smallest (or largest) value of the objective function over the feasible region. However, in case the objective function is vector-valued, it is not obvious anymore how to compare the values of the objective function, usually, no single point will minimize all given objective functions at once. Francis Y. Edgeworth (1845-1926) and Vilfredo Pareto (1848-1923) are credited the first who introduced the optimality concept for vector optimization problems, i.e., for optimization problems with such a vector-valued objective function. Both have studied a particular case of vector optimization problems, called *multiobjective optimization* problems, which is optimization problem with an objective function mapping in the m-dimensional Euclidean space $\mathbb{R}^m$, for $m \geq 2$, i.e., a multiobjective optimization problem can be formulated as

$$\min_{x \in X} F(x),$$

where $F(x) = (f_1(x), \ldots, f_m(x))$, $f_i : \mathbb{R}^n \to \mathbb{R}$ for all$i \in \{1, \ldots m\}$ and X is feasible set. For comparing elements in $\mathbb{R}^m$ they used the natural ordering cone, the nonnegative orthant, which corresponds to the componentwise partial ordering, i.e., given x^1 and $x^2 \in X$, we said that $F(x^1) \preceq F(x^2)$ when $f_i(x^1) \leq f_i(x^2)$ for each $i \in \{1, \ldots m\}$. A point $x^* \in X$ is a optimal point (referred to as Pareto optimal) if there is no other $x \in X$ such that $F(x) \preceq F(x^*)$. In this work, we present three methods to solve vector optimizations problems on different setting.

The first two methods to be studied extend the classical optimization methods to the Riemannian context to solve vector optimization problems. In the last two decades it has appeared a significant number of papers dealing with extensions of optimization methods to the Riemannian context; see [11, 12, 15, 26, 32, 68]. One of the motivations to study this subject is that many optimization problems are naturally posed in this context; see [2,58,61]. Besides, in this setting, it is possible to deal with constrained problems as unconstrained one from the point of view Riemannian; see [36,49]. Over the past few years have appeared some papers on multiobjective optimization problems in the Riemannian context; see [11,12,15,19].

The first method to be worked is the *Trust Region* method for multiobjective optimization. The Trust Region method is recognized as one of the most efficient techniques of global convergence of the Newton Method for nonlinear optimization, which origin can be traced back to Levenberg [44]. This method became popular due to its attractive convergence properties under mild assumptions, for instance, global convergence without assuming any convexity assumption. Early works dealing with this subject include, but are not limited to [27, 52, 55, 56, 64, 65]. In [76] Yuan has been presented a perspective historical and a comprehensive description of the recent advances in this method. Unlikely line search algorithms, where a line search is carried out along a search direction in each iteration, the trust region algorithms obtain the new iterate point by solving a subproblem restricted to a neighborhood centered at the current iterate point, this neighborhood is called a trust region To be more exact, at the *kth* iteration, a trust region algorithm for the general optimization problem

$$\min_{x\in X} f(x),$$

where $f : X \to \mathbb{R}$ is the objective function to be minimized and $X \subset \mathbb{R}^n$ is the feasible set, obtains the trial step s^k by solving the following trust region subproblem

$$\min\left\{m_k(s) : \; s \in X, \; \|s\| \le \Delta_k\right\}, \tag{1.1}$$

where $m_k(s)$ is a model function that approximates the objective function $f(x^k + s)$ near the current iteration point x^k and $\Delta_k > 0$ is the trust region radius. Given a trial step s^k we define the trust ratio

$$r_k = \frac{f(x^k) - f(x^k + s^k)}{m_k(0) - m_k(s^k)},$$

then we verify the agreement between the numerator, called actual reduction, and denominator, called the predicted reduction. Depending on the criteria to be used, if trust ratio is greater than a certain constant the trial step will be accepted, otherwise, the trial step must be reject and the trust region should be shrink by reducing Δ_k, then remake the subproblem (1.1).

Recently this method was extended to solve multiobjective problems. In this new context some distinct quantities, namely, the quadratic model functions and the trust rates, have been introduced in order to design the method; see [21, 40, 57, 70]. Furthermore, the trust region methods were extended to solve scalar problems in the Riemannian context, and its main features were preserved, including its global convergence properties; see [1, 41]. As a natural generalization what is done in the scalar Riemannian version of the trust region method; see [1], the model in our multiobjective version is the maximum of the quadratic models involving the coordinates of the objective function. Our method iteratively obtain a trial step by solving inexactly an optimization problem involving the quadratic model within a trust region in the tangent space at the current iterate. Then, next iterate is obtained by using the exponential mapping in the tangent space at the current iterate evaluated at the trial step and judging this point using the trust rate, the same trust rate present in [70], the formal description of the method is done in Chapter 3. Under mild assumptions, we show that each accumulation point of a generated sequence by the method, if any, is Pareto critical, which retrieves the same convergence results of the scalar Euclidean version; see [25, 53].

The next method to be extended to the Riemannian context is the *Proximal Point* method to solve vector optimization problem. As an extension of the nonnegative orthant for comparing elements in multiobjective optimization, in vector optimization it is often assumed that the partial order of a linear space is induced by a general closed convex pointed cone; see [43, 48]. Let $F : X \to Y$ a vector-value function, where X and Y are to linear space, and C be a closed convex pointed cone which induce a partial order in Y, given x^1 and x^2, we said that $F(x^1) \preceq_C F(x^2)$ if $F(x^2) - F(x^1) \in C$. An element $x^* \in X$ is said to be a minimal element (efficient element) if there is no $x \in X$ such that $F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. The classical Proximal Point method, to solve the minimization problem

$$\min_{x \in X} f(x),$$

introduced by Martinet [50] and Rockafellar [59], generates a sequence $\{x^k\} \subset X$ as follows: given $x_0 \in X$ and some $\hat{\lambda} > 0$

$$x^{k+1} = \operatorname*{argmin}_{x \in X} \left\{ f(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \right\},$$

where the sequence of real numbers $\{\lambda_k\}$ satisfies $0 < \lambda_k \leq \hat{\lambda}$. In the last years researchers began extend the proximal point method to solve to vector optimization see [13, 18, 22, 69]. The method to be developed in this work deals with the extension of the proximal point method for vector optimization from the Euclidean settings to the Riemannian context, which continues the subject addressed in the following papers [14, 16, 32, 45, 72]. To our best knowledge, this is the first extending of the proximal point method and its inexact version for vector optimization to the Riemannian settings. Besides our approach is new even in Euclidean context, since we are dealing with general convex cone and the nonlinear and variable scalarization which is more flexible than the one considered in Bento et al. [13]. Under suitable assumptions on the objective function, the well-definedness of the method and its inexact version are established. Besides, the convergence of any generated sequence, to a weak efficient point, is obtained. It is worth pointing out that under assumption of null sectional curvature, our algorithm retrieves the proximal point method for multiobjective presented in [13] and, somehow, goes further.

The last method to be worked is the *Newton method For Variable Order Vector Optimization Problems*. The idea of variable ordering structures was given by Yu in [75] where it was assumed that it is a set-valued map with cone values that associates with each element of the linear space and ordering, furthermore, Yu defined the notion of nondominated element, a candidate element is called a nondominated element if it is not dominated by their reference elements with respect to the corresponding ordering of these other ones, i.e, given the set-valued application $K : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ where $K(y)$ is a cone which define a partial ordering, it is said that $y^* \in \mathbb{R}^m$ is a *nondominated element* if there is no $y \in \mathbb{R}^m / \{y^*\}$ such that $y - y^* \in K(y)$. In [4] Bao et al. study some necessary conditions for nondominated solutions with respect to variable ordering structures. But in this work we use another notion of optimal elements in variable ordering structures, called *efficient* or *minimal element*, it is said that $y^* \in \mathbb{R}^m$ is a minimal element if there is no $y \in \mathbb{R}^m/\{y^*\}$ such that $y - y^* \in K(y^*)$. It is easy to see, that if $K(\cdot)$ is a constant set-valued application, then the concepts of nondominated solutions and minimal element retrieve the concept of efficient solutions in the case vector optimization and in case the image of $K(\cdot)$ is the ortant non-negative the concept of Pareto minimal. In [28] Eichfelder shows that the nondominated element and minimal element are are connected by duality properties, moreover she proves their characterizations and existence results.

The first applications of variable ordering optimization models appear in Economy and Medicine, see [3,31,73]. They motivated the characterization of the solutions via necessary or sufficient optimality conditions, solution approaches and further applications in Behavioral Sciences, see [5,6,28–30,47,62,63].

Recently, some works have appeared generalizing the classical methods of optimization for vector optimization problem with variable order structures. An extension of the steepest descend method has been proposed in Bello-Cruz et al. [9], in addition they introduced the concept of convexity function for the case variable order. In [10] Bello-Cruz et al. present the extension of the subgradient method with a different structure of variable order present in [9]. In this work, using the variable order present in [9] and their concept of convexity, we generalize Newton method proposed in Fliege et al. [33] and Graña-Drummond et al. [38], for the case vector optimization problem with respect to with variable ordering structures. To this aim we extend the concept of strong convexity and prove some properties of the that in this case Newton's method is well defined and will converge to a solution of the problem. Variable orders under which the rate of convergence is super-linear and under which is quadratic are given.

Chapter 2

Preliminaries results on riemannian geometry

In this section, we introduce some notations about Riemannian geometry and some results related to convex analysis in this setting, which will be used throughout the Chapters 3 and 4.

2.1 Notation and terminology

In this section, we introduce some notations about Riemannian geometry, which can be found in any introductory book on Riemannian geometry, such as in Sakai [60], Do Carmo [20]. From now on, let M be an n-dimensional connected manifold. We denote by T_pM the n-dimensional *tangent space* of M at p, and by $TM = \cup_{p\in M} T_pM$ *tangent bundle* of M and by $\mathcal{X}(M)$ the space of smooth vector fields over M. Suppose that M be endowed with a Riemannian metric $\langle \, , \, \rangle$, with corresponding norm denoted by $\| \; \|$, that is, M is a Riemannian manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining p to q, i.e., such that $\gamma(a) = p$ and $\gamma(b) = q$, by

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt,$$

and, moreover, by minimizing this functional length over the set of all such curves, we obtain a Riemannian distance $d(p, q)$, which induces the original topology on M, and, given a nonempty set $U \subset M$, we can define the distance function associated to U, which is given by:

$$M \ni p \mapsto d(p, U) := \inf\{d(p, q) \ : \ q \in U\}.$$

Let $f : M \to \mathbb{R}$ a differential function, the metric induces a map

$$f \mapsto \operatorname{grad} f \in \mathcal{X}(M),$$

which associates to each scalar function smooth over M its gradient via the rule $\langle \operatorname{grad} f, X\rangle = df(X),\ X \in \mathcal{X}(M)$. If f is a two differentiable function we define, for any $p \in M$, the hessian matrix of the function f as the map $\mathrm{Hess} f(p) : T_pM \to T_pM$. Let ∇ be the Levi-Civita connection associated to $(M, \langle\, ,\, \rangle)$. A vector field V along γ is said to be *parallel* iff $\nabla_{\gamma'} V = 0$. If γ' itself is parallel we say that γ is a *geodesic*. A geodesic $\gamma = \gamma_v(., p)$ is determined by its position p and velocity v at p. We say that γ is *normalized* if $\|\gamma'\| = 1$. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. A geodesic segment joining p to q in M is said to be *minimal* iff its length is equals to $d(p, q)$ and, in this case, the geodesic is called a *minimizing geodesic*. If γ is a curve joining points p and q in M then for each $t \in [a, b]$, ∇ induces a linear isometry, relative to $\langle\, ,\, \rangle$, $P_{\gamma,\gamma(a),\gamma(t)} : T_{\gamma(a)}M \to T_{\gamma(t)}M$, the so-called parallel transport along γ from $\gamma(a)$ to $\gamma(t)$. The inverse map of $P_{\gamma,\gamma(a),\gamma(t)}$ is denoted by $P^{-1}_{\gamma,\gamma(a),\gamma(t)} : T_{\gamma(t)}M \to T_{\gamma(a)}M$. In the particular case of γ is the unique curve joining points p and q in M then parallel transport along γ from p to q is denoted by $P_{\gamma,p,q} : T_pM \to T_qM$. Next we present the definition of Lipschitz continuously of vector field on Riemannian manifolds firstly presented in [26].

Definition 2.1.1 *The gradient vector field of a differentiable function $f : M \to \mathbb{R}$ is said to be Lipschitz continuous with constant $L \geq 0$ if there holds the inequality $\|P_{\gamma,p,q} \operatorname{grad} f(p) - \operatorname{grad} f(q)\| \leq L\, l(\gamma)$, for any $p, q \in M$ and γ any geodesic segment joining p to q.*

Proposition 2.1.2 *If the Hessian of the function $f : M \to \mathbb{R}$ is uniformly bounded, i.e., $\|\mathrm{Hess}\, f(p)\| \leq \beta$, for all $p \in M$, then the gradient vector field of f is Lipschitz continuous with constant β.*

Proof. Let $p, q \in M$ and $\gamma : [0, 1] \to M$ be a geodesic segment such that $\gamma(0) = q$ to $\gamma(1) = p$. From the fundamental theorem of calculus we have

$$P_{\gamma,p,q} \operatorname{grad} f(p) - \operatorname{grad} f(q) = \int_0^1 P_{\gamma,\gamma(t),q} \mathrm{Hess}\, f(t)\gamma'(t)dt.$$

Taking norm in both sides of the last equality, using properties of the transport parallel and the uniformly boundedness of Hessian, we get $\|P_{\gamma,p,q} \operatorname{grad} f(p) - \operatorname{grad} f(q)\| \leq \beta l(\gamma)$. Thus the desired result follows.

■

A Riemannian manifold is *complete* iff geodesics are defined for any values of t. Hopf-Rinow Theorem asserts that, if this is the case then any pair of points, say p and q, in M can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, (M, d) is a complete metric space and bounded and closed subsets are compact. From the completeness of the Riemannian manifold M, for each $p \in M$, the *exponential map* $\exp_p : T_pM \to M$ can be defined by $\exp_p v = \gamma_v(1, p)$, where $\gamma_v(\cdot, p) : \mathbb{R} :\to M$ denotes the unique geodesic such that $\gamma_v(0, p) = p$ and $\gamma_v'(0, p) = v$.

We denote by R the curvature tensor defined by $R(X, Y) = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$ with X, Y, $Z \in \mathcal{X}(M)$ where $[X, Y] = Y X - XY$ Moreover, the sectional curvature with respect to X and Y is given $K(X,Y) = \langle R(X,Y)Y, X\rangle/(\|X\|^2\|Y\|^2 - \langle X,Y\rangle^2)$ where $\|X\| = \langle X,Y\rangle^{1/2}$. If $K(X,Y) \leq 0$ for all X and Y, then M is called a Riemannian manifold of nonpositive curvature and we use the short notation $K \leq 0$.

Theorem 2.1.3 *Let M be a complete,n-dimensional, simply connected Riemannian manifold with nonpositive sectional curvature. Then M is diffeomorphic to the Euclidean space $\mathbb{R}^n$. More precisely, at any point $p \in M$, the exponential map $\exp_p$ is a diffeomorphism.*

Proof. See [20, Lemma 3.2 p. 149] or [60, Theorem 4.1 p. 221] ■

A complete simply connected Riemannian manifold of nonpositive sectional curvature is called a Hadamard manifold. Thus Theorem 2.1.3 states that if M is a *Hadamard manifold* then M has the same topology and differential structure of the Euclidean space $\mathbb{R}^n$. Furthermore, there are known some similar geometrical properties of the Euclidean space $\mathbb{R}^n$ such as, given two points there exists an unique geodesic segment that joins them. Hence, if M is a Hadamard manifold the lenght of the geodesic segment γ joining p to q is equals $d(p,q)$. Moreover, the *exponential map* $exp_p v = \gamma_v(1,p)$ is a diffeomorphism. Let $q \in M$ and $exp_q^{-1} : M \to T_qM$ be the inverse of the exponential map. Note that $d(q\,,\,p) = ||exp_p^{-1} q||$, the function $d_q^2 : M \to \mathbb{R}$ defined by $d_q^2(p) = d^2(q,p)$ is C^∞ and $\operatorname{grad} d_q^2(p) := -2exp_q^{-1}p$.

A geodesic triangle $\Delta(p_1,p_2,p_3)$ of a Riemannian manifold is the set consisting of three distinct points p_1,p_2,p_3 called the vertices and three minimizing geodesic segments γ_{i+1} joining p_{i+1} to p_{i+2} called the sides, where $i = 0,1,2$.

Theorem 2.1.4 *Let M be a Hadamard manifold $\Delta(p_1,p_2,p_3)$ a geodesic triangle. Denote by $\gamma_{i+1} : [0,\ell_{i+1}] \to M$ geodesic segments joining p_{i+1} to p_{i+2} and set $\ell_i = L(\gamma_i)$, $\theta_i = \angle(\gamma'_{i+1}(0), -\gamma'_i(\ell_i))$ where $i \equiv 1,2,3 (\operatorname{mod} 3)$. Then*

$$\theta_1 + \theta_2 + \theta_3 \leq \pi;$$

$$\ell_{i+1}^2 + \ell_{i+2}^2 - 2\ell_{i+1}\ell_{i+2}\cos\theta_i \leq \ell_i^2. \tag{2.1}$$

As $\ell_i = d(p_{i+1},p_{i+2})$, we can rewrite (2.1) as follows

$$d^2(p_i,p_{i+2}) + d^2(p_i,p_{i+1}) - d(p_i,p_{i+2})d(p_i,p_{i+1})\cos\theta_i \leq d^2(p_{i+1},p_{i+2}),$$

where $\theta_i = \angle(\exp_{p_i}^{-1} p_{i+1}, \exp_{p_i}^{-1} p_{i+2})$.

Since $d(p_i,p_{i+2})d(p_i,p_{i+1})\cos\theta_i = \langle \exp_{p_i}^{-1} p_{i+1}, \exp_{p_i}^{-1} p_{i+2}\rangle$, then we have

$$d^2(p_i,p_{i+2}) + d^2(p_i,p_{i+1}) - \langle \exp_{p_i}^{-1} p_{i+1}, \exp_{p_i}^{-1} p_{i+2}\rangle \leq d^2(p_{i+1},p_{i+2}). \tag{2.2}$$

Proof. See [60, Proposition 4.5 p. 223] ■

Since (M,d) is a complete metric space, we recall the notion of Fejér convergence in this context. A sequence $\{p^k\} \subset (M,d)$ is said to be *Fejér convergence* to a nonempty set $W \subset M$ iff $d^2(q,p^{k+1}) \leq d^2(q,p^k)$, for every $q \in W$. We end this section with the following result, whose proof can be found in [32].

Proposition 2.1.5 *Let $\{p^k\}$ be a sequence in (M,d). If $\{p^k\}$ is Fejér convergent to non-empty set $W \subset M$, then $\{p^k\}$ is bounded. If furthermore, an accumulation point p of $\{p^k\}$ belongs to W, then $\lim_{k\to\infty} p^k = p$.*

2.2 Convexity

In this section, some results related to convex analysis in the Riemannian context are presented, which can be found in [45,46,68].

A set $\Omega \subseteq M$ is said to be *convex* if any geodesic segment with end points in Ω is contained in Ω, that is, if $\gamma : [a,b] \to M$ is a geodesic such that $x = \gamma(a) \in \Omega$ and $y = \gamma(b) \in \Omega$; then $\gamma((1-t)a + tb) \in \Omega$ for all $t \in [0,1]$. Let $g : M \to \mathbb{R} \cup \{+\infty\}$ be a function. The *domain* of g is the set $\operatorname{dom} g := \{p \in M \ : \ g(p) < \infty\}$. The function g is said to be proper if $\operatorname{dom} g \neq \varnothing$ and *convex* (resp. *strictly convex, strongly convex*) on a convex set $\Omega \subset \operatorname{dom} g$ if for any geodesic segment $\gamma : [a,b] \to \Omega$ the composition $g \circ \gamma : [a,b] \to \mathbb{R}$ is convex (resp. *strictly convex, , strongly convex*). It is well known that, for each $q \in M$, d_q^2 is strongly convex. Take $p \in \operatorname{dom} g$. A vector $s \in T_pM$ is said to be a *subgradient* of g at p, if $g(q) \geq g(p) + \langle s, \exp_p^{-1} q\rangle$, for $q \in M$. The set $\partial g(p)$ of all subgradients of g at p is called the *subdifferential* of g at p. If g is a differentiable function at $p \in M$ then $\partial g(p) = \{\operatorname{grad} g(p)\}$. The function g is *lower semicontinuous* (lsc) at $\overline{x} \in \operatorname{dom} g$ if for each sequence $\{x^n\}$ converging to $\overline{x}$ we have $\liminf_{n\to\infty} g(x^n) \geq g(\overline{x})$. Let $\Omega \subset M$ be a closed and convex set. The indicator function of Ω, $I_\Omega : M \to \mathbb{R} \cup \{+\infty\}$, is a lower semicontinuous function and, for each $p \in \Omega$, $\partial I_\Omega(p) = N_\Omega(p) \neq \varnothing$, where N_Ω denotes the *normal cone* defined by

$$N_\Omega(p) := \left\{ v \in T_pM \ : \ \langle v, \ \exp_p^{-1} q\rangle \leq 0, \ q \in \Omega \right\}.$$

The next result is the additivity rule for subdifferential, which is a particular instance of [46, Proposition 4.3]. Indeed, since we are in the Hadamard setting, the concept of convex set that we are using merges into the concept of weakly convex set in [46, Definition 3.7].

