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

Flávio Pinto Vieira

# Computing inexact $\mathcal{K}$-steepest descent directions and a new line search procedure for Vector Optimization

Goiânia
2022

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

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

**E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG**

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

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

**1. Identificação do material bibliográfico**

[ ] Dissertação [ X ] Tese

**2. Nome completo do autor**

Flávio Pinto Vieira

**3. Título do trabalho**

Computing inexact K-steepest descent directions and a new line search procedure for Vector Optimization

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

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

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

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

---

Documento assinado eletronicamente por **Orizon Pereira Ferreira**, **Professora do Magistério Superior**, em 24/03/2022, às 14:35, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **FLÁVIO PINTO VIEIRA**, **Discente**, em 25/03/2022, às 11:16, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2783520** e o código CRC **B779F513**.

---

**Referência:** Processo nº 23070.009185/2022-11 SEI nº 2783520

Flávio Pinto Vieira

# Computing inexact $\mathcal{K}$-steepest descent directions and a new line search procedure for Vector Optimization

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. Orizon Pereira Ferreira

**Co-Orientador:** Prof. Dr. Luis Román Lucambio Pérez

Goiânia
2022

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

Vieira, Flávio Pinto
Computing inexact K-steepest descent directions and a new line search procedure for Vector Optimization [manuscrito] / Flávio Pinto Vieira. - 2022.
cxxxii, 132 f.: il.

Orientador: Prof. Dr. Orizon Pereira Ferreira; co-orientador Dr. Luis Román Lucambio Pérez.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2022.
Bibliografia. Apêndice.
Inclui símbolos, gráfico, tabelas, algoritmos.

1. Otimização vetorial. 2. pareto ótimo. 3. otimização irrestrita. 4. busca linear não monotona. I. Ferreira, Orizon Pereira, orient. II. Título.

CDU 517

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

**ATA DE DEFESA DE TESE**

Ata nº 02 da sessão de Defesa de Tese de **Flávio Pinto Vieira,** que confere o título de Doutor em Matemática, **na área de concentração de Otimização.**

Ao vigésimo quarto dia do mês de março do ano de dois mil e vinte e dois, a partir das dez horas, através de web-vídeo-conferência, realizou-se a sessão pública de Defesa de Tese intitulada **“Computing inexact K-steepest descent directions and a new line search procedure for Vector Optimization.”** Os trabalhos foram instalados pelo Orientador e presidente da banca, Professor Doutor **Orizon Pereira Ferreira - IME/UFG** com a participação dos demais membros da Banca Examinadora: Professor Doutor **Luis Román Lucambio Pérez - IME/UFG** Coorientador, Professor Doutor **Leandro da Fonseca Prudente - IME/UFG** membro titular interno, Professor Doutor **Max Leandro Nobre Gonçalves - IME/UFG** membro titular interno, Professora Doutora **Ellen Hidemi Fukuda - Graduate School of Informatics,** membro titular externo e o Professor Doutor **Alfredo Noel Iusem - IMPA** membro titular externo. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese, tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Orizon Pereira Ferreira - IME/UFG**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, Ao vigésimo quarto dia do mês de março do ano de dois mil e vinte e dois.

TÍTULO SUGERIDO PELA BANCA

**Computing inexact K-steepest descent directions and a new line search procedure for Vector Optimization**

---

Documento assinado eletronicamente por **Leandro Da Fonseca Prudente**, **Professor do Magistério Superior**, em 24/03/2022, às 12:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Orizon Pereira Ferreira**, **Professora do Magistério Superior**, em 24/03/2022, às 12:28, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Luis Román Lucambio Pérez**, **Professor do Magistério Superior**, em 24/03/2022, às 13:03, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Ellen Hidemi Fukuda**, **Usuário Externo**, em 24/03/2022, às 13:15, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Max Leandro Nobre Gonçalves**, **Professor do Magistério Superior**, em 24/03/2022, às 14:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Alfredo Noel Iusem**, **Usuário Externo**, em 25/03/2022, às 08:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2710780** e o código CRC **3EC93846**.

**Referência:** Processo nº 23070.009185/2022-11 SEI nº 2710780

Dedico este trabalho a minha família e a minha namorada.

# Agradecimentos

Agradeço acima de tudo a Deus por ter me dado o dom da vida e permitido com que eu chegasse até aqui. A Nossa mãe Maria Santíssima por ter guiado os meus passos e sempre ouvido minhas preces.

A meus pais Alziro Camilo Vieira e Jovely Pinto de Faria por terem me educado e sempre me ensina a importância do conhecimento. A minha namorada Natana Tuanny Pimenta dos Santos pelo amor, carinho e incentivo. Ao restante da minha família pelo exemplo de vida.

A meu coorientador Dr Luis Román Lucambio Pérez pela confiança paciência e dedicação em me orientar por todo o período em que desenvolvi este trabalho. Ao professor Dr Orizon Pereira Ferreira pelo apoio durante todo esse trabalho.

Aos professores da banca, Professor Doutor Orizon Pereira Ferreira, Professor Doutor Leandro da Fonseca Prudente, Professor Doutor Max Leandro Nobre Gonçalves, Professora Doutora Ellen Hidemi Fukuda, Professor Doutor Alfredo Noel Iusem pelas valiosas contribuições a este trabalho. Aos demais professores da pós-graduação do IME-UFG, e em especial aos do grupo de Otimização, Glaudston Bento, Jefferson Divino.

Aos grandes amigos que fiz no doutorado e que foram um pilar de sustentação para que eu conseguisse concluir esse curso, Ana Maria, Angie Tatiana, Elismar Dias, Fernando Santana, Fernando Coltinho, Márcia do Socorro, Jeferson Arley, Milton Javier, Raimundo Cavalcante, Yovani Adolfo.