Proposition 2.2.1 *Let $g_1, g_2 : M \to \mathbb{R} \cup \{+\infty\}$ be proper and convex functions such that* $\operatorname{dom} g_1 \cap \operatorname{dom} g_2$ *is convex. Then, there holds*

$$\partial(g_1 + g_2)(p) = \partial g_1(p) + \partial g_2(p), \ \ p \in \operatorname{int}(\operatorname{dom} g_1) \cap \operatorname{dom} g_2.$$

In particular, for a nonempty and convex set $\Omega \subset M$ and a convex function $g : M \to \mathbb{R} \cup \{+\infty\}$ such that $\operatorname{int}(\operatorname{dom}) g \neq \varnothing$ *and* $\Omega \cap \operatorname{dom} g$ *is convex, it follows that*

$$\partial(g + I_\Omega)(p) = \partial g(p) + N_\Omega(p), \ \ p \in \operatorname{int}(\operatorname{dom} g) \cap \Omega.$$

In [7] Batista et. al. introduced the concept of the enlargement of monotone vector fields, moreover they presented an inexact proximal point method for constrained optimization where they defined the ϵ-subdifferential of $g : M \to \mathbb{R}$ as the set-valued function $\partial_\epsilon g : M \to TM$, which is given by

$$\partial_\epsilon g(p) := \{u \in T_pM \ : \ g(q) \geq g(p) + \langle u, \exp_p^{-1} q\rangle - \epsilon, \ q \in M\},$$

where $\epsilon > 0$. One of its properties that is easy to see is $\partial g(p) \subset \partial_\epsilon g(p)$ for all $p \in M$

Chapter 3

A trust region method for unconstrained mutiobjective problems on Riemannian manifolds

3.1 The multiobjective problem

In this section, we present the multiobjective problem, the first order optimality condition for it and some basic definitions. Let $I := \{1, \ldots, m\}$, $\mathbb{R}^m_+ = \{y \in \mathbb{R}^m \ : \ y_i \geq 0, j \in I\}$ and $\mathbb{R}^m_{++} = \{y \in \mathbb{R}^m \ : \ y_j > 0, j \in I\}$. For y, $z \in \mathbb{R}^m$, $z \succeq y$ (or $y \preceq z$) means that $z - y \in \mathbb{R}^m_+$ and $z \succ y$ (or $y \prec z$) means that $z - y \in \mathbb{R}^m_{++}$. Let M be a Riemannian manifold and $F : M \to \mathbb{R}^m$ a vector function. We denote the *multiobjective unconstrained optimization problem* in the Riemannian context as

$$\min\{F(p) \ : \ p \in M\}. \tag{3.1}$$

A point $p^* \in M$ is an *optimum Pareto point* for (3.1) if there exists no other $p \in M$ with $F(p) \preceq F(p^*)$ and $F(p) \neq F(p^*)$. Letting the jacobian of F at $p \in M$ be denoted by $JF(p)$, the first order optimality condition for (3.1) is stated as

$$\mathrm{Im}(JF(p)) \cap (-\mathbb{R}^m_{++}) = \emptyset, \qquad p \in M. \tag{3.2}$$

A point satisfying (3.2) will be called a *Pareto critical point* for the problem in (3.1).

Let $F := (f_1, \ldots, f_m)$ be a differentiable function and define the auxiliary function $\omega : M \to \mathbb{R}$ by

$$\omega(p) := -\min\left\{\max_{i\in I}\langle \mathrm{grad}\, f_i(p), \nu\rangle \ : \ \nu \in T_pM, \ \|\nu\| \leq 1\right\}. \tag{3.3}$$

For $M = \mathbb{R}^n$ the function (3.3) retrieves the correspondent one in Svaiter [34]. In order to recall some properties of ω, we define $M \ni p \mapsto D(p) \subset T_pM$ as follows

$$D(p) := \operatorname{argmin}\left\{\max_{i\in I}\langle \operatorname{grad} f_i(p), \nu\rangle \ : \ \nu \in T_pM,\ \|\nu\| \le 1\right\}.$$

In the next result, we state some properties of the functions ω and D.

Lemma 3.1.1 *The following statements hold:*

(i) *If p is Pareto critical point, then $0 \in D(p)$ and $\omega(p) = 0$;*

(ii) *If p is not Pareto critical point, then $\omega(p) > 0$;*

(iii) *If $f_i : M \to \mathbb{R}$ has Lipschitz continuous gradient with constant $L_i \ge 0$, for each $i \in I$. Then, ω is Lipschitz continuous with constant $L = \max_{i\in I} L_i$, i.e., $|\omega(p) - \omega(q)| \le L\, d(p,q)$, for all $p, q \in M$.*

Proof. Assume that $p \in M$ is Pareto critical point. Thus (5.2) implies that, for each $\nu \in T_pM$, there exists $i \in I$ such that $\langle \operatorname{grad} f_i(p), \nu\rangle \ge 0$, which implies $\max_{i\in I}\langle \operatorname{grad} f_i, \nu\rangle \ge 0$. Since $0 \in T_pM$ the proof of item (i) is concluded. Now, let us suppose that $p \in M$ is not Pareto critical point. Then, there exists $\nu \in T_pM$ such that $\langle \operatorname{grad} f_i(p), \nu\rangle < 0$ for all $i \in I$ and thus $\omega(p) > 0$. This, together with item (i), proof (ii). For proving the item (iii), let p, q in M, $\gamma : [0,1] \to M$ any geodesic joining p to q and $P_{\gamma,p,q} : T_pM \to T_qM$ the parallel transport along γ. Let $\nu_q \in D(q)$. Hence, (3.3) implies

$$\begin{aligned} -\omega(p) &\le \max_{i\in I}\left\langle \operatorname{grad} f_i(p),\ P^{-1}_{\gamma,p,q}\nu_q\right\rangle \\ &= -\omega(q) + \max_{i\in I}\left\langle \operatorname{grad} f_i(p),\ P^{-1}_{\gamma,p,q}\nu_q\right\rangle - \max_{i\in I}\left\langle \operatorname{grad} f_i(q),\ \nu_q\right\rangle. \end{aligned}$$

After some algebraic manipulations, the latter expression becomes

$$\omega(q) - \omega(p) \le \max_{i\in I}\left\langle \operatorname{grad} f_i(p) - P^{-1}_{\gamma,p,q}\operatorname{grad} f_i(q),\ P^{-1}_{\gamma,p,q}\nu_q\right\rangle.$$

Using Cauchy-Schwarz inequality and taking into account that the parallel transport is an isometry and $\|v_q\| \le 1$, we obtain from last inequality that $\omega(q) - \omega(p) \le \max_{i\in I}\|P_{\gamma,p,q}\operatorname{grad} f_i(p) - \operatorname{grad} f_i(q)\|$. On other hand, by using similar argument and using that the parallel transport is an isometry, it easy to see that for $\nu_p \in D(p)$ there holds $\omega(p) - \omega(q) \le \max_{i\in I}\|P_{\gamma,p,q}\operatorname{grad} f_i(p) - \operatorname{grad} f_i(q)\|$. Combining two last inequalities, using that $f_i : M \to \mathbb{R}$ has Lipschitz continuous gradient with constant $L_i \ge 0$, for $i \in I$ and $L = \max_{i\in I} L_i$, from Proposition 2.1.2 we obtain

$$|\omega(p) - \omega(q)| \le L l(\gamma),$$

for any geodesic segment γ joining p to q, in particular we can take the geodesic segment γ such that $l(\gamma) = d(p,q)$, then the desired result follows. ∎

3.2 The trust region method

In this section, we introduce the trust region method for solving (3.1). As a natural generalization what is done in the scalar version of the method, see [1], which establishes a quadratic model involving the objective function in the tangent space at the current iterate in the multiobjective version the model is the maximum of the quadratic models involving the coordinate of the objective function. More specifically, for the current iterate $p \in M$ the model $m_p : T_pM \to \mathbb{R}$ is defined by

$$m_p(v) := \max_{i\in I} \left\{ f_i(p) + \langle \mathrm{grad} f_i(p), \nu\rangle + \frac{1}{2}\left\langle \mathrm{Hess} f_i(p)\nu, \nu\right\rangle \right\}. \tag{3.4}$$

The method iteratively obtain the *trial step* $\nu_p \in T_pM$ solving inexactly the problem

$$\min\left\{m_p(\nu) \ : \ \nu \in T_pM, \ \|\nu\| \leq \Delta_p\right\},$$

where $\Delta_p > 0$ denotes the *trust radius* associated to p. Obtained the trial step ν_p, the decision to accept or not this candidate, and update the trust radius Δ_p, is based on the *trust ratio* $r(\nu_p)$ given by

$$r(\nu_p) := \frac{\max_{i\in I} f_i(p) - \max_{i\in I} f_i(\exp_p \nu_p)}{m_p(0) - m_p(\nu_p)}. \tag{3.5}$$

Since the trial step ν_p is obtained by minimizing the model m_p over the region defined by the trust radius Δ_p that includes $v = 0$, the quantity $m_p(0) - m_p(\nu_p)$ will always be nonnegative. Hence, if $r(\nu_p)$ is negative or close to zero, then we have a poor correlation between the model m_p and the function $\max_{i\in I} f_i$ over this step. In this case, the trial step ν_p must be rejected and the trust radius is reduced at the next iteration. On the other hand, if $r(\nu_p)$ is close to 1, then there is a good correlation between the model m_p and the function $\max_{i\in I} f_i$ over this step and, consequently, the step must be accepted and it is safe to increase the radius in the next iteration. If $r(\nu_p)$ is positive and smaller than 1, but not as close to zero, then we accepted the step, but do not alter the trust radius. The method describing the above process is as follows:

Multiobjective Trust Region Method

Step 0. Let $\bar{\Delta} > 0$, $0 < c \leq 1/2$, $0 < r' < 1/4$, $p_0 \in M$, $\Delta_0 \in (0, \bar{\Delta})$ and set $k = 0$;

Step 1. Given p^k and $\Delta_k := \Delta_{p^k}$, computes $m_k := m_{p^k}$ using (3.4) and $\nu^k := \nu^{p_k}$ a solution of the problem

$$\min\left\{m_k(v) \ : \ \nu \in T_{p^k}M, \ \|\nu\| \leq \Delta_k\right\}, \tag{3.6}$$

according to the following error criterium

$$m_k(0) - m_k(\nu_k) \geq c\omega(p^k) \min\left\{\Delta_k, \frac{\omega(p^k)}{\max_{i\in I}\|Hessf_i(p^k)\|}\right\}. \tag{3.7}$$

Compute the trust ratio $r_k := r(\nu^k)$ using (3.5). Define the next iterate p^{k+1} by

$$p^{k+1} := \begin{cases} \exp_{p^k} \nu^k, & \text{if } r_k > r', \\ p^k, & \text{otherwise,} \end{cases} \tag{3.8}$$

and computes the trust radius Δ_{k+1} as follows:

$$\Delta_{k+1} := \begin{cases} \dfrac{\Delta_k}{4}, & \text{if} \quad r_k < \dfrac{1}{4}, \\ \min\{2\Delta_k, \bar{\Delta}\}, & \text{if} \quad r_k > \dfrac{3}{4} \text{ and } \|\nu_k\| = \Delta_k, \\ \Delta_k, & \text{otherwise;} \end{cases} \tag{3.9}$$

Step 2. set $k \leftarrow k+1$ and go to step 1.

end

Remark 3.2.1 The multiobjective trust region method becomes the classical Euclidean trust region method when $m = 1$; see [53, page 69]. We also remark that the quadratic model (3.4) generalize to the Riemannian context the corespondent Euclidean one of [57]. However, the trust ratio (3.5) not merge into the one used in [21, 57]. Finally, note that the sequence $\{\max_{i\in I} f_i(p^k)\}$ is non-increasing, where $\{p^k\}$ is generated from the above method. Indeed, given an iterate p^k, if the trust ratio $r_k \leq r'$, from (3.8), we have $p^{k+1} = p^k$. Otherwise, if the trust ratio $r_k > r'$, then from (3.5) we obtain $\max_{i\in I} f_i(p^k) - \max_{i\in I} f_i(p^{k+1}) > r'(m_k(0) - m_k(\nu^k)) \geq 0$, which prove that $\{\max_{i\in I} f_i(p^k)\}$ is in fact non-increasing.

In the scalar case one of strategy to obtaining the global convergence of the method is that the trial step produces a reduction which is at least a fraction of the reduction produced by the Cauchy step, i.e., a minimizer of the model quadratic along the steepest descent direction within the trust region. This method is the determination of the step ν_p that sufficiently reduces the model within the trust region. It is shown in [56] that the trust region method is convergent if the trial step satisfies a sufficient reduction of the model function. Based on that, it is enough, for purposes of global convergence, accept an approximate solution ν_k that lies within the trust region and satisfies a sufficient reduction defined in (3.7). It is worth to mentioning that an exact solution of (3.6) satisfies (3.7) to $c = 1/2$. Therefore, the method is always well-defined. In the next lemma we show that (3.7) is also satisfied for the Cauchy step with $c = 1/2$.

Lemma 3.2.2 *Let $p \in M$, $\nu \in D(p)$ and the trust radius $\Delta_p > 0$. Then, the following inequality holds*

$$m_p(0) - \min_{\alpha \in [0,\Delta_p]} m_p(\alpha\nu) \geq \frac{1}{2}\omega(p) \min\left\{\Delta_p,\ \frac{\omega(p)}{\max_{i\in I} \|\mathrm{Hess} f_i(p)\|}\right\}.$$

Proof. Let $\alpha \in [0, \Delta_p]$ and $\nu \in D(p)$. From (3.4) it follows that

$$m_p(0) - m_p(\alpha\nu) = \max_i f_i(p) - \max_{i\in I}\left\{f_i(p) + \alpha \langle \mathrm{grad} f_i(p), \nu\rangle + \frac{\alpha^2}{2}\langle \mathrm{Hess} f_i(p)\nu, \nu\rangle\right\}.$$

Using properties of the maximum and Cauchy-Schwartz inequality, the last equality becomes

$$m_p(0) - m_p(\alpha\nu) \geq -\alpha \max_{i\in I}\langle \mathrm{grad} f_i(p), \nu\rangle - \frac{\alpha^2}{2}\|\nu\|^2 \max_{i\in I}\|\mathrm{Hess} f_i(p)\|.$$

Since $\nu \in D(p)$, using the definition of $\omega(p)$ in (3.3), we obtain, for any $\alpha \in [0, \Delta_p]$

$$\max_{\alpha\in[0,\Delta_p]}[m_p(0) - m_p(\alpha\nu)] \geq \max_{\alpha\in[0,\Delta_p]}\left[\alpha\omega(p) - \frac{\alpha^2}{2}\max_{i\in I}\|\mathrm{Hess} f_i(p)\|\right]. \tag{3.10}$$

Now, analyzing the right side of the inequality above, realize the function $\alpha \mapsto \alpha\omega(p) - \alpha^2/2 \max_{i\in I}\|\mathrm{Hess} f_i(p)\|$ is concave and non-negative, then we need verifying the two possibilities for the maximum above. First, we denote the unconstrained maximizer of this function by

$$\alpha^* = \frac{\omega(p)}{\max_{i\in I}\|\mathrm{Hess} f_i(p)\|} \geq 0, \tag{3.11}$$

and with maximum value equal to

$$\frac{1}{2}\frac{\omega(p)}{\max_{i\in I}\|\mathrm{Hess} f_i(p)\|} \geq 0.$$

Then, we have two possibilities

a) If $0 \leq \alpha^* < \Delta_p$, then

$$\max_{\alpha\in[0,\Delta_p]}\left[\alpha\omega(p) - \frac{\alpha^2}{2}\max_{i\in I}\|\mathrm{Hess} f_i(p)\|\right] = \frac{1}{2}\frac{\omega(p)}{\max_{i\in I}\|\mathrm{Hess} f_i(p)\|} \geq 0. \tag{3.12}$$

b) If $\alpha^* \geq \Delta_p$, then

$$\max_{\alpha\in[0,\Delta_p]}\left[\alpha\omega(p) - \frac{\alpha^2}{2}\max_{i\in I}\|\mathrm{Hess} f_i(p)\|\right] = \Delta_p\omega(p) - \frac{\Delta_p^2}{2}\max_{i\in I}\|\mathrm{Hess} f_i(p)\|.$$

From (3.11) and item b) we have $\omega(p)/\max_{i\in I}\|\mathrm{Hess}f_i(p)\| \geq \Delta_p$, as a consequence

$$\Delta_p\omega(p) - \frac{\Delta_p^2}{2}\max_{i\in I}\|\mathrm{Hess}f_i(p)\| \geq \Delta_p\omega(p) - \frac{\Delta_p}{2}\omega(p) = \frac{\Delta_p}{2}\omega(p). \tag{3.13}$$

Hence, from (3.12) and (3.13) we obtain

$$\max_{\alpha\in[0,\Delta_p]}\left[\alpha\omega(p) - \frac{\alpha^2}{2}\max_{i\in I}\|\mathrm{Hess}f_i(p)\|\right] \geq \frac{1}{2}\omega(p)\min\left\{\Delta_p,\ \frac{\omega(p)}{\max_{i\in I}\|Hessf_i(p)\|}\right\},\ \alpha\in[0,\Delta_p].$$

Since $\max_{\alpha\in[0,\Delta_p]}[m_p(0) - m_p(\alpha\nu)] = m_p(0) - \min_{\alpha\in[0,\Delta_p]} m_p(\alpha\nu)$ the results follows from (3.10). ∎

3.3 Convergence analysis

In this section, we present a convergence analysis for the multiobjective trust region method under the following assumptions:

(A1). There exists $i_0 \in I$ such that f_{i_0} is bounded below.

(A2). The Hessian matrices are uniformly bounded, i.e, $\|\mathrm{Hess}f_i(p)\| \leq \beta$, for all $p \in M$ and $i = 1,\ldots,m$.

Remark 3.3.1 The assumptions **A1** and **A2** are generalizations to the mutiobjective setting of the corespondent one scalar version; see, for example, [25, page 121].

From now on, we assume that $\{p^k\}$, $\{\nu^k\}$, $\{r_k\}$ and $\{\Delta_k\}$ are the sequences generated by the multiobjective trust region method. In the next result, we study an important property of the trust radius sequence.

Lemma 3.3.2 *There holds*

$$|r_k - 1| \leq \frac{\beta\Delta_k^2}{c\omega(p^k)\min\{\Delta_k,\ \omega(p^k)/\beta\}}, \qquad k = 0,1,2,\ldots.$$

Proof. First note that direct manipulations on the function $t \mapsto f_i(\exp_{p^k}(t\nu^k))$ yield

$$f_i(\exp_{p^k}(\nu^k)) = f_i(p^k) + \left\langle \mathrm{grad}\, f_i(p^k), \nu^k\right\rangle + A_i(k), \quad k = 0,1,\ldots, \tag{3.14}$$

where the quantity $A_i(k)$ is defined by

$$A_i(k) := \int_0^{\|\nu^k\|}\left\langle P_{\gamma(t)\gamma(0)}\,\mathrm{grad}\, f_i\left(\exp_{p^k}\left(t\frac{\nu^k}{\|\nu^k\|}\right)\right) - \mathrm{grad}\, f_i(p^k), \frac{\nu^k}{\|\nu^k\|}\right\rangle dt.$$

and $\gamma(t) = \exp_{p^k}(t\nu^k/\|\nu^k\|)$. Thus, using (3.14) and considering the definition of m_k, we can conclude

$$\max_{i\in I} f_i(\exp_{p^k}(\nu^k)) - m_k(\nu^k) \leq \max_{i\in I} \left[A_i(k) - \left\langle \mathrm{Hess} f_i(p^k)\nu^k, \nu^k\right\rangle/2\right], \quad k = 0, 1, \ldots.$$

Again, from some manipulations in (3.14) and taking into account the definition of m_k, we have

$$\max_{i\in I} f_i(\exp_{p^k}(\nu^k)) - m_k(\nu^k) \geq \max_{i\in I}\left[-A_i(k)\right] - \max_{i\in I}\left\langle \mathrm{Hess} f_i(p^k)\nu^k, \nu^k\right\rangle/2, \quad k = 0, 1, \ldots.$$

Combining two latter inequalities, we obtain

$$\left|\max_{i\in I} f_i(\exp_{p^k}(\nu^k)) - m_k(\nu^k)\right| \leq$$
$$\max\left\{\left|\max_{i\in I}\left\{A_i(k) - \left\langle \mathrm{Hess} f_i(p^k)\nu^k, \nu^k\right\rangle/2\right\}\right|, \ \left|\max_{i\in I}\left[-A_i(k)\right] - \max_{i\in I}\left\langle \mathrm{Hess} f_i(p^k)\nu^k, \nu^k\right\rangle/2\right|\right\},$$

for $k = 0, 1, \ldots$. Using properties of the maximum and triangular inequality, last inequality reduces to

$$\left|\max_{i\in I} f_i(\exp_{p^k}(\nu^k)) - m_k(\nu^k)\right| \leq \max_{i\in I}|A_i(k)| + \frac{1}{2}\max_{i\in I}\left|\left\langle \mathrm{Hess} f_i(p^k)\nu^k, \nu^k\right\rangle\right|, \quad k = 0, 1, \ldots.$$

Taking into account **A2**, Proposition 2.1.2 and that $\|\nu^k\| \leq \Delta_k$, for all k, the last inequality implies

$$\left|\max_{i\in I} f_i(\exp_{p^k}(\nu^k)) - m_k(\nu^k)\right| \leq \beta\Delta_k^2, \quad k = 0, 1, \ldots. \tag{3.15}$$

On the other hand, the combination of the definition of m_k with (3.5) yield, for k=0, 1,. . . ,

$$|r_k - 1| = \left|\frac{\max_{i\in I} f_i(p^k) - \max_{i\in I} f_i(\exp_{p^k}(\nu^k))}{m_k(0) - m_k(\nu^k)} - 1\right| = \left|\frac{m_k(\nu^k) - \max_{i\in I} f_i(\exp_{p^k}(\nu^k))}{m_k(0) - m_k(\nu^k)}\right|.$$

Since we are under the condition (3.7), the desired result follows from **A2**, (3.15) and the last equality.

■

We ready to prove the main result of the chapter, whose the proof uses similar techniques from linear scalar setting adapted to the Riemannian multiobjective context; see, for example, [53, Theorem 4.5 and 4.6] .

Theorem 3.3.3 *There holds* $\lim_{k\to\infty} \omega(p^k) = 0$. *Moreover, each cluster point of* $\{p^k\}$, *if any, is Pareto critical.*

Proof. For proving the result, let $r' > 0$ the constant defined on the Algorithm, let us to define the set $\Gamma := \{k \ : \ r_k > r'\}$. First, we assume that Γ is a finite set and let $\bar{k} = \max \Gamma$. It is immediate from (3.8) that $p^k = p^{\bar{k}+1}$, for $k = \bar{k}+1, \bar{k}+2, \ldots$. Let us to suppose, by contradiction, that $p^{\bar{k}+1}$ is not a Pareto critical point. Thus, the item (ii) of Lemma 3.1.1 implies

$$\omega(p^k) = \omega(p^{\bar{k}+1}) > 0, \quad k = \bar{k}+1, \bar{k}+2, \ldots.$$

Using Lemma 3.3.2 we have

$$|r_k - 1| \leq \frac{\beta \Delta_k^2}{c\omega(p^{\bar{k}+1}) \min\left\{\Delta_k, \dfrac{\omega(p^{\bar{k}+1})}{\beta}\right\}}, \quad k = \bar{k}+1, \bar{k}+2, \ldots. \tag{3.16}$$

Since $r_k \leq r'$, for $k = \bar{k}+1, \bar{k}+2, \ldots$, (3.9) implies that the trust radius sequence $\{\Delta_k\}$ converges to 0. Thus, using (3.16) we conclude that there exists $\hat{k} > \bar{k}$ such that $|r_{\hat{k}} - 1| \leq 1/2$. Last inequality implies that $r_{\hat{k}} \geq 1/2 > r'$. But it is a contradiction because $\hat{k} > \bar{k} = \max \Gamma$. Therefore, $p^{\bar{k}+1}$ is Pareto critical.

Let us to assume that Γ is an infinite set. Suppose that there exist $\epsilon > 0$ and $\tilde{k}$ satisfying

$$\omega(p^k) \geq \epsilon, \quad k \geq \tilde{k}. \tag{3.17}$$

First of all we are going to prove that the following inequality holds

$$\Delta_k \geq \min\left\{\Delta_{\tilde{k}}, \frac{\hat{\Delta}}{4}\right\} > 0, \quad \hat{\Delta} := \min\left\{\frac{c\epsilon}{2\beta}, \frac{\epsilon}{\beta}\right\}, \qquad k \geq \tilde{k}. \tag{3.18}$$

We proceed by induction on k. The inequality holds for $k = \tilde{k}$. Assume that (3.18) holds for $k > \tilde{k}$. We have two possibilities, either $\Delta_k \leq \hat{\Delta}$ or $\Delta_k > \hat{\Delta}$. We begin analyzing the possibility $\Delta_k \leq \hat{\Delta}$. Since we are under the assumption $\Delta_k \leq \hat{\Delta}$, using Lemma 3.3.2, (3.17) and definition of $\hat{\Delta}$ we have

$$|r_k - 1| \leq \frac{\beta \Delta_k^2}{c\epsilon \min\left\{\Delta_k, \dfrac{\epsilon}{\beta}\right\}} \leq \frac{\Delta_k}{2 \min\left\{\Delta_k, \frac{\epsilon}{\beta}\right\}} = \frac{1}{2}.$$

Therefore, $r_k \geq 1/2 > 1/4$ and it follows from the Algorithm that $\Delta_{k+1} \geq \Delta_k$ and (3.18) holds for $k+1$. Now, assume that $\Delta_k > \hat{\Delta}$. From (3.9) we have, either $\Delta_{k+1} \geq \Delta_k$ or $\Delta_{k+1} = \Delta_k/4$. If $\Delta_{k+1} \geq \Delta_k$ then using the induction assumption we conclude that (3.18) holds for $k+1$. If $\Delta_{k+1} = \Delta_k/4$ then $\Delta_{k+1} > \hat{\Delta}/4$ because $\Delta_k > \hat{\Delta}$ and (3.18) holds for $k+1$. Therefore, (3.18) holds for all $k \geq \tilde{k}$. Consider the index set $J = \{k_j \in \Gamma \ : \ k_j \geq \tilde{k}, \ j \in \mathbb{N}\}$ and note that from (3.8) we have

$$p^{k_{j+1}} = p^{k_j+1} = \exp_{p^{k_j}} \nu^{k_j}, \quad k_j \in J.$$

Hence, from last equality, (3.5), (3.7), assumption **A2** and (3.17) we obtain

$$\max_{i\in I} f_i(p^{k_j}) - \max_{i\in I} f_i(p^{k_{j+1}}) \geq r' \left(m_{k_j}(0) - m_{k_j}(\nu^{k_j})\right) \geq r'c\epsilon \min\left\{\Delta_{k_j}, \frac{\epsilon}{\beta}, \right\} \geq 0, \quad k_j \in J.$$

The Remark 3.2.1 implies that the sequence $\{\max_{i\in I} f_i(p^{k_j})\}$ is nonincreasing and, using **A1**, it is also bounded below. Thus, $\{\max_{i\in I} f_i(p^{k_j})\}$ converges and, using last inequality, we conclude that

$$\lim_{k_j\to\infty} \Delta_{k_j} = 0,$$

contradicting the inequality (3.18). Therefore the assumption (3.17) is false. Consequently, in view of the assumption that Γ is an infinite set, using Lemma 3.1.1 we have

$$\liminf_{k\to\infty} \omega(p^k) = 0. \tag{3.19}$$

For proving that in fact $\lim_{k\to\infty} \omega(p^k) = 0$, fix an arbitrary $s \geq 0$. For simplifying the notations, we define

$$\epsilon := \frac{1}{2}\omega(p^s), \qquad R := \frac{\epsilon}{\beta}, \qquad B(p^s, R) := \{p \in M \ : \ d(p^s, p) < R\}. \tag{3.20}$$

For any $p \in B(p^s, R)$ we can prove that $\omega(p) = \omega(p^s) + \omega(p) - \omega(p^s) \geq \omega(p^s) - |\omega(p) - \omega(p^s)|$. Thus, combining **A2**, Proposition 2.1.2 and Lemma 3.1.1 item (iii) with the definitions of ϵ and R in (3.20), the last inequality becomes $\omega(p) \geq \omega(p^s) - \beta d(p, p^s) > \omega(p^s) - \beta R = \epsilon$, for all $p \in B(p^s, R)$. This inequality together with (3.19) implies that there exists $n > s$ such that $p^n \notin B(p^s, R)$. Take $\ell \geq s$ such that $p^{\ell+1}$ is the first iterate satisfying $p^{\ell+1} \notin B(p^s, R)$. We claim that

$$\max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) \geq r'c\epsilon\frac{\epsilon}{\beta}. \tag{3.21}$$

Using (3.5) and (3.8), after some simple algebraic manipulations we have

$$\max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) = \sum_{k=s}^{\ell} \max_{i\in I} f_i(p^k) - \max_{i\in I} f_i(p^{k+1}) \geq \sum_{k=s,\ k\in J}^{\ell} r' \left[m_k(0) - m_k(\nu^k)\right],$$

where $J := \{k \ : \ p^k \neq p^{k+1}\}$. Since $\omega(p^k) > \epsilon$, for $k = s, s+1, \ldots, \ell$, using (3.7) and taking into account assumption **A2**, the latter inequality becomes

$$\max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) \geq \sum_{k=s,k\in J}^{\ell} r'c\epsilon \min\left\{\Delta_k, \frac{\epsilon}{\beta}\right\}. \tag{3.22}$$

Note that, if there exists $k' \in \{s, s+1, \ldots \ell\} \cap J$ such that $\Delta_{k'} > \epsilon/\beta$, then (3.21) is an immediate consequence of (3.22) . Now, assume that $\Delta_k \leq \epsilon/\beta$, for all $k \in \{s, s+1, \ldots \ell\} \cap J$. Hence, (3.22) yields

$$\max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) \geq r'c\epsilon \sum_{k=s,k\in J}^{\ell} \Delta_k.$$

Considering that (3.6) and (3.8) imply, respectively, that $\|\nu^k\| \leq \Delta_k$ and $\|\nu^k\| = d(p^k, p^{k+1})$, for all $k \in J$, the above inequality becomes

$$\max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) \geq r'c\epsilon \sum_{k=s,k\in J}^{\ell} d(p^k, p^{k+1}).$$

Using triangular inequality, last inequality implies $\max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) \geq r'c\epsilon d(p^s, p^{\ell+1})$ and due to $p^{\ell+1} \notin B(p^s, R)$, we conclude from (3.20) that (3.21) holds, which prove the claim. Since Remark 3.2.1 implies that $\{\max_{i\in I} f_i(p^k)\}$ is non-increasing and bounded below, thus we have $\lim_{k\to\infty} \max_{i\in I} f_i(p^k) = \tau > -\infty$. Therefore, using (3.21) we obtain that

$$\max_{i\in I} f_i(p^s) - \tau \geq \max_{i\in I} f_i(p^s) - \max_{i\in I} f_i(p^{\ell+1}) \geq r'c\epsilon\frac{\epsilon}{\beta}.$$

Combining the equalities in (3.20) with the latter inequality, it follows that

$$\max_{i\in I} f_i(p^s) - \tau \geq r'c\frac{\omega(p^s)}{2}\frac{\omega(p^s)}{2\beta} \geq 0.$$

Then, letting s goes to infinity and considering that $\lim_{s\to\infty} \max_{i\in I} f_i(p^s) = \tau$, the first statement of the theorem follows. The second one follows from item (ii) of Lemma 3.1.1. ■

Remark 3.3.4 We show that, unlikely the line search method used in [33], the trust region method ensure the global convergence of the Newton method without assuming any hypothesis of convexity of the objective function.

Chapter 4

The proximal point method for vector optimization on Hadamard Manifolds

In this chapter, we present the vector optimization problem, some concepts and results related to this problem, and introduce the proximal point method for this problem and its inexact version.

Let $C \subset \mathbb{R}^m$ be a closed pointed convex cone, with nonempty interior. We will use the binary relations $\preceq_C$ and $\prec_C$ defined, respectively, by $p \preceq_C q$ means $q - p \in C$ and $p \prec_C q$ means $q - p \in \text{int}C$, for all $p, q \in \mathbb{R}^m$. Given a vector-valued function $F : M \to \mathbb{R}^m$, we consider the problem of finding an *efficient point* of F, i.e., a point $p^* \in \mathbb{R}^n$ such that there exists no $p \in M$ with $F(p) \preceq_C F(p^*)$ and $F(p) \neq F(p^*)$. We denote this unconstrained problem as

$$C - \text{Min}_{p \in M} F(p). \tag{4.1}$$

We say that $p^* \in M$ is a *weakly efficient point* of (4.1) if there is no $p \in M$ such that $F(p) \prec_C F(p^*)$. The set of the weakly efficient points of (4.1) is denoted by $C - \text{argmin}_w\{F(p) : p \in M\}$. Let define by C^* the positve dual set of the cone C, i.e.

$$C^* = \{w \in \mathbb{R}^n \ : \ \langle w, y \rangle \geq 0, \text{ for all}\, y \in C\}.$$

Since C is a closed convex cone, then $C = C^{**}$, as a consequence

$$C = \{y \in \mathbb{R}^n \ : \ \langle y, w \rangle \geq 0, \text{ for all}\, w \in C^*\},$$

and

$$\text{int}C = \{y \in \mathbb{R}^n \ : \ \langle y, w \rangle > 0, \text{ for all}\, w \in C^*\}.$$

From now on, we take a compact set $Z \subset \mathbb{R}^m \backslash \{0\}$ such that the cone generated by the convex hull of Z will be the dual set C^*, i.e., $\text{cone}(\text{conv}Z) = C^*$. As $\text{int}(C) \neq \emptyset$ and $Z \subset C^* \backslash \{0\}$, it follows that $0 \in conv(Z)$, therefore

$$C = \{y \in \mathbb{R}^m \ : \ \langle y, z \rangle \geq 0, \text{ for all}\, z \in Z\}, \tag{4.2}$$

and

$$\mathrm{int}C = \{y \in \mathbb{R}^n \ : \ \langle y, z\rangle > 0, \text{ for all}\, z \in Z\}\,.$$

Remark 4.0.1 In the particular case of classical optimization, namely, $C = \mathbb{R}_+$, we can take $Z = \{1\}$. For multiobjective optimization, i.e., when C is the nonnegative orthant of $\mathbb{R}^m$, we can take Z as the canonical base of $\mathbb{R}^m$. For a generic cone C we can take $Z = \{z \in C^* \ : \ \|z\|_1 = 1\}$, where $C^* := \{y \in \mathbb{R}^m \ : \ \langle y, x\rangle \geq 0,\ x \in C\}$ and $\|z\|_1 := |z_1| + |z_2| + \cdots + |z_m|$.

We consider the following nonlinear scalar function $f_e : \mathbb{R}^m \to \mathbb{R}$, which will play an important role in our analysis, defined by

$$f_e(y) := \inf\{t \in \mathbb{R} \ : \ \ te \in y + C\}, \tag{4.3}$$

where e is any fixed point in int$C$; see, for example, [37, page 39]. In [23, Proposition 1.44] it was proved that the nonlinear function above can be rewritten as follow.

Proposition 4.0.2 *Assuming that int$C \neq \emptyset$. Let $e \in$ int$C$, then for all $y \in \mathbb{R}^m$ the function $f_e(y) := \inf\{t \in \mathbb{R} \ : \ \ te \in y + C\}$ is equivalent to*

$$f_e(y) = \max_{z \in Z} \frac{\langle y, z\rangle}{\langle e, z\rangle}. \tag{4.4}$$

Proof. From definition of (4.3), for any $e \in \mathrm{int}C$ and $y \in \mathbb{R}^m$ we have $f_e(y)e - y \in C$, then, for any $z \in Z$ we obtain

$$\langle t_0 e - y, z\rangle \geq 0.$$

As $\langle e, z\rangle > 0$, it follows for linearly that

$$f_e(y) \geq \frac{\langle y, z\rangle}{\langle e, z\rangle}.$$

Since z has been taken any in Z, we conclude

$$f_e(y) \geq \max_{z \in Z} \frac{\langle y, z\rangle}{\langle e, z\rangle}. \tag{4.5}$$

Now, define

$$t_0 = \max_{z \in Z} \frac{\langle y, z\rangle}{\langle e, z\rangle}.$$

Hence, we obtain

$$\langle t_0 e - y, z\rangle \geq 0, \text{ for all}\, z \in Z.$$

From (4.2), the inequality above implies

$$t_0 e - y \in C.$$

It follows from (4.3) that $f_e(y) \leq t_0$, which together with (4.5) concludes our proof.

■

Remark 4.0.3 Note that, for C being the nonnegative orthant of $\mathbb{R}^m$, (4.4) becomes $f(y) = \max_{i\in I}\langle y, e_i\rangle$, where $\{e_i\} \subset \mathbb{R}^m$ is the canonical base of the space $\mathbb{R}^m$, which has been used in [13].

Next lemma, given in [37, Corollary 2.3.5], gives us some properties of (4.3) that will be useful through this chapter.

Lemma 4.0.4 *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a nonlinear function defined in* (4.3)*, then following properties hold:*

i) Given $y \in \mathbb{R}^m$, $\alpha \in \mathbb{R}$ and $t > 0$, $f_e(ty + \alpha e) = tf_e(y) + \alpha$;

ii) If $y \preceq_C w$ then $f(y) \leq f(w)$ for any y, $w \in \mathbb{R}^m$.

Proof. Since the equation (4.3) is equivalent to (4.4), to prove item *i)*, we have directly,

$$f_e(ty + \alpha e) = \max_{z\in Z} \frac{\langle ty + \alpha e, z\rangle}{\langle e, z\rangle} = tf_e(y) + \alpha.$$

Now, using the definition of partial order induced by the cone C, the inequality $y \preceq_C w$ implies, for any $z \in Z$,

$$\langle y, z\rangle \leq \langle w, z\rangle.$$

Then multiplying the inequality above by $1/\langle e, z\rangle$ and taking the maximum in $z \in Z$ we obtain the proof of *ii)*. ■

Now we introduce the proximal point method for vector optimization. Let $\{\lambda^k\}$ be a sequence of positive numbers and $\{e^k\} \subset \operatorname{int} C$ such that $\|e^k\| = 1$, for $k = 0, 1, \ldots$. Consider the sequence of functions

$$f_k(y) := \max_{z\in Z} \frac{\langle y, z\rangle}{\langle e^k, z\rangle}, \qquad k = 0, 1, \ldots. \tag{4.6}$$

The *proximal point method* for solving (4.1), with starting point $p^0 \in M$, is defined by

$$p^{k+1} := \operatorname{argmin}_{p\in\Omega_k} f_k \left(F(p) + \frac{\lambda_k}{2} d^2(p, p^k) e^k \right), \tag{4.7}$$

for $k = 0, 1, \ldots$, where $\Omega_k := \{p \in M \ : \ F(p) \preceq_C F(p^k)\}$. From now on, $\{p^k\}$ denotes the sequence generated by the proximal point method with starting point $p^0 \in M$.

4.1 Convergence analysis

In this section, we prove the convergence of any generated sequence by the proximal point method to a weak efficient point. For this purpose, we need to define the convexity of a function with respect to the order induced by C. A vector-valued function $F : M \to \mathbb{R}^m$ is called C- *convex* if, for $p, q \in M$ and $\gamma : [0,1] \to M$ a geodesic segment joining p to q, there holds $F(\gamma(t)) \preceq_C (1-t)F(p) + tF(q)$, for all $t \in [0,1]$; see [19]. In the next proposition, we show that the scalar function (4.4) composed with a C-*convex* function is a convex function.

Proposition 4.1.1 *Let $F : M \to \mathbb{R}^n$ be a C-*convex *function , then the compound function $f_e \circ F : \mathbb{R} \to \mathbb{R}^n$ is a convex function.*

Proof. From definition of C-convexity, for any geodesic $\gamma : [0,1] \to M$ joining p to q, for all $t \in [0,1]$ and for any $z \in Z$ we have

$$\langle F(\gamma(t)), z\rangle \leq \langle (1-t)F(p) + tF(q), z\rangle.$$

Then multiplying the inequality above by $1/\langle e, z\rangle$, taking the maximum in $z \in Z$ and using the properties of the maximum we obtain

$$\max_{z\in Z} \frac{\langle F(\gamma(t)), z\rangle}{\langle e, z\rangle} \leq (1-t) \max_{z\in Z} \frac{\langle F(p), z\rangle}{\langle e, z\rangle} + t \max_{z\in Z} \frac{\langle F(q), z\rangle}{\langle e, z\rangle},$$

which implies

$$(f_e \circ F)(\gamma(t)) \leq (1-t)(f_e \circ F)(p) + t(f_e \circ F)(q).$$

■

We also need of the following assumption:

(A1) $\bar{\Omega} := \bigcap_{k=0}^{\infty} \Omega_k \neq \varnothing$.

Remark 4.1.2 In general the set $\bar{\Omega}$ in **(A1)** can be an empty set. One way to guarantee that $\bar{\Omega}$ is nonempty is to assume: for each $p^0 \in M$ the set $(F(p^0) - C) \cap F(M)$ is C-*complete* (see [48, Section 19]), meaning that each sequence $\{q^k\} \subset M$, with $q^0 = p^0$, such that $F(q^{k+1}) \preceq_C F(q^k)$, for $k = 0, 1, \ldots$, there exists $q \in M$ such that $F(q) \preceq_C F(q^k)$, for $k = 0, 1, \ldots$. This assumption is standard to ensure the convergence of descent methods in vector optimization; see, for instance, [18, 22, 35, 39, 69].

Now we are ready to state and prove the main result of this section.

Theorem 4.1.3 *Let $F : M \to R^m$ be a C-convex function, and assume that* **(A1)** *hold and let $\{\lambda^k\}$ be a sequence such that $\sum_{k=0}^{\infty}(1/\lambda_k) = +\infty$. Then, $\{p^k\}$ is well defined and converges to a weakly efficient point.*

Proof. Let $\{f_k\}$ be the sequence of functions defined in (4.6), and define $\varphi_k : M \to \mathbb{R} \cup \{+\infty\}$ by

$$\varphi_k(p) = f_k\left(F(p) + I_{\Omega_k}(p)e^k + \frac{\lambda_k}{2} d^2(p, p^k)e^k\right),$$

for all $k = 0, 1, \ldots$, where I_{Ω_k} is the indicator function of Ω_k. From item $i)$ of Lemma 4.0.4 we have

$$\varphi_k(p) = f_k\left(F(p)\right) + I_{\Omega_k}(p) + \frac{\lambda_k}{2} d^2(p, p^k), \tag{4.8}$$

for all $k = 0, 1, \ldots$. Since F is C- convex, then $f_k \circ F$ is convex and I_{Ω_k} are convex, for $k = 0, 1, \ldots$. Consequently, taking into account that $d^2(\cdot, p^k)$ is strongly convex and Ω_k is closed, we conclude that φ_k is strongly convex and lower semicontinuous on Ω_k, for $k = 0, 1, \ldots$. Thus, there exists a unique $p^{k+1} \in \Omega_k$ such that

$$p^{k+1} = \operatorname{argmin}_{p \in M} \varphi_k(p), \qquad k = 0, 1, \ldots, \tag{4.9}$$

which implies that $\{p^k\}$ is well defined and the first part of the proposition is proved.

Using the convexity of the function φ_k and (4.9) we conclude that $0 \in \partial \varphi_k(p^{k+1})$, which from (4.8) yields

$$0 \in \partial \left(f_k \circ F(\cdot) + I_{\Omega_k}(\cdot) + \frac{\lambda_k}{2} d^2(\cdot, p^k) \right) (p^{k+1}),$$

for all $k = 0, 1, \ldots$. Note that $\operatorname{int}(\operatorname{dom} f_k \circ F) = M$, $\operatorname{dom}(\lambda_k)/2)d^2(\cdot, p^k) = M$ and $\operatorname{dom} I_{\Omega_K} = \Omega_k$. Owing to F be C- convex and **(A1)** holds, then it follows that Ω_k is convex and nonempty. Thus, we can apply Proposition 2.2.1 with $g = f_k \circ F + (\lambda_k/2)d^2(\cdot, p^k)$ and $\Omega = \Omega_k$ to obtain that

$$0 \in \partial \left(f_k \circ F(\cdot) + \frac{\lambda_k}{2} d^2(\cdot, p^k) \right) (p^{k+1}) + N_{\Omega_k}(p^{k+1}),$$

for all $k = 0, 1, \ldots$. Again applying Proposition 2.2.1 with $g_1 = f_k \circ F$ and $g_2 = (\lambda_k/2)d^2(\cdot, p^k)$ in last inclusion and considering that $\operatorname{grad} d^2_{p^k}(p^{k+1}) = -2exp^{-1}_{p^{k+1}} p^k$, we conclude that there exist $w^{k+1} \in \partial(f_k \circ F)(p^{k+1})$ and $v^{k+1} \in N_{\Omega_k}(p^{k+1})$ such that

$$\exp^{-1}_{p^{k+1}} p^k = \frac{1}{\lambda_k} \left(w^{k+1} + v^{k+1} \right), \qquad k = 0, 1, \ldots. \tag{4.10}$$

On the other hand, using inequality (2.2) with $p_1 = p \in M$, $p_2 = p^k$ and $p_3 = p^{k+1}$, we have

$$d^2(p, p^{k+1}) + d^2(p^{k+1}, p^k) - \langle \exp^{-1}_{p^{k+1}} p, \exp^{-1}_{p^{k+1}} p^k \rangle \leq d^2(p, p^k),$$

for each $k = 0, 1, \ldots$. Substituting the equality in (4.10) into the last inequality, we obtain

$$d^2(p, p^{k+1}) \leq d^2(p, p^k) - d^2(p^{k+1}, p^k) + \frac{1}{\lambda_k} \langle \exp^{-1}_{p^{k+1}} p, w^{k+1} + v^{k+1} \rangle, \tag{4.11}$$

for all $k = 0, 1, \ldots$. Considering that $f_k \circ F$ is a convex function and $w^{k+1} \in \partial(f_k \circ F)(p^{k+1})$ then assumption **(A1)** implies

$$\left\langle \exp^{-1}_{p^{k+1}} p,\, w^{k+1} \right\rangle \leq f_k(F(p)) - f_k(F(p^{k+1})) \leq 0, \tag{4.12}$$

for all $p \in \bar{\Omega}$ and $k = 0, 1, \ldots$, where the last inequality follows from item *ii*) of Lemma 4.0.4. Now, taking into account that $\bar{\Omega} \subset \Omega_k$ and $v^{k+1} \in N_{\Omega_k}(p^{k+1})$, for $k = 0, 1, \ldots$, we have

$$\langle \exp^{-1}_{p^{k+1}} p,\, v^{k+1} \rangle \leq 0, \;\; p \in \bar{\Omega}, \qquad k = 0, 1, \ldots. \tag{4.13}$$

Therefore, taking into account **(A1)**, we can combine (4.11), (4.12) and (4.13) to conclude that $\{p^k\}$ is Fejér convergence to $\bar{\Omega}$. In particular, Proposition 2.1.5 implies that $\{p^k\}$ is bounded. Let $\bar{p}$ be a cluster point of $\{p^k\}$ and $\{p^{k_j}\}$ a subsequence of $\{p^k\}$ such that $\lim_{k\to\infty} p^{k_j} = \bar{p}$. Note that (4.7) yields

$$F(p^{k+1}) \preceq_C F(p^k), \qquad k = 0, 1, \ldots.$$

Hence, the continuity of F implies that $F(\bar{p}) \preceq_C F(p^k)$, for $k = 0, 1, \ldots$, which is equivalent to say $\bar{p} \in \bar{\Omega}$. Using Proposition 2.1.5 we conclude that $\{p^k\}$ converges to $\bar{p}$.

It remains to prove that $\bar{p}$ is a weakly efficient point. Let us suppose, by contradiction, that there exists $\hat{p} \in M$ such that $F(\hat{p}) \prec_C F(\bar{p})$ (note that, in particular, $\hat{p} \in \bar{\Omega}$). Since $\bar{p} \in \bar{\Omega}$, using item *ii*) of Lemma 4.0.4 we have

$$f_k(F(\hat{p})) - f_k(F(\bar{p})) \geq f_k(F(\hat{p})) - f_k(F(p^{k+1})).$$

Owing that (4.10) implies $\lambda_k \exp^{-1}_{p^{k+1}} p^k - v^{k+1} \in \partial(f_k \circ F)(p^{k+1})$, the last inequality and the definition of subdifferential yield

$$f_k(F(\hat{p})) - f_k(F(\bar{p})) \geq \langle \exp^{-1}_{p^{k+1}} \hat{p}, \lambda_k \exp^{-1}_{p^{k+1}} p^k - v^{k+1} \rangle.$$

Since $F(\hat{p}) \prec_C F(\bar{p}) \preceq_C F(p^k)$ then $\hat{p} \in \Omega_k$ for any k and, consequently $\langle \exp^{-1}_{p^{k+1}} \hat{p}, v^{k+1} \rangle \leq 0$, which implies

$$\frac{1}{\lambda_k}\left(f_k(F(\hat{p})) - f_k(F(\bar{p}))\right) \geq \langle \exp^{-1}_{p^{k+1}} \hat{p}, \exp^{-1}_{p^{k+1}} p^k \rangle. \tag{4.14}$$

On the other hand, applying inequality (2.2) with $p_1 = \hat{p}$, $p_2 = p^k$ and $p_3 = p^{k+1}$ we conclude that

$$d^2(\hat{p}, p^{k+1}) - d^2(\hat{p}, p^k) \leq 2\langle \exp^{-1}_{p^{k+1}} \hat{p}, \exp^{-1}_{p^{k+1}} p^k \rangle - d^2(p^{k+1}, p^k).$$

Using (4.6) we have $f_k(F(\bar{p}) - F(\hat{p})) \geq f_k(F(\bar{p})) - f_k(F(\hat{p}))$, for $k = 0, 1, \ldots$. Thus, the latter inequality and (4.14) give us

$$d^2(\hat{p}, p^{k+1}) - d^2(\hat{p}, p^k) \leq \frac{2}{\lambda_k} f_k(F(\bar{p}) - F(\hat{p})).$$

Since $F(\hat{p}) \prec_C F(\bar{p})$, we conclude that there exist $\delta > 0$ such that $\langle F(\hat{p}), z \rangle < \langle F(\bar{p}), z \rangle - \delta$, for all $z \in Z$. Hence, using (4.6) and the last inequality we obtain

$$d^2(\hat{p}, p^{k+1}) - d^2(\hat{p}, p^k) \leq \frac{2}{\lambda_k} \max_{z \in Z} \frac{\langle F(\bar{p}) - F(\hat{p}), z \rangle}{\langle e_k, z \rangle} \leq \frac{2}{\lambda_k} \max_{z \in Z} \frac{-\delta}{\langle e_k, z \rangle}.$$

After some algebraic manipulations, and taking account $\{e^k\} \subset \operatorname{int} C$ and $\|e^k\| = 1$, the last inequality becomes

$$\frac{1}{\lambda_k} \leq \frac{\delta}{2}\left[d^2(\hat{p}, p^k) - d^2(\hat{p}, p^{k+1})\right]\|z\|.$$

Summing the inequality from $k = 0$ to j we have

$$\sum_{k=0}^{j} \frac{1}{\lambda_k} \leq \frac{\delta}{2}\left(d^2(\hat{p}, p^o) - d^2(\hat{p}, p^{j+1})\right)\|z\| \leq \frac{\delta}{2} d^2(\hat{p}, p^o)\|z\|,$$

which contradicts the assumptions $\sum_{k=0}^{\infty}(1/\lambda_k) = +\infty$ and the desired result follows. ■

4.2 Inexact version of the proximal method

In this section we present the inexact version of the proximal point method for vector optimization and prove that any generated sequence convergence to the weakly efficient point. Inexact versions of the proximal point method for optimization problem have been studied in [7, 67, 71]. It is worth mentioning that the error criterium of our method merge into the criterium used in [18], that is different from the used in [71], which solves the proximal subproblem exactly and chooses as iterate point close to that solution, in many cases, solving the proximal subproblem exactly is either impossible or as difficult as solving the minimization problem. Before stating the method we choose $0 < \sigma < 1$, $\{\lambda_k\} \subset \mathbb{R}_{++}$ and $\{e^k\} \subset \operatorname{int} C$ such that $\|e^k\| = 1$, for all $k = 0, 1, \ldots$. The *inexact proximal point method* for solving (4.1) starts choosing $p^0 \in M$. In the iterative step it takes as p^{k+1} any $p \in M$ such that there exists $\epsilon_k \geq 0$ satisfying,

$$0 \in \partial_{\epsilon_k}(f_k \circ F)(p) + N_{\Omega_k}(p) + \lambda_k \exp_p^{-1} p^k, \tag{4.15}$$

$$\epsilon_k \leq \sigma \lambda_k d^2(p, p^k)/2, \tag{4.16}$$

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

Theorem 4.2.1 *Let $F : M \to R^m$ be a C-convex function, and assume that* **(A1)** *holds and let $\{\lambda^k\}$ be a sequence such that $\sum_{k=0}^{\infty}(1/\lambda_k) = +\infty$. Then $\{p^k\}$ generated by* (4.15) *and* (4.16) *is well defined and converges to a weakly efficient point.*

Proof. In Theorem 4.1.3, we proved that there exists p^{k+1} such that $0 \in \partial\varphi_k(p^{k+1})$, where

$$\varphi_k(p) = f_k\left(F(p) + I_{\Omega_k}(p)e^k + \lambda_k/2\, d^2(p, p^k)e^k\right).$$

From Proposition 2.2.1 we have

$$0 \in \partial(f_k \circ F)(p) + N_{\Omega_k}(p) - \lambda_k \exp_p^{-1} p^k, \qquad k = 0, 1, \ldots.$$

Since $\partial(f_k \circ F)(p^{k+1}) \subset \partial_\epsilon(f_k \circ F)(p^{k+1})$. Thus, we conclude that there exists p^{k+1} satisfying (4.15) and (4.16). From (4.15) we obtain

$$0 \in \partial_{\epsilon_k}(f_k \circ F)(p^{k+1}) + N_{\Omega_k}(p^{k+1}) - \lambda_k \exp^{-1}_{p^{k+1}} p^k,$$

for all $k = 0, 1, \ldots$. Taking $\nu^k \in N_{\Omega_k}(p^{k+1})$, the above inclusion becomes

$$\lambda_k \exp^{-1}_{p^{k+1}} p^k - \nu^k \ \in \partial_{\epsilon_k}(f_k \circ F)(p^{k+1}), \qquad k = 0, 1, \ldots.$$

Take $p \in \bar{\Omega}$. In view of the definition of $\partial_{\epsilon_k}(f_k \circ F)$, **(A1)** and Lemma 4.0.4 we have

$$-\epsilon_k + \lambda_k \langle \exp^{-1}_{p^{k+1}} p^k, \exp^{-1}_{p^{k+1}} p\rangle - \langle \nu^k, \exp^{-1}_{p^{k+1}} p\rangle \leq f_k \circ F(p) - f_k \circ F(p^{k+1}) \leq 0,$$

for all $k = 0, 1, \ldots$. Since $\nu^k \in N_{\Omega_k}(p^{k+1})$, we obtain from the last inequality that

$$0 \leq \epsilon_k - \lambda_k \langle \exp^{-1}_{p^{k+1}} p^k, \exp^{-1}_{p^{k+1}} p\rangle, \qquad k = 0, 1, \ldots.$$

Applying (2.2) with $p_1 = p$, $p_2 = p^k$ and $p_3 = p^{k+1}$, the last inequality implies

$$0 \leq \epsilon_k + \frac{\lambda_k}{2}\left(d^2(p, p^k) - d^2(p, p^{k+1}) - d^2(p^{k+1}, p^k)\right),$$

for all $k = 0, 1, \ldots$. Combining (4.16) with the last inequality we obtain

$$0 \leq \frac{\lambda_k}{2}\left(d^2(p, p^k) - d^2(p, p^{k+1}) - (1-\sigma) d^2(p^{k+1}, p^k)\right),$$

for all $k = 0, 1, \ldots$. Thus, in view of $0 < \sigma < 1$ and $\lambda_k > 0$, it follows that $d^2(p, p^{k+1}) \leq d^2(p, p^k)$, for all $k = 0, 1, \ldots$. Hence $\{p^k\}$ is Fejér convergence to $\bar{\Omega}$. Therefore, by using similar argument as in Theorem 4.1.3, we can prove that $\{p^k\}$ converges to some $\bar{p} \in \bar{\Omega}$, and $\bar{p}$ is a weakly efficient point of F.

■

Chapter 5

The Newton method for variable order structure

In this chapter we present the Newton method for solving vector optimization problems with respect to the variable order structure. But first we introduce some notations, definitions and results which will be used throughout this chapter. Moreover, we present the vector optimization problem with respect to the variable ordering structure, the first-order optimality condition for it and we extend the concept of convexity and strong convexity of a function to the variable ordered case and obtain some properties of this classes of functions. Then we introduce the concept of Newton direction and its properties. We propose the algorithm and we prove the convergence of the sequence generated for this method to efficient point. And finally we make a study of the convergence rate of this algorithm.

5.1 Notation and auxiliary results

We denote $\langle\cdot,\cdot\rangle$ as the usual inner product in $\mathbb{R}^n$ and $\|\cdot\|$ the Euclidean norm. We denote A^c the complement of A and its convex hull as $\mathrm{conv}(A)$. $B(x,r) := \{y : \|x-y\| \leq r\}$ represents the ball centered at x with radius r and $S(x,r) := \{y : \|x-y\| = r\}$, its boundary. Given two sets A and B, $\mathrm{dist}(A,B)$ is the Hausdorff's distance, *i.e.*, $\mathrm{dist}(A,B) := \max\Big\{\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(a,b)\Big\}$.

5.2 Vector optimization problem

The variable order structure in $\mathbb{R}^m$ is given by the set-valued application $K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, where $K(x)$ is a pointed, convex and closed cone with nonempty interior for each $x \in \mathbb{R}^n$. Given $x \in \mathbb{R}^n$, we define the dual cone of $K(x)$ by $K^*(x) := \{w \in \mathbb{R}^m : \langle w, y\rangle \geq 0$, for all $y \in K(x)\}$. Then, we denote as $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ the set value map, which for each x, defines a compact set $G(x) \subset K^*(x) \cap S(0;1)$, such that the cone generated by the convex hull of the set $G(x)$ coincide with $K^*(x)$, i.e., $0 \notin G(x)$ and $\text{cone}(\text{conv}G(x^*)) = K^*(x)$.

Given an open and convex set $U \subset \mathbb{R}^n$ and a vector function $F : U \to \mathbb{R}^m$, $F(x) = (F_1(x), \ldots, F_m(x))$, we consider the problem of finding an efficient point of F in U with respect to a variable order structure, i.e., a point $x^* \in U$ such that there is no $x \in U \backslash \{x^*\}$ with $F(x^*) - F(x) \in K(x^*)$. We denote this problem as follows:

$$K - \min_{x \in U} F(x). \tag{5.1}$$

A point $x^* \in U$ is said to be weakly efficient for the problem defined in (5.1) iff $F(x) - F(x^*) \notin -\text{int}K(x^*)$, for all $x \in U$.

Assume that F is a differentiable function for all $x \in U$, let us denote the Jacobian of F by $JF(x) := (\nabla F_1(x), \ldots, \nabla F_m(x))$, where $\nabla F_i : \mathbb{R}^n \to \mathbb{R}^n$ is the gradient map of each coordinate function $F_i : \mathbb{R}^n \to \mathbb{R}$ for $i \in \{1, \ldots, m\}$. In this case, the image of the Jacobian of F at a point $x \in \mathbb{R}^n$ is given by $\text{Im}(JF(x)) := \{JF(x)v = (\langle \nabla F_1(x), v\rangle, \ldots, \langle \nabla F_m(x), v\rangle) : v \in \mathbb{R}^n\}$. A first-order necessary optimality condition for the model presented in (5.1), see Bello Cruz and Bouza Allende [9], is stated as

$$x^* \in U, \qquad \text{Im}(JF(x^*)) \cap -\text{int}K(x^*) = \emptyset. \tag{5.2}$$

A point of U satisfying (5.2) is called *critical or stationary point*. In general, (5.2) is a necessary optimality condition. In the next section we will present functions such that this condition is also sufficient.

5.3 Convexity

In this section, the definitions and properties of convexity in a variable order setting are presented. We introduce the definition of strong convexity for variable order vector optimization and its relations with other concepts related with convexity. First we recall the notion of convex function introduced in [9].

Definition 5.3.1 *Let $U \subset \mathbb{R}^n$ be a non-empty, open and convex set and $F : U \to \mathbb{R}^m$ a vector function.*

i) F is said to be convex with respect to mapping K iff for each $x, y \in U$ and for all $\lambda \in [0,1]$ hold,

$$F(\lambda x + (1-\lambda)y) \in \lambda F(x) + (1-\lambda)F(y) - K(\lambda x + (1-\lambda)y);$$

ii) F is strongly convex with respect to K iff there exists $\hat{u} \in \cap_{z \in U} \text{int}K(z) \cap S(0,1)$ and $\gamma > 0$ such that, for each $x, y \in U$ and for all $\lambda \in [0,1]$,

$$F(\lambda x+(1-\lambda)y) \in \lambda F(x)+(1-\lambda)F(y)-K(\lambda x+(1-\lambda)y)-\frac{1}{2}\lambda(1-\lambda)\|x-y\|^2\gamma\hat{u}. \tag{5.3}$$

Note that, in the particular case where the set-valued application K is constant, the definition of strongly convexity retrieves the notion of strong convexity introduced by Graña Drummond *et al.* in [38]. In the case in which the function is continuously differentiable, the following relations hold.

Proposition 5.3.2 *Assume that K is a closed cone-valued mapping and $F : U \to \mathbb{R}^m$ is a continuously differentiable function, where U is non-empty, open and convex set. Then, F is convex if and only if, for each $x, y \in U$ hold*

$$F(y) - F(x) - JF(x)(y - x) \in K(x). \tag{5.4}$$

Furthermore, F is strongly convex if and only if, for each $x, y \in U$

$$JF(x)(y - x) \in (F(y) - F(x)) - K(x) - \frac{1}{2}\|x - y\|^2\gamma\hat{u}. \tag{5.5}$$

Proof. First suppose that F is convex, then we have

$$F(\lambda x + (1 - \lambda)y) \in \lambda F(x) + (1 - \lambda)F(y) - K(\lambda x + (1 - \lambda)y),$$

after some algebraic manipulations, taking into account that $K(\lambda x + (1 - \lambda)y)$ is a cone, we obtain,

$$\frac{F(\lambda x + (1 - \lambda)y) - F(x)}{1 - \lambda} \in F(x) - F(y) - K(\lambda x + (1 - \lambda)y).$$

As F is a continuously differentiable function, it holds that

$$\lim_{\lambda \to 1} \frac{F(\lambda x + (1 - \lambda)y) - F(x)}{1 - \lambda} \to JF(x)(y - x).$$

Letting $\lambda \to 1$, recall that K is a map closed, follows (5.4).

Conversely, suppose holds (5.4). Let $x(\lambda) = \lambda x + (1 - \lambda)y$, making $x = x(\lambda)$ in (5.4) we obtain

$$F(y) - F(x(\lambda)) - \lambda JF(x(\lambda))(y - x) \in K(x(\lambda)). \tag{5.6}$$

Similarly, making $x = x(\lambda)$ and $y = x$ in (5.4), we obtain

$$F(x) - F(x(\lambda)) - (1 - \lambda)JF(x(\lambda))(y - x) \in K(x(\lambda)). \tag{5.7}$$

Multiplying (5.6) by $1 - \lambda$, (5.7) by λ and summing up the resulting expressions, and taking account that $K(x(\lambda))$ is a cone, follows the desired result.

For the second part, let us firstly suppose that F is strongly convex. Thus, there exist $\hat{u} \in \cap_{z \in U} \mathrm{int}K(z) \cap S(0, 1)$ and $\gamma > 0$ such that, for each $x, y \in U$ and for all $\lambda \in [0, 1]$.

$$F(\lambda x + (1 - \lambda)y) \in \lambda F(x) + (1 - \lambda)F(y) - K(\lambda x + (1 - \lambda)y) - \frac{1}{2}\lambda(1 - \lambda)\|x - y\|^2\gamma\hat{u}.$$

Recalling that $K(\lambda x+(1-\lambda)y)$ is a cone, for all $\lambda \in]0,1[$ and $x, y \in U$, last inclusion implies that

$$\frac{F(\lambda x+(1-\lambda)y)-F(x)}{1-\lambda} \in F(y)-F(x)-K(\lambda x+(1-\lambda)y)-\frac{1}{2}\lambda\|x-y\|^2\gamma\hat{u}.$$

Thus, recall that K is a closed mapping, taking limits for $\lambda \to 1$ we obtain (5.5).

Now, let us assume that there exists $\hat{u} \in \cap_{z\in U}\mathrm{int}K(z) \cap S(0,1)$ such that (5.5) holds for each $x, y \in U$. Take $\lambda \in [0,1]$ and $x(\lambda) := \lambda x + (1-\lambda)y$. Making $x = x(\lambda)$ in (5.5), it is easy to see that

$$-\lambda JF(x(\lambda))(x-y) \in (F(y)-F(x(\lambda))-K(x(\lambda))-\lambda^2\frac{1}{2}\|x-y\|^2\gamma\hat{u}. \tag{5.8}$$

Similarly, making $x = x(\lambda)$ and $y = x$, we obtain

$$(1-\lambda)JF(x(\lambda))(x-y) \in (F(x)-F(x(\lambda)))-K(x(\lambda))-(1-\lambda)^2\frac{1}{2}\|x-y\|^2\gamma\hat{u}. \tag{5.9}$$

Therefore, taking into account that $K(x)$ is a convex cone, the desired result follows by multiplying (5.8) by $1-\lambda$, (5.9) by λ and summing up the resulting expressions. ■

Since the previously result will be widely needed is the rest of the work, we assume, from now on, that K is a closed mapping. In the next proposition we demonstrate a condition of the twice continuously differentiable strongly convex function.

Proposition 5.3.3 *Let $U \subset \mathbb{R}^n$ be an open and convex set, $F : U \to \mathbb{R}^m$ a twice continuously differentiable and strongly convex function on U. Then, for each $x \in U$ and $d \in \mathbb{R}^n$,*

$$\langle w, d^T D^2F(x)d\rangle \geq \gamma\langle w, \hat{u}\rangle\|d\|^2, \text{ for all } w \in G(x),$$

where $d^T D^2F(x)d := (d^T\nabla^2F_1(x)d, \dots, d^T\nabla^2F_m(x)d)$, with $\nabla^2F_i(x)$ being the Hessian of $F_i : \mathbb{R}^n \to \mathbb{R}$, for all $i = 1, \dots, m$ and $\hat{u} \in \cap_{z\in U}\mathrm{int}K(z) \cap S(0,1)$ fulfills the strong convexity condition given in (5.3).

Proof. By the Taylor's second order expansion in each coordinate function F_i, we have

$$F_i(y)-F_i(x)-\nabla F_i(x)(y-x) = \frac{1}{2}(y-x)^T\nabla^2F_i(x)(y-x)+o_i(\|x-y\|^2),$$

for all $i = 1, \dots, m$. Comparing with (5.5), we get

$$\frac{1}{2}(y-x)^TD^2F(x)(y-x)+o(\|x-y\|^2) \in \frac{1}{2}\|x-y\|^2\gamma\hat{u}+K(x), \text{ for all } y \in \mathbb{R}^n, \tag{5.10}$$

where $\hat{u} \in \cap_{x\in U}\mathrm{int}K(x) \cap S(0,1)$ and $o(\|x-y\|^2) := (o_1(\|x-y\|^2), \dots, o_m(\|x-y\|^2))$. Now, taking $y = x+td$ for all $t \in \mathbb{R}_{++}$ and $d \in \mathbb{R}^n$, inclusion (5.10) becomes $t^2d^TD^2F(x)d+o(t^2) \in$

$t^2\|d\|^2\ \gamma\hat{u} + K(x)$. Since $K(x)$ is a closed and convex cone, dividing the last inclusion by t^2 and letting t goes to zero, we obtain

$$d^T D^2 F(x) d \in \|d\|^2\ \gamma\hat{u} + K(x), \text{ for all } d \in \mathbb{R}^n, \tag{5.11}$$

from which follows the desired result. ■

Next proposition shows some properties of critical and efficient points of strongly convex functions.

Proposition 5.3.4 *Let $U \subset \mathbb{R}^n$ be an open and convex set and $F : U \to \mathbb{R}^m$ be a continuously differentiable function. Then, the following results hold:*

(i) *If F is convex and $x^* \in U$ is a critical point for F, then x^* is a weakly efficient point.*

(ii) *If F is strongly convex and $x^* \in U$ is a critical point for F, then x^* is an efficient point.*

(iii) *Let x_1^* and x_2^* be two distinct critical points. If F is strongly convex, then $F(x_1^*) \neq F(x_2^*)$.*

Proof. To prove (i), let $x^* \in U$ a critical point and let suppose too there exist x such that

$$F(x) - F(x^*) = k_1 \in -\text{int}K(x^*)$$

Since F is convex, from proposition 5.3.2 we have

$$F(x) - F(x^*) - JF(x^*)(x^* - x) = k_2 \in K(x^*).$$

Then we have $JF(x^*)(x^* - x) = k_1 - k_2 \in$

For (ii), let us suppose that there is a $\bar{x} \in U\backslash\{x^*\}$ satisfying

$$F(\bar{x}) - F(x^*) \in -K(x^*). \tag{5.12}$$

Since F is strongly convex, as $\bar{x} \neq x^*$, Proposition 5.3.2 implies that for all $\bar{x} \in U \setminus \{x^*\}$

$$JF(x^*)(\bar{x} - x^*) - F(\bar{x}) + F(x^*) \in -K(x^*) - \frac{1}{2}\|x^* - \bar{x}\|^2\gamma\hat{u} \subset -\text{int}K(x^*). \tag{5.13}$$

Combining (5.12) with (5.13) it follows that $JF(x^*)(\bar{x}-x^*) \in -\text{int}K(x^*)$. But this is absurd, because x^* is a critical point, which concludes the proof.

For the last part, take two distinct critical points x_1^* and x_2^* such that $F(x_1^*) = F(x_2^*)$. By the strong convexity of F, $JF(x_1^*)(x_2^* - x_1^*) \in F(x_2^*) - F(x_1^*) - K(x_1^*) - 1/2\|x_1^* - x_2^*\|^2\gamma\hat{u}$, see (5.5). As $F(x_2^*) = F(x_1^*)$ and $\hat{u} \in \cap_{x\in U}\text{int}(K(x))$, it holds $JF(x_1^*)(x_2^* - x_1^*) \in -\text{int}K(x_1^*)$, contradicting that x_1^* is a critical point. ■

As expected, strong convexity provides a sufficient condition for the efficiency. Next example shows that the natural extension of the strict convexity, namely, $F(\lambda x+(1-\lambda)x^*) \in \lambda F(x) + (1-\lambda)F(x^*) - \text{int}(K(\lambda x + (1-\lambda)x^*))$, for all $\lambda \in (0,1)$, is not a sufficient condition for the efficiency, as holds in the particular case where the order is given by a fixed cone (see [38]).

Example 5.3.5 Let $F : [0,1] \to \mathbb{R}^2$, where $F(x) = (x^2 - 2x, x^3 - 2x)$ and the ordering structure given by $K : [0,1] \rightrightarrows \mathbb{R}^2$, where $K(x)$ is the cone generated by the convex hull of the vectors $[1, x]$ and $[1, 2x + 1]$.

After some calculations it holds that

$$F(\lambda x + (1-\lambda)x^*) - \lambda F(x) - (1-\lambda)F(x^*) = -\lambda(1-\lambda)(x - x^*)^2 \left[1 \ [(\lambda+1)x + (2-\lambda)x^*]\right]^T.$$

Then $\lambda x + (1-\lambda)x^* \leq (\lambda+1)x + (2-\lambda)x^*$ and $(\lambda+1)x + (2-\lambda)x^* \leq 2[\lambda x + (1-\lambda)x^*] + 1$ holds, recall that $0 \leq x, x^* \leq 1$. As the equality in the first relation holds if and only if $x = x^* = 0$ and the second if $x = x^* = 1$, $\left[1 \ \ [(\lambda+1)x + (2-\lambda)x^*]\right]^T \in \operatorname{int}(K(\lambda x + (1-\lambda)x^*)$ and, therefore, F is a strictly convex function.

Note that since $JF(0) = (-2, -2)$ and -$K(0)$ is generated by $[-1, 0]$, $[-1, -1]$, 0 is a stationary point. However $F(1) - F(0) = [-1 \ \ -1]^T \in -K(0)$. Hence it is not an efficient point.

The previous example shows that the strong convexity is the natural assumptions for algorithms using second order derivatives. This is the main goal of the next section.

5.4 Newton direction for variable order

In this section we will extend the formula of Newton's method to the case of minimizing an strongly convex function according to the variable ordering application. A quadratic approximation of the scalarized objective function will be used. It is proven that this problem has a unique solution and, hence, that the Newton's direction is well defined. Further properties of the direction such as the characterization of efficiency, the continuity of the direction, upper bounds and descend of the objective functions will be obtained. From now on, we will assume that F is strongly convex and twice continuously differentiable on the open convex set U.

Given $x \in U$ and $w \in G(x)$, let us consider the function $\varphi_w : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$\varphi_w(x, v) := \langle w, JF(x)v \rangle + \frac{1}{2}\langle w, v^T D^2 F(x) v \rangle, \tag{5.14}$$

where $v^T D^2 F(x) v := (v^T \nabla^2 F_1(x) v, \ldots, v^T \nabla^2 F_m(x) v)$ and $w \in G(x)$. We define the Newton direction $s : U \to \mathbb{R}^n$ as

$$s(x) := \operatorname*{argmin}_{v \in \mathbb{R}^n} \max_{w \in G(x)} \varphi_w(x, v). \tag{5.15}$$

In the next result, we show the well-definedness of the Newton direction defined above.

Proposition 5.4.1 *Given $x \in U$ and $w \in G(x)$, then the function $\varphi_w(x, \cdot)$ is strongly convex with modulus $\rho(x) := \min_{w \in G(x)} \gamma\langle w, \hat{u} \rangle$, where $\hat{u} \in \cap_{x \in U} \operatorname{int} K(x)$ is such that* (5.3) *holds. Moreover, the Newton direction defined in* (5.15) *is well-defined.*

Proof. First note that since $G(x)$ is compact and $\hat{u} \in \mathrm{int}K(x)$, $\rho(x)$ is well-defined and positive.

Now take $v_1,\ v_2 \in \mathbb{R}^n$, $\lambda \in [0,1]$ and define $v(\lambda) := \lambda v_1 + (1-\lambda)v_2$. From the definition of φ_w, we have

$$\varphi_w(x, v(\lambda)) = \langle w, JF(x)[v(\lambda)]\rangle + \frac{1}{2}\left\langle w, [v(\lambda)]^T D^2F(x)[v(\lambda)]\right\rangle.$$

Simple algebraic manipulations allow us to obtain,

$$\begin{aligned}\varphi_w(x, v(\lambda)) &= \lambda\langle w, JF(x)v_1\rangle + (1-\lambda)\langle w, JF(x)v_2\rangle \\ &\qquad + \frac{1}{2}\lambda^2\langle w, v_1^T D^2F(x)v_1\rangle + \frac{1}{2}(1-\lambda)^2\langle w, v_2^T D^2F(x)v_2\rangle.\end{aligned}$$

Adding and subtracting $\frac{1}{2}\lambda(1-\lambda)\langle w, (v-s)^T D^2F(x)(v-s)\rangle$ to the last equality, and again using simple algebraic manipulations, we get

$$\begin{aligned}\varphi_w(x, v(\lambda)) &= \lambda\langle w, JF(x)v_1\rangle + (1-\lambda)\langle w, JF(x)v_2\rangle + \frac{1}{2}\lambda\langle w, v_1^T D^2F(x)v_1\rangle \\ &\quad + \frac{1}{2}(1-\lambda)\langle w, v_2^T D^2F(x)v_2\rangle - \frac{1}{2}\lambda(1-\lambda)\langle w, (v_1-v_2)^T D^2F(x)(v_1-v_2)\rangle.\end{aligned}$$

Taking into account that F is strongly convex and using Proposition 5.3.3 with $d = v_1 - v_2$, last equality yields

$$\varphi_w(x, v(\lambda)) \leq \lambda\varphi_w(x, v_1) + (1-\lambda)\varphi(x, v_2) - \frac{1}{2}\lambda(1-\lambda)\gamma\langle w, \hat{u}\rangle\|v_1 - v_2\|^2, \text{ for all } w \in G(x).$$

As by the definition of $\rho(x)$

$$\gamma\langle w, \hat{u}\rangle \geq \rho(x) > 0, \tag{5.16}$$

the first part of the proposition holds.

For proving the second part, note that there exists $\bar{w} \in G(x)$ such that $\max_{w\in G(x)} \varphi_w(x, v) = \varphi_{\bar{w}}(x, v)$, because $G(x)$ is compact. Furthermore, since $\varphi_w(x, \cdot)$ is strongly convex for each $w \in G(x)$, in particular, $\varphi_{\bar{w}}(x, \cdot)$ is also strongly convex, which is enough to conclude the proof of the proposition. ∎

Denote by $\theta(x)$ the optimal value of the minimization problem in (5.15), i.e.,

$$\begin{aligned}\theta(x) &= \min_{v\in\mathbb{R}^n} \max_{w\in G(x)} \left[\langle w, JF(x)v\rangle + \frac{1}{2}\langle w, v^T D^2F(x)v\rangle\right] \\ &= \max_{w\in G(x)} \left[\langle w, JF(x)s(x)\rangle + \frac{1}{2}\langle w, s(x)^T D^2F(x)s(x)\rangle\right].\end{aligned} \tag{5.17}$$

Let us give some results relating a critical point x to its corresponding Newton direction $s(x)$ and $\theta(x)$.

Proposition 5.4.2 *Let $\theta : U \to \mathbb{R}$ and $s : U \to \mathbb{R}^n$ be given by* (5.17) *and* (5.15) *respectively. Then the following holds*

(i) *$\theta(x) \leq 0$ for all $x \in U$.*

(ii) *x is not a critical point iff $\theta(x) < 0$, furthermore $s(x) \neq 0$.*

(iii) *If x is a critical point then $\theta(x) = 0$ and $s(x) = 0$.*

Proof. Since we have $\theta(x) \leq \max_{w \in G(x)} \langle w, JF(x)v\rangle + 1/2\langle w, v^T D^2F(x)v\rangle$ for all $x \in U$ and $v \in \mathbb{R}^n$, making $v = 0$ we obtain *(i)*.

In order to prove *(ii)* first we suppose that x is not critical. Then there is a $\hat{v} \in \mathbb{R}^n$ such that $JF(x)\hat{v} \in -\mathrm{int}K(x)$, which means that $\langle w, JF(x)\hat{v}\rangle < 0$ for all $w \in G(x)$. By the definition of $\theta(x)$ we have, for $v = t\hat{v}$ with $t > 0$,

$$\theta(x) \leq t \max_{w \in G(x)} \langle w, JF(x)\hat{v}\rangle + \frac{t^2}{2}\langle w, \hat{v}^T D^2F(x)\hat{v}\rangle.$$

For $t > 0$ small enough the right hand side of the last inequality is negative and therefore $\theta(x) < 0$. Conversely supposing that $\theta(x) < 0$, from Proposition 5.3.3 we have

$$\langle w, JF(x)s(x)\rangle \leq \langle w, J(x)s(x)\rangle + \langle w, s(x)^T D^2F(x)s(x)\rangle \leq \theta(x) < 0, \quad \text{for all}\, w \in G(x).$$

The above inequality implies $JF(x)s(x) \in -\mathrm{int}K(x)$, concluding that x is not a critical point, and since $\theta(x) < 0$, the solution of that problem cannot be $s(x) = 0$, concluding *(ii)*.

The last part is a clear consequence of *(i)* and *(ii)*. ■

The next result is used to proof the continuity of the functions $\theta(x)$ and $s(x)$.

Lemma 5.4.3 *If $\mathrm{int}K(x) \neq \emptyset$ for all $x \in U$, then the following holds:*

(i) *$\|s(x)\| \leq \frac{2}{\gamma\langle w, \hat{u}\rangle}\|JF(x)\|$, for all $w \in G(x)$.*

(ii) *Suppose that $W \subset \mathbb{R}^m$ is a bounded set. If for some $L \geq 0$, $\mathrm{dist}(G(x_1), G(x_2)) \leq L\|x_1 - x_2\|$ for all x_1, $x_2 \in U$, then the auxiliary function $\eta(x, z) := \max_{w \in G(x)}\langle w, z\rangle$ is a Lipschitz function for all $(x, z) \in U \times W$, i.e., for all z_1, $z_2 \in W$*

$$|\eta(x_1, z_1) - \eta(x_2, z_2)| \leq L\|x_1 - x_2\| + \|z_1 - z_2\|. \tag{5.18}$$

Proof. To prove the first part we use the fact that $\theta(x) \leq 0$. Then

$$\langle w,\, JF(x)s(x)\rangle + \frac{1}{2}\langle w,\, s(x)^T \nabla^2 F(x)s(x)\rangle \leq \theta(x) \leq 0.$$

By Proposition 5.3.3 and Holder inequality, we obtain

$$\frac{\gamma\langle w, \hat{u}\rangle}{2}\|s(x)\|^2 \leq \frac{1}{2}\langle w,\, s(x)^T \nabla^2 F(x)s(x)\rangle \leq \|w\|\|JF(x)s(x)\| \leq \|JF(x)\|\|s(x)\|.$$

Dividing by $\frac{\gamma\langle w,\hat{u}\rangle}{2} > 0$, recall (5.16), the first part is obtained.

Now to prove (ii), since $\eta(x,z)$ is a sublinear function, holds the following inequation.

$$\eta(x_1,z_2) - \eta(x_1,z_2 - z_1) \leq \eta(x_1,z_1) \leq \eta(x_1,z_2) + \eta(x_1,z_1 - z_2). \tag{5.19}$$

We first analyze the left side of the inequality above, which can be equivalently written as

$$\eta(x_1,z_2) - \eta(x_2,z_2) - \eta(x_1,z_2 - z_1) \leq \eta(x_1,z_1) - \eta(x_2,z_2). \tag{5.20}$$

From the CauchySchwartz inequality and definition of $G(x)$ we have

$$\eta(x,z) = \max_{w\in G(x)} \langle w,z\rangle \leq \max_{w\in G(x)} \|w\|\|z\| = \|z\|. \tag{5.21}$$

Thus, combining (5.20) and (5.21), we obtain

$$\eta(x_1,z_2) - \eta(x_2,z_2) - \|z_2 - z_1\| \leq \eta(x_1,z_1) - \eta(x_2,z_2). \tag{5.22}$$

Now we pay attention on the expression $\eta(x_1,z_2) - \eta(x_2,z_2)$ As $G(x_1)$ and $G(x_2)$ are compacts, then there exists $y_1 \in G(x_1)$ and $y_2 \in G(x_2)$, satisfying respectively $\eta(x_1,z_2) = \langle y_1,z_2\rangle$ $\eta(x_2,z_2) = \langle y_2,z_2\rangle$. Taking into account that $\operatorname{dist}(G(x_1),G(x_2)) \leq L\|x_1 - x_2\|$, there exist $w_1 \in G(x_1)$ and $\mu_1 \in B(0,1)$ such that $y_2 = w_1 + L\|x_1 - x_2\|\mu_1$. Hence,

$$\eta(x_1,z_2) - \eta(x_2,z_2) = \langle y_1,z_2\rangle - \langle y_2,z_2\rangle = \langle y_1 - w_1 - L\|x_1 - x_2\|\mu_1, z_2\rangle .$$

Recalling that $w_1 \in G(x1)$ and $\eta(x_1,z_2) = \langle y_1,z_2\rangle$, from definition of ρ, we have $\langle y_1,z_2\rangle \geq \langle w_1,z_2\rangle$, this implies the last inequality becomes

$$\eta(x_1,z_2) - \eta(x_2,z_2) \geq -\langle L\|x_1 - x_2\|\mu_1, z_2\rangle \geq -L\|x_1 - x_2\|\|z_2\|. \tag{5.23}$$

Note that as W is bounded, then for some $M > 0$, $\|w\| \leq M$ for all $w \in W$.Combining (5.22) and (5.23), and defining $\hat{L} := LM$, we obtain

$$\eta(x_1,z_1) - \eta(x_2,z_2) \geq -\hat{L}\|x_1 - x_2\| - \|z_2 - z_1\|. \tag{5.24}$$

On the other hand, analyzing the right side of inequality (5.19) leads to

$$\eta(x_1,z_1) - \eta(x_2,z_2) \leq \eta(x_1,z_2) - \eta(x_2,z_2) + \eta(x_1,z_1 - z_2).$$

Using the Cauchy-Schwartz inequality, we have

$$\eta(x_1,z_1) - \eta(x_2,z_2) \leq \eta(x_1,z_2) - \eta(x_2,z_2) + \|z_1 - z_2\|. \tag{5.25}$$

Analogously, taking $w_2 \in G(x_2)$ and $\mu_2 \in B(0,1)$ such that $y_1 = w_2 + L\|x_1x_2\|\mu_2$, we obtain

$$\eta(x_1,z_2) - \eta(x_2,z_2) = \langle y_1,z_2\rangle - \langle y_2,z_2\rangle = \langle w_2 - y_2 + L\|x_1 - x_2\|\mu_2, z_2\rangle ,$$

since $\rho(x_2, w_2) = \langle y_2, z_2 \rangle$, this implies $\langle y_2, z_2 \rangle \geq \langle w_2, z_2 \rangle$, hence

$$\eta(x_1, z_2) - \eta(x_2, z_2) \leq \langle L\|x_1 - x_2\|\mu_2, z_2 \rangle \leq L\|x_1 - x_2\|\|\mu_2\|. \tag{5.26}$$

Taking $\hat{L} = LM$, the combination of (5.25) and (5.26) implies

$$\eta(x_1, z_1) - \eta(x_2, z_2) \leq \hat{L}\|x_1 - x_2\|\| + \|z_1 - z_2\|.$$

Together with (5.24), evidently it follows that

$$|\eta(x_1, z_1) - \eta(x_2, z_2)| \leq \hat{L}\|x_1 - x_2\|\| + \|z_1 - z_2\|.$$

■

Corollary 5.4.4 *Let $\{x^k\}$ be a sequence which satisfies $x^k \in U$, for all k. If G is a closed map and $x^k \to x^*$, $x^* \in U$, then there exists $\hat{\rho} > 0$ such that $\rho(x^k) > \hat{\rho}$ for all k. Moreover, $\{s(x^k)\}$ is bounded.*

Proof. Since $G(x^k)$ is compact, there exists w^k such that $\rho(x^k) = \langle w^k, \hat{u} \rangle$. As $\hat{u} \in \cap_{x \in U} K(x)$ and $\{w^k\}$ is a bounded sequence, the accumulation points of $\rho(x^k)$ are non-negative. Due to $x^k \to x^*$ and G is closed, the limit points of w^k are elements of $G(x^*)$. Hence

$$\lim_k \rho(x^k) = \lim_k \langle w^k, \hat{u} \rangle = \langle \hat{w}, \hat{u} \rangle > 0.$$

Recalling that the set of accumulation points of a sequence is closed, it is clear that $\rho(x^k) > \hat{\rho}$, for some $\hat{\rho} > 0$.

For the second part, by Lemma 5.4.3(i) we have, $\|s(x^k)\| \leq (2/\gamma\langle w^k, u \rangle)\|JF(x^k)\|$, for all k. As proven before $\gamma\langle w^k, \hat{u} \rangle \geq \rho(x^k) > \hat{\rho}$. Then we have that $\{s(x^k)\}$ is also bounded. ■

Proposition 5.4.5 *The following holds*

(i) *If G is a closed map, then $\theta(x)$ is an upper semi-continuous function.*

(ii) *If for some $L \geq 0$, $\operatorname{dist}(G(x_1), G(x_2)) \leq L\|x_1 - x_2\|$, for all $x_1, x_2 \in U$, G is a closed map, then $\theta(\cdot)$ is a lower semi-continuous function.*

(iii) *If for some $L \geq 0$, $\operatorname{dist}(G(x_1), G(x_2)) \leq L\|x_1 - x_2\|$, for all $x_1, x_2 \in U$ and $\theta(\cdot)$ is a continuous function, then $s(\cdot)$ is closed. Furthermore, $s(\cdot)$ is a continuous function.*

Proof. Let $\{x^k\}$ be a sequence in U that converges to $\bar{x} \in U$. For each k, we consider $w^k \in G(x^k)$ such that

$$\langle w^k, JF(x^k)s(\bar{x}) \rangle + \frac{1}{2}\langle w^k, s(\bar{x})^T D^2 F(x^k) s(\bar{x}) \rangle =$$
$$\max_{w \in G(x^k)} \langle w, JF(x^k)s(\bar{x}) \rangle + \frac{1}{2}\langle w, s(\bar{x})^T D^2 F(x^k) s(\bar{x}) \rangle.$$

It is clear that, by the definition of θ, we have

$$\theta(x^k) \leq \langle w^k, JF(x^k)s(\bar{x})\rangle + \frac{1}{2}\langle w^k, s(\bar{x})^T D^2F(x^k)s(\bar{x})\rangle, \text{ for all } k. \tag{5.27}$$

As $\{w^k\}$ is bounded, it has a convergent sub-sequence. Since G is a closed map, we will suppose without loss of generality that $\{w^k\}$ converges to $\bar{w} \in G(\bar{x})$. Thus, taking limits in (5.27),

$$\begin{aligned}
\limsup_{k\to\infty} \theta(x^k) \quad &\leq \quad \limsup_{k\to\infty} \left[\langle w^k, JF(x^k)s(\bar{x})\rangle + \frac{1}{2}\langle w^k, s(\bar{x})^T D^2F(x^k)s(\bar{x})\rangle\right] \\
&= \langle \bar{w}, JF(\bar{x})s(\bar{x})\rangle + \frac{1}{2}\langle \bar{w}, s(\bar{x})^T D^2F(\bar{x})\rangle \\
&\leq \max_{w\in G(\bar{x})} \langle w, JF(\bar{x})s(\bar{x})\rangle + \frac{1}{2}\langle w, s(\bar{x})^T D^2F(\bar{x})s(\bar{x})\rangle = \theta(\bar{x}),
\end{aligned}$$

which proves (i).
Now for (ii), take $w^k \in G(x^k)$ as

$$\begin{aligned}
&\langle w^k, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle w^k, s(x^k)^T D^2F(x^k)s(x^k)\rangle = \\
&\qquad\qquad \max_{w\in G(x^k)} \langle w, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle w, s(x^k)^T D^2F(x^k)s(x^k)\rangle,
\end{aligned}$$

and $u \in \cap_{x\in U}\ \mathrm{int}K(x) \cap S(0,1)$ such that the condition given in (5.3) is fulfilled. By Corollary 5.4.4 we have $\{s(x^k)\}$ is also bounded. Then by the definition of the application $\theta(\cdot)$, for all k

$$\begin{aligned}
\theta(\bar{x}) \quad &\leq \quad \max_{w\in G(\bar{x})} \langle w, JF(\bar{x})s(x^k)\rangle + \frac{1}{2}\langle w, s(x^k)^T D^2F(\bar{x})s(x^k)\rangle \\
&= \quad \theta(x^k) + \max_{w\in G(\bar{x})} \langle w, JF(\bar{x})s(x^k)\rangle + \frac{1}{2}\langle w, s(x^k)^T D^2F(\bar{x})s(x^k)\rangle \\
&- \quad [\max_{w\in G(x^k)} \langle w, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle w, s(x^k)^T D^2F(x^k)s(x^k)\rangle].
\end{aligned} \tag{5.28}$$

From Lemma 5.4.3(ii), making $x_1 = \bar{x}$, $x_2 = x^k$, $z_1 = JF(\bar{x})s(x^k) + 1/2s(x^k)^T D^2F(\bar{x})s(x^k)$ and $z_2 = JF(x^k)s(x^k) + 1/2s(x^k)^T D^2F(x^k)s(x^k)$, for all k, we have

$$\begin{aligned}
&\left|\eta(\bar{x}, JF(\bar{x})s(x^k) + \frac{1}{2}s(x^k)^T D^2F(\bar{x})s(x^k)) - \eta(x^k, JF(x^k)s(x^k) + \frac{1}{2}s(x^k)^T D^2F(x^k)s(x^k))\right| \\
&\quad \leq L\|\bar{x} - x^k\| \left\|JF(\bar{x})s(x^k) + \frac{1}{2}s(x^k)^T D^2F(\bar{x})s(x^k)\right\| + \left\|\big(JF(\bar{x}) - JF(x^k)\big)s(x^k)\right\| \\
&\qquad\qquad + \left\|s(x^k)^T \big(D^2F(\bar{x}) - D^2F(x^k)\big)s(x^k)\right\|.
\end{aligned} \tag{5.29}$$

Substituting the above inequality in (5.28) and recalling that $\eta(x,z) := \max_{w\in G(x)}\langle w, z\rangle$, we get

$$\begin{aligned}\theta(\bar{x}) &\leq \theta(x^k) + L\|\bar{x}-x^k\|\big\|JF(\bar{x})v^k + \frac{1}{2}{v^k}^T D^2F(\bar{x})s(x^k)\big\| \\ &+ \big\|\big(JF(\bar{x}) - JF(x^k)\big)s(x^k)\big\| + \big\|s(x^k)^T\big(D^2F(\bar{x}) - D^2F(x^k)\big)s(x^k)\big\|. \quad (5.30)\end{aligned}$$

As $s(x^k)$ is a bounded sequence and F is twice continuously differentiable in U, then taking limit in (5.30), we obtain

$$\begin{aligned}\theta(\bar{x}) &\leq \liminf_{k\to\infty} \theta(x^k) + L\|\bar{x}-x^k\|\big\|JF(\bar{x})s(x^k) + \frac{1}{2}s(x^k)^T D^2F(\bar{x})s(x^k)\big\| \\ &+ \big\|\big(JF(\bar{x}) - JF(x^k)\big)s(x^k)\big\| + \big\|s(x^k)^T\big(D^2F(\bar{x}) - D^2F(x^k)\big)s(x^k)\big\| \\ &= \liminf_{k\to\infty} \theta(x^k),\end{aligned}$$

concluding the lower semi-continuity of θ.

Finally, to prove (iii), let $\{x^k\} \to x^*$. Note that, by Corollary 5.4.4 we obtain that $\{s(x^k)\}$ is bounded. Hence, it has a convergent sub-sequence and, without loss of generality, we suppose that $s(x^k) \to \bar{s}$. We must show that $\bar{s}$ solves (5.15). Let $\hat{w} \in G(\bar{x})$ be such that,

$$\langle \hat{w}, JF(\bar{x})\bar{v}\rangle + \frac{1}{2}\langle \hat{w}, \bar{v}^T D^2F(\bar{x})\bar{v}\rangle = \eta\big(x, JF(\bar{x})\bar{v} + \frac{1}{2}\bar{v}^T D^2F(\bar{x})\bar{v}\big).$$

By the hypothesis of this part, we know that $\mathrm{dist}(G(x^k), G(\bar{x})) \leq L\|x^k - \bar{x}\|$, for each $k \in \mathbb{N}$. Then, there is $\hat{w}^k \in G(x^k)$ such that $\hat{w} = \hat{w}^k + L\|\bar{x} - x^k\|\mu^k$, where $\mu^k \in B(0,1)$. Hence for all k it holds

$$\begin{aligned}&\theta(x^k) \geq \langle \hat{w}^k, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle \hat{w}^k, s(x^k)^T D^2F(x^k)s(x^k)\rangle = \\ &\quad \langle \hat{w} - L\|\bar{x}-x^k\|\mu^k, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle \hat{w} - L\|\bar{x}-x^k\|\mu^k, s(x^k)^T D^2F(x^k)s(x^k)\rangle. \quad (5.31)\end{aligned}$$

Since μ^k is bounded and θ is continuous by hypothesis, then taking limits in (5.31) we obtain,

$$\begin{aligned}\lim_{k\to\infty}\theta(x^k) &\geq \lim_{k\to\infty}\langle \hat{w} - L\|\bar{x}-x^k\|\mu^k, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle \hat{w} - L\|\bar{x}-x^k\|\mu^k, s(x^k)^T D^2F(x^k)s(x^k)\rangle \\ \theta(\bar{x}) &\geq \langle \hat{w}, JF(\bar{x})\bar{s}\rangle + \frac{1}{2}\langle \hat{w}, \bar{s}^T D^2F(\bar{x})\bar{s}\rangle = \eta\big(x, JF(\bar{x})\bar{s} + \frac{1}{2}\bar{s}^T D^2F(\bar{x})\bar{s}\big),\end{aligned}$$

which prove the closeness of $s(\cdot)$ as a map.

Since all accumulation points of $\{s(x^k)\}$ solve (5.15) and this problem has a unique solution, recall Proposition 5.4.1, $s(x^*)$ is the limit of the whole sequence $\{s(x^k)\}$. Therefore $s(\cdot)$ is a continuous function. ■

Remark 5.4.6 It is clear that the continuity of the function $\theta(\cdot)$ can be substituted by the hypotheses needed for its lower and upper semi-continuity. Using (i) and (ii), the result stated in (iii) also holds true if G is closed and $dist(G(x_1), G(x_2)) \leq L\|x_1 - x_2\|$ for some $L \geq 0$ and for all $x_1, x_2 \in U$.

The classical Armijo rule for the Newton direction $s(x)$ at a noncritical point for a scalar function $F : U \to \mathbb{R}$ reads

$$F(x + ts(x)) \leq F(x) + \beta t JF(x)^T s(x),$$

with $\beta \in]0, \frac{1}{2}[$. Note that, in this case, since $JF(x) + D^2F(x)s(x) = 0$, $\theta(x) = 1/2JF(x)^T s(x)$. Then the Armijo rule can be rewritten as

$$F(x + ts(x)) \leq F(x) + \sigma\theta(x),$$

where $\sigma = 2\theta \in (0, 1)$.

Combining the results proven in the previous propositions and generalizing the Armijo-rule to variable order optimization problems, next we show that, given a non critical point, the Armijo-rule is well-defined. This fact guarantees that for an arbitrary s such that $JF(x)s \in -\text{int}K(x)$, it is possible to compute a new feasible solution in which the value of the objective function decreases.

Proposition 5.4.7 *Let $x \in U$ and $s \in \mathbb{R}^n$ be such that $JF(x)s \in -intK(x)$. Then, for any $0 < \sigma < 1$ there exist $0 < \bar{t} < 1$ such that*

$$x + ts \in U \quad \text{and} \quad F(x + ts) \in F(x) + \sigma t JF(x)s - K(x),$$

hold for all $t \in (0, \bar{t}]$.

Proof. Since F is differentiable, the Taylor's expansion leads to

$$F(x + ts) = F(x) + tJF(x)s + o(t), \tag{5.32}$$

where $\lim_{t\to 0} \|o(t)\|/t = 0$.

Since U is an open set, we have for t small enough that $x + ts \in U$. As $-\text{int}K(x)$ is a cone then, by the hypothesis $(1 - \sigma)JF(x)s \in -\text{int}K(x)$. On the other hand, there is $\bar{t} > 0$ such that for all $t \in (0, \bar{t}]$, $\|o(t)\|$ is small enough and $(1 - \sigma)JF(x)s + o(t) \in -K(x)$. This result, together with (5.32), implies $F(x + ts) \in F(x) + \sigma tJF(x)s - K(x)$, concluding the proof. ∎

Now, under assuming that F is a twice continuously differentiable and strongly convex function, we showed the well-definedness of the backtracking procedure for the Armijo-like rule with the Newton direction.

Corollary 5.4.8 *If $x \in U$ is not a critical point for F, then for $0 < \sigma < 1$ there is $0 < \bar{t} < 1$ such that*

$$\langle w, F(x + ts(x))\rangle \leq \langle w, F(x)\rangle + \sigma t\theta(x)$$

holds for all $w \in G(x)$ and for all $t \in (0, \bar{t}]$.

Proof. By Proposition 5.4.7, we have

$$\langle w, F(x + ts(x))\rangle \leq \langle w, F(x)\rangle + \sigma t\langle w, JF(x)s(x)\rangle$$

for all $w \in G(x)$ and for some $t \in (0, \bar{t}]$. Besides, as $\langle w, s(x)D^2F(x)s(x)\rangle > 0$ for all $w \in G(x)$ and by the definition of $\theta(\cdot)$, we obtain the required result . ■

Due to the already proven properties of $s(\cdot)$, we can use it as a descent direction if the function F is strongly convex. Next section we will present Newton's method and analyze its convergence.

5.5 The Newton method

In this section we present the Newton's method for variable order optimization problems and study the sequence $\{x^k\}$ generated by the approach. First the convergence $\{F(x^k)\}$ is shown. This result is used to ensure that the accumulation points of the sequence $\{x^k\}$, if any, are efficient. Then, it is proven that the algorithm converges.

We start presenting the algorithm.

Algorithm 5.5.1

(i) First choose $\sigma \in (0, 1)$, $x^0 \in U$ and set $k := 0$.

(ii) Given x^k, compute $s^k := s(x^k)$ and $\theta(x^k)$ solution and optimal value respectively of the problem given in (5.15) for $x = x^k$.

(iii) If $s^k = 0$, then stop. Otherwise, compute a step-length t_k such that $x^k + t_k s^k \in U$ and

$$t_k = \max\{2^{-j},\ j \in \mathbb{Z}_+;\ \langle w, F(x^k + 2^{-j}s^k)\rangle \leq \langle w, F(x^k)\rangle + \sigma 2^{-j}\theta(x^k), \forall w \in G(x^k)\}. \tag{5.33}$$

(iv) Define $x^{k+1} = x^k + t_k s^k$, and go to step (ii).

The well-definedness of Algorithm 5.5.1 was proven in Proposition 5.4.7 and Corollary 5.4.8.

Note that whenever Newton's algorithm for vector optimization has a finite termination, the last iteration computes a stationary point. So, from now on, we assume that $\{x^k\}$, $\{s^k\}$ and $\{t^k\}$ are infinite sequences generated by Newton's algorithm. Before we started the convergence analysis, we present the following lemma that, under certain hypotheses, guarantee the full convergence of the sequence $\{F(x^k)\}$ to $F(x^*)$, where x^* is an accumulation point of the sequence $\{x^k\}$.

Lemma 5.5.2 *Assume that*

(i) $\bigcup_{x\in R^n} K(x) \subset \mathcal{K}$, *where* $\mathcal{K}$ *is a closed, convex and pointed cone.*

(ii) *The application* $G(\cdot)$ *is closed.*

(iii) $dist(G(x), G(y)) \leq L\|x-y\|$, *for all,* x, $y \in \mathbb{R}^n$.

If x^* *is an accumulation point of the sequence* $\{x^k\}$, *then* $\lim_{k\to\infty} F(x^k) = F(x^*)$.

Proof. Take $\{x^{i_k}\}$ a sub-sequence of $\{x^k\}$ that converges to x^*. By the Armijo-step rule we have

$$\langle w, F(x^{k+1}) - F(x^k)\rangle \leq \sigma t_k \theta(x^k),$$

for all $w \in G(x^k)$. Hence, x^k is not a critical point and, therefore, by Proposition 5.4.2 we have $\theta(x^k) < 0$. So, we obtain

$$\langle w, F(x^{k+1}) - F(x^k)\rangle < 0, \quad \text{for all} \quad w \in G(x^k) \quad \Rightarrow \quad F(x^{k+1}) - F(x^k) \in -\text{int}K(x^k).$$

As $\bigcup_{x\in R^n} K(x) \subset \mathcal{K}$, we have, for all x, $\text{int}K(x) \subset \mathcal{K}$. So

$$F(x^k) - F(x^{k+1}) \in \text{int}K(x^k)$$

or equivalently: $F(x^k)$ is a decreasing sequence with respect to the cone $\mathcal{K}$.

By the continuity of F, we have

$$\lim_{k\to\infty} F(x^{i_k}) = F(x^*).$$

Particularly, $F(x^*)$ is an accumulation point of $\{F(x^k)\}$. So, due to $\mathcal{K}$ is a closed, convex and pointed cone, $F(x^k) \to F(x^*)$ because if $y^k - y^{k+1} \in \mathcal{K}$, for all k, $\mathcal{K}$ satisfies (i) and $\{y^k\}$ has an accumulation point, y^*, then $\{y^k\} \to y^*$, see [54] for more details. ■

Now we prove that every accumulation point of $\{x^k\}$, if any, is an efficient point.

Proposition 5.5.3 *Suppose that*

(i) $\bigcup_{x\in R^n} K(x) \subset \mathcal{K}$, *where* $\mathcal{K}$ *is a closed, convex and pointed cone.*

(ii) *The application* $G(\cdot)$ *is closed.*

(iii) $\text{dist}(G(x), G(y)) \leq L\|x-y\|$, *for all,* x, $y \in \mathbb{R}^n$.

If $x^* \in U$ *is an accumulation point of* $\{x^k\}$, *then* $\theta(x^*) = 0$ *and* x^* *is an efficient point.*

Proof. Suppose that $x^* \in U$ is an accumulation point of $\{x^k\}$ and let $\{w^k\}$ be a bounded sequence where $w^k \in G(x^k)$. As in the previous lemma, we have for all $w^k \in G(x^k)$

$$\langle w^k, F(x^{k+1}) - F(x^k)\rangle \leq \sigma t_k \theta(x^k) < 0.$$

Since $t^k \in (0,1]$, for all k, the sequence $\{t^k\}$ is bounded. Let us consider the sub-sequences $\{x^{i_k}\}$, $\{t^{i_k}\}$ and $\{w^{i_k}\}$. Without loss of generality we assume that these sub-sequences converge respectively to x^*, t^* and w^*. Note that $w^* \in G(x^*)$, since G is a closed map. Now we have

$$\lim_{k\to\infty} \langle w^{i_k}, F(x^{i_k+1}) - F(x^{i_k})\rangle \leq \lim_{k\to\infty} \sigma t_{i_k}\theta(x^{i_k}) \leq 0. \tag{5.34}$$

By Lemma 5.5.2 we have $\lim_{k\to\infty} F(x^{i_k+1}) - F(x^{i_k}) = 0$. Then, together with (5.34), we obtain

$$\lim_{k\to\infty} \sigma t_{i_k}\theta(x^{i_k}) = 0. \tag{5.35}$$

Now we have two cases: $t^* > 0$ and $t^* = 0$.

Case 1. If $t^* > 0$, then, using (5.35), we have $\theta(x^*) = 0$. By Proposition 5.4.2 we conclude that x^* is critical.

Case 2. Now we suppose that $t^* = 0$. First let us define the sequence $\{s^{i_k}\}$, where s^{i_k} is the solution of the optimization model introduced in (5.15) for $x = x^{i_k}$. From Corollary 5.4.4, this sequence is bounded, recall that $x^{i_k} \to x^*$. Without loss of generality let us suppose that $\{s^{i_k}\}$ converges to s^*, which is a solution of the problem in (5.15) for $x = x^*$ because $s(\cdot)$ is closed. Take $r \in \mathbb{N}$. As $t^* = 0$ there is a $k \in \mathbb{N}$ large enough such that $t_{i_k} < 2^{-r}$. This means the condition of Armijo rule is not satisfied for $t = 2^{-r}$, i.e., there is a $w^{i_k} \in G(x^k)$ such that

$$\langle w^{i_k}, F(x^{i_k} + 2^{-r}s^{i_k}) - F(x^{i_k})\rangle > \sigma 2^{-r}\theta(x^{i_k}).$$

Letting $k \to \infty$ in the above inequality, taking into account that $w^{i_k} \to w^* \in G(x^*)$, since w^{i_k} is a bounded sequence, $G(\cdot)$ is a closed map and $\theta(\cdot)$ is continuous, we obtain

$$\frac{\langle w^*, F(x^* + 2^{-r}s^*) - F(x^*)\rangle}{2^{-r}} \geq \sigma\theta(x^*).$$

Letting $r \to \infty$, we have

$$\langle w^*, JF(x^*)s^*\rangle \geq \sigma\theta(x^*).$$

As $F(\cdot)$ is a strong convex function, $\langle w, s^T D^2F(x^*)s\rangle > 0$ for all $w \in G(x^*)$, see Proposition 5.3.3. Due to $s \in \mathbb{R}^n$, then

$$\max_{w\in G(x^*)} \langle w, JF(x^*)s^*\rangle + \langle w, {s^*}^T D^2F(x^*)s^*\rangle \geq \langle w^*, JF(x^*)s^*\rangle \geq \sigma\theta(x^*).$$

Since s^* is solution of (5.15) and $0 < \sigma < 1$, then we conclude

$$\theta(x^*) \geq 0. \tag{5.36}$$

Therefore by Proposition 5.4.2 we have x^* is critical. ■

We end this section with the convergence of the method.

Theorem 5.5.4 *Suppose that there exists $f \in \mathbb{R}^m$ such that the set $\mathcal{F} = \{x \in \mathbb{R}^n : F(x) \in f - \mathcal{K}\}$ is compact, where $\mathcal{K}$ is a convex, pointed and closed cone which contains $\cup_{x \in U} K(x)$. If x^0 belongs to $\mathcal{F}$, then the sequence $\{x^k\}$ converges to an efficient point $x^* \in U$.*

Proof. Let $x^0 \in \mathcal{F}$. By Lemma 5.5.2, the sequence $\{F(x^k)\}$ is $\mathcal{K}$-decreasing. So, we have

$$x^k \in \mathcal{F}, \text{ for all } k.$$

By the compactness of $\mathcal{F}$, the accumulation points of $\{x^k\}$ belong to $\mathcal{F}$ and, using Proposition 5.5.3, we conclude that all accumulation points of $\{x^k\}$ are efficient. Moreover, since $\{F(x^k)\}$ is K-decreasing, all these accumulation points have the same value at F. As U is convex and F is strongly K-convex, by Proposition 5.3.4(iii) there exists just one accumulation point of $\{x^k\}$ as desired. ■

This section shows that Newton's method is well defined for optimization problems in which a strongly convex function is minimized with respect to a variable ordering structure. Moreover, the generated sequence converges to a solution. It is important to study the speed of this convergence. Next section is devoted to analyze under which condition super-linear or quadratic convergence can be ensured.

5.6 Convergence rate

In this section we analyze the convergence rate of the sequence $\{x^k\}$ generated by Newton's method for variable order vector optimization problems proposed in the previous section. As usual a characterization of $s(x^k)$ is needed. To this aim the first lemma provides a closed formula of the Newton direction which uses elements of the set of active vectors $w \in G(x)$ for $\varphi_w(x, v)$, where $x \in U$ and $v \in \mathbb{R}^n$. That is

$$W(x, v) := \{\bar{w} \in G(x);\ \varphi_{\bar{w}}(x, v) = \max_{w \in G(x)} \varphi_w(x, v)\}.$$

Lemma 5.6.1 *For any $x \in U$, there exists an integer $r := r(x)$ and*

$$w^i := w^i(x) \in W(x, s(x)), \quad \alpha_i := \alpha_i(x) \in [0, 1],\ i = 1, \dots, r,$$

such that

$$\sum_{j=1}^{m} \left(\tilde{w}_j^k \nabla F_j(x^k) + \tilde{w}_j^k \nabla^2 F_j(x^k) s(x^k)\right) = 0. \tag{5.37}$$

Hence,

$$s(x) = -\left[\sum_{i=1}^{r} \sum_{j=1}^{m} \alpha_i w_j^i \nabla^2 F_j(x)\right]^{-1} \left[\sum_{i=1}^{r} \sum_{j=1}^{m} \alpha_i w_j^i \nabla F_j(x)\right], \qquad \sum_{i=1}^{r} \alpha_i = 1.$$

Proof. Take $x \in U$ and let us define $g : \mathbb{R}^n \to \mathbb{R}$ as

$$g(v) := \max_{w \in G(x)} h(w, v),$$

where $h : G(x) \times \mathbb{R}^n \to \mathbb{R}$ is given by $h(w, v) := \varphi_w(x, v)$ and $\varphi_w(\cdot, \cdot)$ is defined in (5.14). Since $G(x)$ is compact, h is continuous, $h(w, \cdot)$ is a continuously differentiable convex function for all $w \in G(x)$ and $\nabla_v h(\cdot, v)$ is continuous on $G(x)$ for all $v \in \mathbb{R}^n$, using Danskin's Theorem, see [17], we conclude that $\partial g(v) = \operatorname{conv}\{\nabla_v h(w, v);\, w \in W(x, v)\}$. Now calculating the v-gradient of $h(w, \cdot)$ and evaluating in $v = s(x)$, we have

$$\nabla_s h(w,\, s(x)) = \nabla_s \left(\langle w,\, JF(x)s(x)\rangle + \frac{1}{2}\langle w,\, s(x)^T D^2 F(x) s(x)\rangle \right).$$

Making $w = (w_1, \ldots w_m)$, it follows

$$\nabla_s h(w,\, s(x)) = \nabla_s \left(\sum_{j=1}^{m} w_j \nabla F_j(x)^T s(x) + \frac{1}{2} s^T(x) w_j \nabla^2 F_j(x) s(x) \right),$$

which is equal to

$$\nabla_s h(w,\, s(x)) = \left(\sum_{j=1}^{m} w_j \nabla F_j(x) + w_j \nabla^2 F_j(x) s(x) \right).$$

Since $s(x)$ minimizes $g(\cdot)$, $0 \in \partial g(s(x))$. Hence there exists $r = r(x)$, $w^i = w^i(x) \in W(x; s(x))$ and $\alpha_i = \alpha_i(x) \in [0;1]$ $(1 \le i \le r)$ such that

$$0 = \sum_{i=1}^{r} \alpha_i \nabla_s h(w^i,\, s(x)) = \sum_{i=1}^{r} \sum_{j=1}^{m} \alpha_i \left(w_j^i \nabla F_j(x) + w_j^i \nabla^2 F_j(x) s(x) \right),$$

with $\sum_{i=1}^{r} \alpha_i = 1$. Then (5.37) holds.

Since F is strongly convex, by Proposition 5.3.3 $\sum_{m}^{j=1} w_j \nabla^2 F_j(x)$ is positive definite for all $w = (w_1, \ldots, w_m) \in G(x)$. Therefore the positive combination $\sum_{i=1}^{r} \sum_{j=1}^{m} \alpha_i w_j^i \nabla^2 F_j(x)$, is regular and the result follows directly from (5.37). ■

This result is now applied for obtaining an upper bound of the norm of this distance based on the value of $\theta(x)$.

Lemma 5.6.2 *For any $x \in \mathbb{R}^n$ it holds*

$$\frac{\rho(x)}{2} \|s(x)\|^2 \le -\theta(x).$$

Proof. Consider r, α^i, w^i, $i = 1, \ldots, r$ given as in Lemma 5.6.1. Then we have

$$\theta(x) = \sum_{i=1}^{r} \alpha^i \left(\langle w^i, JF(x)s(x)\rangle + \frac{1}{2}\langle w^i, s(x)^T D^2 F(x) s(x)\rangle \right) s(x)$$

or, equivalently:

$$\theta(x) = \sum_{i=1}^{r}\sum_{j=1}^{m} \alpha^i w_j^i \nabla F_j(x)^T s(x) + \frac{1}{2} s(x)^T \left(\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha^i w_j^i \nabla^2 F_j(x) \right) s(x).$$

As by Lemma 5.6.1, $\left[\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha_i w_j^i \nabla^2 F_j(x)\right] s(x) = -\left[\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha_i w_j^i \nabla F_j(x)\right]$, it is equivalent with

$$\theta(x) = -\left(\left[\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha_i w_j^i \nabla^2 F_j(x) \right] s(x) \right)^T s(x) + \frac{1}{2} s(x)^T \left(\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha^i w_j^i \nabla^2 F_j(x) \right) s(x).$$

That is

$$\theta(x) = -s(x)^T \left[\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha_i w_j^i \nabla^2 F_j(x) \right]^T s(x) + \frac{1}{2} s(x)^T \left(\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha^i w_j^i \nabla^2 F_j(x) \right) s(x).$$

Therefore, recalling that $\nabla^2 F_j(x),\ j = 1, \ldots, m$ are symmetric, the expression above becomes

$$\theta(x) = -\frac{1}{2} s(x)^T \left(\sum_{i=1}^{r}\sum_{j=1}^{m} \alpha^i w_j^i \nabla^2 F_j(x) \right) s(x) = -\frac{1}{2} \left\langle \bar{w}, s(x)^T D^2 F(x) s(x) \right\rangle,$$

where $\bar{w} = \sum_{i=1}^{r} \alpha_i w^i$. From Proposition 5.3.3, we have

$$\theta(x) \leq -\frac{\rho(x)}{2} \|s(x)\|^2,$$

which completes the proof. ■

After having a bound for $\|s(\cdot)\|$ depending on θ, last lemma shows how an interval where this value shall lay where its lower extreme depends on the evaluation of F at the iteration point.

Lemma 5.6.3 *For any $w \in \text{conv}G(x^{k+1})$, it follows,*

$$\theta(x^{k+1}) \geq -\frac{1}{2\rho(x^{k+1})} \| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \|^2. \tag{5.38}$$

Proof. Since $G(x^{k+1}) \subset \text{conv}G(x^{k+1})$, we obviously have

$$\max_{w \in convG(x^{k+1})} \varphi_w(x^{k+1}, s(x^{k+1})) \geq \max_{w \in G(x^{k+1})} \varphi_w(x^{k+1}, s(x^{k+1})).$$

The linearity of the function $\varphi_w(x^{k+1}, s(x^{k+1}))$ with respect to w leads to the equality.

From the definition of θ, for any $w \in \text{conv}G(x^{k+1})$

$$\theta(x^{k+1}) \geq \min_s \langle w, JF(x^{k+1})s\rangle + \frac{1}{2}\langle w, s^T D^2 F(x^{k+1})s\rangle.$$

Using Proposition 5.3.3 and the definition of $\rho(x^{k+1})$ given in Proposition 5.4.1, the above inequality becomes

$$\theta(x^{k+1}) \geq \min_s \langle w, JF(x^{k+1})s\rangle + \frac{1}{2}\rho(x^{k+1})\|s\|^2,$$

or equivalently

$$\theta(x^{k+1}) \geq \min_s \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1})s + \frac{1}{2}\rho(x^{k+1})\|s\|^2.$$

As the minimum is attained at $s(x) = -\frac{\sum_{j=1}^{m} w_j \nabla F_j(x^{k+1})}{\rho(x^{k+1})}$,

$$\theta(x^{k+1}) \geq -\frac{\left\|\sum_{j=1}^{m} w_j \nabla F_j(x^{k+1})\right\|^2}{2\rho(x^{k+1})},$$

as desired. ∎

Now we can present the superlinear convergence of the sequence $\{s(x^k)\}$.

Theorem 5.6.4 *Suppose that for some $f \in \mathbb{R}^m$, $x^0 \in \mathcal{F}$, where $\mathcal{F} := \{x \in \mathbb{R}^n : F(x) \in f - \mathcal{K}\}$ is a compact set. Assume that there exists $L > 0$ such that $\text{dist}(G(x^{k+1}), G(x^k)) \leq o(\|s(x^k)\|)$, for all k. Then $\{x^k\}$ converges to an efficient point x^*. Moreover, $t_k = 1$ for k large enough and the convergence of $\{x^k\}$ to x^* is superlinear.*

Proof. From Theorem 5.5.4, the whole sequence $\{x^k\}$ converges to the efficient point x^*. Since F is twice continuously differentiable, for any $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that

$$B(\bar{x}, \delta_\epsilon) \subset U, \quad \|D^2F(x) - D^2F(y)\| < \epsilon, \tag{5.39}$$

for all x, $y \in B(x^*, \delta_\epsilon)$.

From Proposition 5.4.2 and 5.4.5, we can assure that the sequence $\{s(x^k)\}$ converges to $s(x^*) = 0$. Therefore there exists $k_\epsilon > 0$, such that for $k \geq k_\epsilon$ we obtain $x^k + s(x^k) \in B(x^*, \delta_\epsilon)$. Using the Taylor's second order expansion at x^k for each coordinate function F_j, we have

$$F_j(x^k + s(x^k)) = F_j(x^k) + \nabla F_j(x^k)s(x^k) + \frac{1}{2}s(x^k)^T \nabla^2 F_j(\eta^k)s(x^k), \text{ for all } k \geq k_\epsilon,$$

where $\eta^k \in [x^k, x^k + s(x^k)]$. Taking $w \in G(x^k)$, multiplying the expression above for w_j and summing up in j we obtain

$$\sum_{j=1}^{m} w_j \left(F_j(x^k + s(x^k)) - F_j(x^k)\right) = \sum_{j=1}^{m} \left(w_j \nabla F_j(x^k)s(x^k) + \frac{1}{2}w_j s(x^k)^T \nabla^2 F_j(\eta^k)s(x^k)\right),$$

which becomes,

$$\langle w, F(x^{k+1})\rangle - \langle w, F(x^k)\rangle = \langle w, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle w, s(x^k)^T D^2 F(\eta^k)s(x^k)\rangle.$$

Using some algebraic manipulations and (5.39) we have

$$\langle w, F(x^{k+1})\rangle - \langle w, F(x^k)\rangle \leq \langle w, JF(x^k)s(x^k)\rangle + \frac{1}{2}\langle w, s(x^k)^T D^2 F(x^k)s(x^k) + \frac{\epsilon}{2}\|s(x^k)\|^2.$$

Taking the maximum in $w \in G(x^k)$ in the second part of the above inequality, we obtain

$$\langle w, F(x^{k+1})\rangle - \langle w, F(x^k)\rangle \leq \theta(x^k) + \frac{\epsilon}{2}\|s(x^k)\|^2.$$

Writing $\theta(x^k) = \sigma\theta(x^k) + (1-\sigma)\theta(x^k)$, by Lemma 5.6.2 we obtain

$$\langle w, F(x^{k+1})\rangle - \langle w, F(x^k)\rangle \leq \sigma\theta(x^k) + \big(\epsilon - (1-\sigma)\rho(x^k)\big)\,\frac{\|s(x^k)\|^2}{2}.$$

This means that if $\epsilon < (1-\sigma)\rho(x^k)$ then $t_k = 1$ fulfills the Armijo rule for $k \geq k_\epsilon$, see (5.33). Since the sequence $\{x^k\}$ converges, by Corollary 5.4.4, $\rho(x^k)$ is lower bounded.

From now on we will assume that $\epsilon < (1-\sigma)\rho(x^k)$. From Lemma 5.6.1, there exists $\tilde{w}^k \in \mathrm{conv}W(x^k, s(x^k))$, such that $\sum_{j=1}^m \big(\tilde{w}_j^k \nabla F_j(x^k) + \tilde{w}_j^k \nabla^2 F_j(x^k)s(x^k)\big) = 0$. For any $w \in \mathrm{conv}G(x^{k+1})$ we have

$$\left\|\sum_{j=1}^m w_j \nabla F_j(x^{k+1})\right\| = \left\|\sum_{j=1}^m w_j \nabla F_j(x^{k+1}) - \sum_{j=1}^m \big(\tilde{w}_j^k \nabla F_j(x^k) + \tilde{w}_j^k \nabla^2 F_j(x^k)s(x^k)\big)\right\|. \quad (5.40)$$

Let $\{w^1(k)\ldots, w^l(k)\} \subset W(x^k, s(x^k))$ such that $\sum_{i=1}^l \alpha_i w^i(k) = \tilde{w}^k$, where $\sum_{i=1}^l \alpha_i = 1$. But we assumed that $\mathrm{dist}(G(x^{k+1}), G(x^k)) \leq o(\|s(x^k)\|)$. Then, for each $w^i(k) \in W(x^k, s(x^k))$ there is $w^i(k+1) \in G(x^{k+1})$ such that

$$w^i(k+1) = w^i(k) + L\|s(x^k)\|\mu^i(k), \quad (5.41)$$

where $\mu^i(k) \to 0$. Making $w = \sum_{i=1}^l \alpha_i w^i(k+1)$ in (5.40) it follows

$$\begin{aligned}\left\|\sum_{j=1}^m w_j \nabla F_j(x^{k+1})\right\| = \Bigg\|\sum_{j=1}^m \sum_{i=1}^l \alpha_i w_j^i(k+1)\nabla F_j(x^{k+1}) - \sum_{j=1}^m \sum_{i=1}^l \alpha_i w_j^i(k) \nabla F_j(x^k) \\ - \sum_{j=1}^m \sum_{i=1}^l \alpha_i w_j^i(k) \nabla^2 F_j(x^k)s(x^k)\Bigg\|.\end{aligned}$$

Substituting (5.41) in the above expression we have

$$\begin{aligned}\left\|\textstyle\sum_{j=1}^m w_j \nabla F_j(x^{k+1})\right\| = \Big\|\textstyle\sum_{j=1}^m \sum_{i=1}^l \alpha_i w_j^i(k)(\nabla F_j(x^{k+1}) - \nabla F_j(x^k)) \\ +L\|s(x^k)\| \textstyle\sum_{j=1}^m \sum_{i=1}^l \alpha_i \mu_j^i(k) \nabla F_j(x^{k+1}) - \sum_{j=1}^m \sum_{i=1}^l \alpha_i w_j^i(k) \nabla^2 F_j(x^k)s(x^k)\Big\|.\end{aligned} \quad (5.42)$$

Using the expansion of Taylor, we get

$$\left\| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \right\| \leq \left\| \sum_{j=1}^{m} \sum_{i=1}^{l} \alpha_i w_j^i(k) H_j(x^k) s(x^k) + L\|s(x^k)\| \sum_{j=1}^{m} \sum_{i=1}^{l} \alpha_i \mu_j^i(k) \nabla F_j(x^{k+1}) \right\|,$$

where $H_j(x^k) = \int_0^1 \nabla^2 F_j(x^k + ts(x^k)) - \nabla^2 F_j(x^k))dt$. Furthermore, from (5.39) we obtain $\|H_j(x^k)\| < \epsilon$. Due to $\alpha_i \geq 0$, $\sum_{i=1}^{l} \alpha_i = 1$ and $w_j^i(k) \in G(x^k) \subset S(0,1)$, it follows

$$\left\| \sum_{j=1}^{m} \sum_{i=1}^{l} \alpha_i w_j^i(k) H_j(x^k) s(x^k) \right\| \leq m\epsilon \|s(x^k)\|. \tag{5.43}$$

On the other hand, from the convergence of the sequence $\{x^k\}$ we can guarantee the existence of $c > 0$ such that

$$\|\nabla F_j(x^{k+1})\| \leq c,$$

for all $j = \{1, \ldots, m\}$. Therefore, for all $\eta > 0$ there exists k_η such that

$$L\|s(x^k)\| \left\| \sum_{j=1}^{m} \sum_{i=1}^{l} \alpha_i \mu_j^i(k) \nabla F_j(x^{k+1}) \right\| \leq \eta \|s(x^k)\|, \text{ for all } k \geq k_\eta,$$

recall that $\alpha_i \geq 0$, $\sum_{i=1}^{l} \alpha_i = 1$ and for all $j = 1, 2, \ldots, m$, $\mu_j(k) \to 0$. Combining (5.42), (5.43) and the previous inequality, we obtain

$$\left\| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \right\| \leq (m\epsilon + \eta) \|s(x^k)\|.$$

Taking $m\epsilon + \eta < \varepsilon$, for $k \geq k_\varepsilon := \max\{k_\epsilon, k_\eta\}$, the last inequality becomes

$$\left\| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \right\| \leq \varepsilon \|s(x^k)\|.$$

Using the last inequality and Lemma 5.6.3, we obtain

$$-\theta(x^{k+1}) \leq \frac{1}{2\rho(x^{k+1})} \left\| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \right\|^2 \leq \frac{\varepsilon^2}{2\rho(x^{k+1})} \|s(x^k)\|^2.$$

But, from Lemma 5.6.2 we have

$$\frac{\rho(x^{k+1})}{2} \|s(x^{k+1})\|^2 \leq -\theta(x^{k+1}).$$

So, the last two relations lead to

$$\frac{\rho(x^{k+1})}{2} \|s(x^{k+1})\|^2 \leq \frac{\varepsilon^2}{2\rho^2(x^{k+1})} \|s(x^k)\|^2.$$

Then
$$\|x^{k+2}-x^{k+1}\| = \|s(x^{k+1})\| \leq \frac{\varepsilon}{\rho(x^{k+1})}\|s(x^k)\| = \frac{\varepsilon}{\rho(x^{k+1})}\|x^{k+1}-x^k\|.$$
Since $\{x^k\}$ is convergent, there is ρ^* such that $\rho_(x^k) > \rho^*$ for all k. Then for $k \geq k_\varepsilon$ and $j \geq 1$ we have
$$\|x^{k+j+1}-x^{k+j+1}\| \leq \left(\frac{\varepsilon}{\rho^*}\right)^j \|x^{k+1}-x^k\|. \tag{5.44}$$

To prove superlinear convergence, we use the same ideas of the proof of [38, Theorem 6.2]. We take $0 < \tau < 1$ and define
$$\bar{\epsilon} = \min\left\{1-\sigma,\ \frac{\tau}{1+2\tau}\right\}\rho^*.$$

If $\epsilon < \bar{\epsilon}$ and $k \geq k_\varepsilon$, using (5.44) and the convergence of the sequence $\{x^k\}$ to x^*, we have
$$\|x^{k+1}-x^*\| \leq \sum_{j=1}^{\infty}\|x^{k+j+1}-x^{k+j+1}\| \leq \sum_{j=1}^{\infty}\left(\frac{\tau}{1+2\tau}\right)^j\|x^{k+1}-x^k\| = \frac{\tau}{1+\tau}\|x^{k+1}-x^k\|.$$
Hence,
$$\|x^k-x^*\| \geq \|x^k-x^{k+1}\| - \|x^{k+1}-x^*\| \geq \frac{1}{1+\tau}\|x^k-x^{k+1}\|.$$
Combining the two above inequalities we conclude that, if $\epsilon < \bar{\epsilon}$ and $k \geq k_\varepsilon$ then
$$\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|} \leq \tau.$$
Recalling that τ is an arbitrary element of $(0,1)$, it can be taken as small as desired. So,
$$\frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|} \to 0,$$
and the super-linear convergence is obtained. ■

The super-linear convergence is the first step of the proof of the quadratic convergence of the method. The next theorem generalizes this result, already proven in [33, 38] for multi-objective and vector optimization problems, to the variable ordering case.

Theorem 5.6.5 *Suppose that for some $f \in \mathbb{R}^m$, $\mathcal{F} := \{x \in \mathbb{R}^n : F(x) \in f - \mathcal{K}\}$ is a compact set. Assume that $x^0 \in \mathcal{F}$, D^2F is Lipschitz continuous on U and that* $\mathrm{dist}(G(x), G(y)) \leq o(\|x-y\|^2)$. *Then the sequence $\{x^k\}$ converges quadratically to an efficient solution x^*.*

Proof. From the previous theorem, we have that the sequence $\{x^k\}$ converges superlinearly to an efficient point $x^* \in U$ and that $t_k = 1$ for k sufficiently large. Let $\mathcal{L}$ be the Lipschitz constant for D^2F. By Lemma 5.6.1, there exist $\tilde{w}^k \in \text{conv}W(x^k, s(x^k))$, $k = 1, \ldots, m$, such that

$$\sum_{j=1}^{m} \left(\tilde{w}_j^k \nabla F_j(x^k) + \tilde{w}_j^k \nabla^2 F_j(x^k) s(x^k) \right) = 0.$$

Hence, let $\{w^1(k) \ldots, w^l(k)\} \subset W(x^k, s(x^k))$ such that $\sum_{i=1}^{l} \alpha_i w^i(k) = \tilde{w}^k$, where $\sum_{i=1}^{l} \alpha_i = 1$. Then, due to we assumed that $\text{dist}(G(x^{k+1}), G(x^k)) \leq L\|s(x^k)\|^2$, for each $w^i(k) \in W(x^k, s(x^k))$ there exists $w^i(k+1) \in G(x^{k+1})$ such that

$$w^i(k+1) = w^i(k) + L\|s(x^k)\|^2 \mu^i(k),$$

where $\mu^i(k) \in B(0, 1)$. Using the Taylor expansion, after some algebraic manipulations we obtain

$$\left\| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \right\| \leq \left\| \sum_{j=1}^{m} \sum_{i=1}^{l} \alpha_i w_j^i(k) H_j(x^k) s(x^k) + L\|s(x^k)\|^2 \sum_{j=1}^{m} \sum_{i=1}^{l} \alpha_i \mu_j^i(k) \nabla F_j(x^{k+1}) \right\|, \tag{5.45}$$

where $H_j(x^k) = \int_0^1 \nabla^2 F_j(x^k + ts(x^k)) - \nabla^2 F_j(x^k))dt$. As the function D^2F is Lipschitz continuous

$$H_j(x^k) \leq \int_0^1 t\mathcal{L} dt \leq \mathcal{L}/2.$$

On the other hand, due to the sequence $\{x^k\}$ converges and ∇F_j is continuous for $j = 1, \ldots, m$, we can assume, as in the proof of Theorem 5.6.4, that $\|\nabla F_j(x^{k+1})\| \leq c$, for some $c > 0$. Then, from (5.45) we have

$$\left\| \sum_{j=1}^{m} w_j \nabla F_j(x^{k+1}) \right\| \leq \left(\frac{\mathcal{L}}{2} + cLm \right) \|s(x^k)\|^2.$$

Using the above inequality and Lemma 5.6.3, we obtain

$$|\theta(x^{k+1})| \leq \frac{1}{\rho(x^{k+1})} \left(\frac{\mathcal{L}}{2} + mcL \right)^2 \|s(x^k)\|^4,$$

which combined with Lemma 5.6.2 yields

$$\|s(x^{k+1})\| \leq \frac{2(\frac{\mathcal{L}}{2} + mcL)}{\rho(x^{k+1})} \|s(x^k)\|^2. \tag{5.46}$$

From now on the proof follows the same steps of the proof of [38, Theorem 6.3]. We include it here for sake of completeness.

Take $\tau \in (0, 1)$. Since the sequence $\{x^k\}$ converges superlinearly to x^*, there exists $\bar{k}$ such that for $k \geq \bar{k}$, (5.46) is satisfied and

$$\|x^* - x^{k+1}\| \leq \tau \|x^* - x^k\|.$$

Therefore, from the triangle inequality, for $\ell \leq \bar{k}$ we obtain

$$(1-\tau)\|x^* - x^\ell\| \leq \|x^\ell - x^{\ell+1}\|,$$

$$\|x^\ell - x^{\ell+1}\| \leq (1+\tau)\|x^* - x^\ell\|.$$

Taking $\ell = k+1$, in the first inequality and $\ell = k$ in the second, the following holds

$$(1-\tau)\|x^* - x^{k+1}\| \leq \|s(x^{k+1})\|. \tag{5.47}$$

$$\|s(x^k)\| \leq (1+\tau)\|x^* - x^k\| \tag{5.48}$$

Combining (5.46),(5.48),(5.47), we finally obtain

$$\frac{\|x^* - x^{k+1}\|}{\|x^* - x^k\|^2} \leq \frac{(1+\tau)^2}{(1-\tau)} \frac{2(\frac{\mathcal{L}}{2} + mcL)}{\rho(x^{k+1})},$$

which means that the sequence converges quadratically. ■

Chapter 6

Final remarks

In the Chapter 3, the main purpose has been to introduce and discuss the trust region method for multiobjective problems on the Riemannian context. Our method considered the trust rate function (3.5), which retrieves the trust rate function proposed in [70]. It is worth to mentioning that the trust rate function introduced in Liang el al. [57] and Carrizo et al. [21] can also be generalized to the Riemannian context and, as a consequence, new variants of the trust region method comes up. Furthermore, it is worth to investigate the relationship and performance of these methods from the numerical viewpoint. It is well known that the trust region method is one of the most important iterative methods for solving unconstrained nonlinear optimization problems, due its global convergence and superlinear convergence rate. In this work, we only study its global convergence properties of the method. In future work, we expect to prove the superlinear convergence of the trust region method in addition, extend other variations of trust region method for the multiobjective case for instance, constrained case and non-monotone case.

In relation to the Chapter 4, it is worth noting that the nonlinear scalar function, see (4.6), considered in the iterative step process of the algorithm, see (4.7), allows a relationship between the weak sharp minima set of the vector optimization problem and the weak sharp minima set of the scalarized problem. For state this relationship, we need some definitions and results. Let $G : M \to R^m$, $\eta \in \mathbb{R}^m$ and let us define the following level set

$$W_\eta := \{p \in M \ : \ G(p) = \eta\}.$$

We denote by Min$G$ (resp. WMin$G$) the *set of the efficient points* (resp. *weak efficient points*) associated to (4.1).

Definition 6.0.1 *A point $\hat{p} \in M$ is said to be weak sharp minimum to* (4.1)*, if there is a constant $\tau > 0$ such that*

$$G(p) - G(\hat{p}) \notin B(0, \tau d(p, W_{G(\hat{p})})) - C, \quad p \in M \backslash W_{G(\hat{p})}, \tag{6.1}$$

The set of all weak sharp minimum to (4.1) *is denoted by* WSMin_G.

The above definition has appeared in several contexts, see for example, [8, 13, 66, 74]. Note that the relationship (6.1) can be expressed in following equivalent form

$$d(G(p) - G(\hat{p}), -C) \geq \tau d(p, W_{G(\hat{p})}), \qquad p \in M,$$

and there holds $WSMin_G \subset \mathrm{Min}G$. In the particular case m=1 and $C = \mathbb{R}_+$, the last inequality becomes to the well-known inequality

$$G(p) - G(\hat{p}) \geq \tau d(p, W_{G(\hat{p})}), \qquad p \in M,$$

introduced in [46], defining weak sharp minimizer in Riemannian context.

Next result establishes the above mentioned relationship between WSMin$F$ and the weak sharp minimum associated to the nonlinear scalar function defined in (4.3), the proof follows by using similar arguments used in the proof of [74, Theorem 3.4].

Theorem 6.0.2 *Let $F : M \to R^m$ and $\hat{p} \in M$. Suppose that $W_{F(\hat{p})}$ is closed set and define $\tilde{F} : M \to \mathbb{R}^n$ by $\tilde{F}(p) = F(p) - F(\hat{p})$. If $\hat{p} \in \mathrm{WSMin}_F$ then $\hat{p} \in \mathrm{WSMin}_{f\circ\tilde{F}}$, where f is given by* (4.4).

We expect that the Theorem 6.0.2 constitutes a first step towards to establish the following result: "If $\hat{p} \in WSMin_F$, then $\{p^k\}$ converges, in a finite number of iterations". We foresee further progress along these line in the nearby future. Similar result has been proven in the Euclidean context; see [13].

At last, we have introduced extended Newton's method to variable ordering structures. To his aim the concept of strong convexity has been generalized and results involving first and second derivatives are shown. Although in the case of convex functions, stationarity implies weakly-efficiency in variable order optimization problems, we observed that stationary points may not be efficient if the objective function is strictly convex with respect to the variable ordering as it is in the case of vector optimization models. The variability of the order leads to examples where the difference of the values of the function lay on the boundary of the cone. Efficiency is only guaranteed if the function is strongly convex.

For the method, the Newton direction is defined using a quadratic approximation of the objective function and the dual cone at the point. Uniqueness, continuity, characterization of stationary points are satisfied under Lipschitz continuity of the distance of the generator of the cones and strong convexity of the objective function. The convergence of the algorithm and its rate of convergence are also obtained. We want to point out that in this approach computes efficient points if the function is strongly convex and other natural hypotheses on the ordering structure. This is an advantage with respect to other solutions methods which can only guarantee the calculation of weakly efficient point. In fact, as noted in [51, page 20], "Although weakly Pareto optimal solutions are important for theoretical considerations, they are not always useful in practice, because of the large size of the weakly Pareto optimal set." For an application where Pareto optimality arises as a central concept we mentioned, for example, "new welfare economics". In this case, a social state is Pareto optimal if no individual can be made better off without making at least one other individual worse off; for more details, see [24, page 220].

The hypothesis needed for the convergence of the approach are not too strong. Indeed, compared with the steepest descent method, proposed in [9], only one extra assumption on the objective function have to be asked. On the other hand, the hypotheses are clear generalizations of the suppositions need for the convergence of the Newton scheme in vector optimization models as reported in [38].

For the rate of convergence, if the distance of the generator of the cone is only Lipschitz continuous, the scalarizing vector providing the direction may vary on a linear range. So, not even super-linear convergence can be proven. For that reason, a stronger condition on the rate of variability of the cone is asked for obtaining a quadratic rate of convergence.

The computational implementation of the algorithm is a line of future research. The numerical behavior of the algorithm shall be tested at non-convex, convex and strongly convex functions. Using the ideas of [42] for generating convex functions, strongly convex functions can be obtained.

Bibliography

[1] Absil, P.A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. Found. Comput. Math. **7**(3), 303–330 (2007).

[2] Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ (2008).

[3] Baatar, D., Wiecek, M.M.: Advancing equitability in multiobjective programming. Comput. Math. Appl. **52**(1-2), 225–234 (2006).

[4] Bao, T.Q., Mordukhovich, B.S.: Necessary nondomination conditions in set and vector optimization with variable ordering structures. J. Optim. Theory Appl. **162**(2), 350–370 (2014).

[5] Bao, T.Q., Mordukhovich, B.S., Soubeyran, A.: Fixed points and variational principles with applications to capability theory of wellbeing via variational rationality. Set-Valued Var. Anal. **23**(2), 375–398 (2015).

[6] Bao, T.Q., Mordukhovich, B.S., Soubeyran, A.: Variational analysis in psychological modeling. J. Optim. Theory Appl. **164**(1), 290–315 (2015).

[7] Batista, E.E.A., Bento, G.d.C., Ferreira, O.P.: Enlargement of monotone vector fields and an inexact proximal point method for variational inequalities in Hadamard manifolds. J. Optim. Theory Appl. **170**(3), 916–931 (2016).

[8] Bednarczuk, E.: On weak sharp minima in vector optimization with applications to parametric problems. Control Cybernet. **36**(3), 563–570 (2007)

[9] Bello Cruz, J.Y., Bouza Allende, G.: A steepest descent-like method for variable order vector optimization problems. J. Optim. Theory Appl. **162**(2), 371–391 (2014).

[10] Bello Cruz, J.Y., Bouza Allende, G., Lucambio Pérez, L.R.: Subgradient algorithms for solving variable inequalities. Appl. Math. Comput. **247**, 1052–1063 (2014).

[11] Bento, G.C., Cruz Neto, J.X.: A subgradient method for multiobjective optimization on Riemannian manifolds. J. Optim. Theory Appl. **159**(1), 125–137 (2013).

[12] Bento, G.C., da Cruz Neto, J.X., Santos, P.S.M.: An inexact steepest descent method for multicriteria optimization on Riemannian manifolds. J. Optim. Theory Appl. **159**(1), 108–124 (2013).

[13] Bento, G.C., Cruz Neto, J.X., Soubeyran, A.: A proximal point-type method for multicriteria optimization. Set-Valued Var. Anal. **22**(3), 557–573 (2014).

[14] Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds. Nonlinear Anal. **73**(2), 564–572 (2010).

[15] Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. J. Optim. Theory Appl. **154**(1), 88–107 (2012).

[16] Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Proximal point method for a special class of nonconvex functions on Hadamard manifolds. Optimization **64**(2), 289–319 (2015).

[17] Bertsekas, D.P.: Convex analysis and optimization. Athena Scientific, Belmont, MA (2003).

[18] Bonnel, H., Iusem, A.N., Svaiter, B.F.: Proximal methods in vector optimization. SIAM J. Optim. **15**(4), 953–970 (electronic) (2005).

[19] Bonnel, H., Todjihoundé, L., Udrişte, C.: Semivectorial bilevel optimization on Riemannian manifolds. J. Optim. Theory Appl. **167**(2), 464–486 (2015).

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

[21] Carrizo, G.A., Lotito, P.A., Maciel, M.C.: Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem. Math. Program. **159**(1-2, Ser. A), 339–369 (2016).

[22] Ceng, L.C., Yao, J.C.: Approximate proximal methods in vector optimization. European J. Oper. Res. **183**(1), 1–19 (2007).

[23] Chen, G.y., Huang, X., Yang, X.: Vector optimization, *Lecture Notes in Economics and Mathematical Systems*, vol. 541. Springer-Verlag, Berlin (2005). Set-valued and variational analysis

[24] Cohon, J. L.: Multiobjective programming and planning. Mathematics in Science and Engineering, 140. New York-London, 1978.

[25] Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-region methods. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA (2000).

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

[27] Dennis Jr., J.E., Mei, H.H.W.: Two new unconstrained optimization algorithms which use function and gradient values. J. Optim. Theory Appl. **28**(4), 453–482 (1979).

[28] Eichfelder, G.: Optimal elements in vector optimization with a variable ordering structure. J. Optim. Theory Appl. **151**(2), 217–240 (2011).

[29] Eichfelder, G.: Variable ordering structures in vector optimization. Vector Optimization. Springer, Heidelberg (2014).

[30] Eichfelder, G., Ha, T.X.D.: Optimality conditions for vector optimization problems with variable ordering structures. Optimization **62**(5), 597–627 (2013).

[31] Engau, A.: Variable preference modeling with ideal-symmetric convex cones. J. Global Optim. **42**(2), 295–311 (2008).

[32] Ferreira, O.P., Oliveira, P.R.: Proximal point algorithm on Riemannian manifolds. Optimization **51**(2), 257–270 (2002).

[33] Fliege, J., Graña Drummond, L.M., Svaiter, B.F.: Newton's method for multiobjective optimization. SIAM J. Optim. **20**(2), 602–626 (2009).

[34] Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000).

[35] Fukuda, E.H., Graña Drummond, L.M.: Inexact projected gradient method for vector optimization. Comput. Optim. Appl. **54**(3), 473–493 (2013).

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

[37] Göpfert, A., Riahi, H., Tammer, C., Zălinescu, C.: Variational methods in partially ordered spaces. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 17. Springer-Verlag, New York (2003)

[38] Graña Drummond, L.M., Raupp, F.M.P., Svaiter, B.F.: A quadratically convergent Newton method for vector optimization. Optimization **63**(5), 661–677 (2014).

[39] Graña Drummond, L.M., Svaiter, B.F.: A steepest descent method for vector optimization. J. Comput. Appl. Math. **175**(2), 395–414 (2005).

[40] Grapiglia, G.N., Yuan, J., Yuan, Y.x.: On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization. Math. Program. **152**(1-2, Ser. A), 491–520 (2015).

[41] Grohs, P., Hosseini, S.: Nonsmooth trust region algorithms for locally Lipschitz functions on Riemannian manifolds. IMA J. Numer. Anal. **36**(3), 1167–1192 (2016).

[42] Hernandez, D.: Funciones k-convexas con Órdenes variables (2015). http://dspace.uh.cu:8080/xmlui/handle/123456789/2653

[43] Jahn, J.: Vector optimization. Springer-Verlag, Berlin (2004). Theory, applications, and extensions

[44] Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Quart. Appl. Math. **2**, 164–168 (1944)

[45] Li, C., López, G., Martín-Márquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. (2) **79**(3), 663–683 (2009).

[46] Li, C., Mordukhovich, B.S., Wang, J., Yao, J.C.: Weak sharp minima on Riemannian manifolds. SIAM J. Optim. **21**(4), 1523–1560 (2011).

[47] Li, X.B., Lin, Z.: Convergence for vector optimization problems with variable ordering structure. Optimization (2016).

[48] Luc, D.T.: Theory of vector optimization, *Lecture Notes in Economics and Mathematical Systems*, vol. 319. Springer-Verlag, Berlin (1989)

[49] Luenberger, D.G.: The gradient projection method along geodesics. Management Sci. **18**, 620–631 (1972)

[50] Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Française Informat. Recherche Opérationnelle **4**(Sér. R-3), 154–158 (1970)

[51] Miettinen, K.: Nonlinear multiobjective optimization. International Series in Operations Research & Management Science, 12. Kluwer Academic Publishers, Boston, MA, 1999.

[52] Moré, J.J.: The Levenberg-Marquardt algorithm: implementation and theory. In: Numerical analysis (Proc. 7th Biennial Conf., Univ. Dundee, Dundee, 1977), pp. 105–116. Lecture Notes in Math., Vol. 630. Springer, Berlin (1978)

[53] Nocedal, J., Wright, S.J.: Numerical optimization, second edn. Springer Series in Operations Research and Financial Engineering. Springer, New York (2006)

[54] Peressini, A.L.: Ordered topological vector spaces. Harper & Row, Publishers, New York-London (1967)

[55] Powell, M.J.D.: A new algorithm for unconstrained optimization. In: Nonlinear Programming (Proc. Sympos., Univ. of Wisconsin, Madison, Wis., 1970), pp. 31–65. Academic Press, New York (1970)

[56] Powell, M.J.D.: Convergence properties of a class of minimization algorithms. In: Nonlinear programming, 2 (Proc. Sympos. Special Interest Group on Math. Programming, Univ. Wisconsin, Madison, Wis., 1974), pp. 1–27. Academic Press, New York (1974)

[57] Qu, S., Goh, M., Liang, B.: Trust region methods for solving multiobjective optimisation. Optim. Methods Softw. **28**(4), 796–811 (2013).

[58] Rapcsák, T.: Smooth nonlinear optimization in $\mathbf{R}^n$, *Nonconvex Optimization and its Applications*, vol. 19. Kluwer Academic Publishers, Dordrecht (1997).

[59] Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optimization **14**(5), 877–898 (1976).

[60] Sakai, T.: Riemannian geometry, *Translations of Mathematical Monographs*, vol. 149. American Mathematical Society, Providence, RI (1996).

[61] Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Hamiltonian and gradient flows, algorithms and control, *Fields Inst. Commun.*, vol. 3, pp. 113–136. Amer. Math. Soc., Providence, RI (1994)

[62] Soleimani, B.: Characterization of approximate solutions of vector optimization problems with a variable order structure. J. Optim. Theory Appl. **162**(2), 605–632 (2014).

[63] Soleimani, B., Tammer, C.: Concepts for approximate solutions of vector optimization problems with variable order structures. Vietnam J. Math. **42**(4), 543–566 (2014)

[64] Sorensen, D.C.: Newton's method with a model trust region modification. SIAM J. Numer. Anal. **19**(2), 409–426 (1982).

[65] Sorensen, D.C.: Trust region methods for unconstrained minimization. In: Nonlinear optimization, 1981 (Cambridge, 1981), NATO Conf. Ser. II: Systems Sci., pp. 29–38. Academic Press, London (1982)

[66] Studniarski, M.: Weak sharp minima in multiobjective optimization. Control Cybernet. **36**(4), 925–937 (2007)

[67] Tang, G.j., Huang, N.j.: An inexact proximal point algorithm for maximal monotone vector fields on Hadamard manifolds. Oper. Res. Lett. **41**(6), 586–591 (2013).

[68] Udrişte, C.: Convex functions and optimization methods on Riemannian manifolds, *Mathematics and its Applications*, vol. 297. Kluwer Academic Publishers Group, Dordrecht (1994).

[69] Villacorta, K.D.V., Oliveira, P.R.: An interior proximal method in vector optimization. European J. Oper. Res. **214**(3), 485–492 (2011).

[70] Villacorta, K.D.V., Oliveira, P.R., Soubeyran, A.: A trust-region method for unconstrained multiobjective problems with applications in satisficing processes. J. Optim. Theory Appl. **160**(3), 865–889 (2014).

[71] Wang, J., Li, C., Lopez, G., Yao, J.C.: Convergence analysis of inexact proximal point algorithms on Hadamard manifolds. J. Global Optim. **61**(3), 553–573 (2015).

[72] Wang, X.M., Li, C., Yao, J.C.: Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures. J. Optim. Theory Appl. **164**(1), 202–217 (2015).

[73] Wiecek, M.M.: Advances in cone-based preference modeling for decision making with multiple criteria. Decis. Mak. Manuf. Serv. **1**(1-2), 153–173 (2007)

[74] Xu, S., Li, S.J.: Weak ψ-sharp minima in vector optimization problems. Fixed Point Theory Appl. pp. Art. ID 154,598, 10 (2010)

[75] Yu, P.L.: Cone convexity, cone extreme points, and nondominated solutions in decision problems with multiobjectives. J. Optimization Theory Appl. **14**, 319–377 (1974)

[76] Yuan, Y.x.: Recent advances in trust region algorithms. Math. Program. **151**(1, Ser. B), 249–281 (2015).