Ao IF goiano pela licença para a conclusão desse curso. Aos meus colegas do IF goiano Campus Iporá pelo apoio.

"A matemática é o alfabeto no qual Deus escreveu o universo."

**Galileu Galilei**

# Resumo

Vieira, Flávio Pinto. **Computing inexact $\mathcal{K}$-steepest descent directions and a new line search procedure for Vector Optimization**. Goiânia, 2022. 135p. Tese de Doutorado Relatório de Graduação. Programa de Pós-Graduação em Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Neste trabalho, propomos uma nova busca linear para otimização vetorial e uma forma de calcular a direção $\sigma-$aproximada de máxima descida. Yunda Dong, em 2010 e 2012, introduziu um procedimento de busca linear para o método de Gradiente Conjugado usando apenas informações de primeira ordem, ou seja, sem utilizar valores funcionais. Estenderemos seus trabalhos para Otimização Vetorial. Estudaremos o método de gradiente conjugado, mostrando a convergência quando são utilizados os seguintes $\beta_k$'s: Fletcher-Reeves, conjugate descent, Dai-Yuan, Polak-Ribière-Polyak e Hestenes-Stiefel. Também usamos essa mesma busca linear para o método tipo-gradiente, mostrando sua convergência. Em 2004, Iusem e Graña Drummond introduziram o conceito de $\sigma$-aproximada $\mathcal{K}$-diereção de máxima descida. Eles mostraram que ao substituir a direção de Cauchy por essas direções, o resultado de convergência da sequência gerada é o mesmo: todo ponto de acumulação é crítico. Apresentaremos um procedimento eficiente para calcular essas direções quando o cone $\mathcal{K}$ for finitamente gerado.

**Palavras–chave**

Otimização vetorial, pareto ótimo, otimização irrestrita, busca linear não monotona, direção $\sigma$-aproximada.

# Abstract

Vieira, Flávio Pinto. **Computing inexact $K$-steepest descent directions and a new line search procedure for Vector Optimization**. Goiânia, 2022. 135p. PhD. Thesis Relatório de Graduação. Programa de Pós-Graduação em Matemática, Instituto de Matemática e Estatística, Universidade Federal de Goiás.

In this work, we proposes a new linear search and a way for the computation of $\sigma$-approximate direction. Yunda Dong, in 2010 and 2012, introduced a new linear search procedure for Conjugated Gradient methods using only first-order information, i.e., without working with functional values. We extend his works to Vector Optimization. We stud conjugate gradient methods, showing convergence when the following $\beta_k$'s are used: Fletcher-Reeves, conjugate descent, Dai-Yuan, Polak-Ribière-Polyak, and Hestenes-Stiefel. We also use this line search in the gradient method, showing its convergence. In 2004, Iusem and Graña Drummond introduced the concept of $\sigma$-approximate $\mathcal{K}$-steepest descent direction. They showed that by replacing the Cauchy direction with these directions, the convergence result of the generated sequence is the same: every accumulation point is critical. We will present an efficient procedure for computing these directions when the cone $\mathcal{K}$ is finitely generated.

**Keywords**

Vector optimization, Pareto-optimality, unconstrained optimization, nonmonotone line search, $\sigma$-approximate direction.

# Contents


Introduction **13**
0.1 Thesis outline 20

1 Preliminary **21**
1.1 Basic Results 21
1.2 Vector Optimization 22

2 Conjugate gradient methods with a new line search **30**
2.1 Convex case 30
2.2 Non-convex case 37
2.3 General Case 39
2.3.1 Convergence Analysis 40
2.4 Analysis of convergence for specific $\beta_k$'s 43
2.4.1 Fletcher-Reeves 45
2.4.2 Conjugate Descent 46
2.4.3 Dai-Yuan 47
2.4.4 Polak-Rivière-Polyak e Hestenes-Stiefel 50
2.5 Complexity 51
2.6 Computational experiments 53
2.6.1 Constants 54
2.6.2 Numerical Results 55

3 Steepest descent method with a new line search **73**
3.1 Modifying the new line search 73
3.2 Convex case 74
3.2.1 Rate of convergence 77
3.3 Lipschitz case 78
3.3.1 Rate of convergence 81
3.4 General case 83
3.5 Numerical experiments 85
3.5.1 Finding Pareto points 85
3.5.2 Building Pareto fronts 87
3.6 A new group of testing problems 89
3.6.1 Non-monotone algorithms 96
3.6.2 Four new problems 97
3.7 Chapter conclusion 103

4 How to compute inexact $\mathcal{K}$-steepest descent directions **104**
4.1 $\sigma$-Approximate $\mathcal{K}$-steepest descent direction 104
4.2 How to compute a $\sigma$-$\mathcal{K}$-descent direction 106
4.3 Computational experiments 109
4.3.1 An algorithm using an Armijo-type line-search along $\sigma$-approximate $K$–steepest descent direction 109
4.3.2 Computational experiments 111
4.3.3 First group of problems 114
4.3.4 Second group of problems 119
4.3.5 Third group of problems 122
4.3.6 Chapter conclusion 127

Final remarks **128**

Bibliography **129**

A Metric **134**
A.1 Purity metric 134
A.2 Spread metrics 134

# Introduction

Given $f\colon R^n \to R$, the (scalar or common) optimization problem, denoted by

$$\min f(x), \tag{0-1}$$

consists in the searching of $x^* \in R^n$ such that $f(x) \geq f(x^*)$ for all $x \in R^n$. A point like $x^*$ is called optimal solution of (0-1). Now, consider $F\colon R^n \to R^m$ and $f_i$, $i = 1, 2, \ldots, m$, its coordinate functions. Problem

$$\min_{R^n_+} F(x) \tag{0-2}$$

consists in the searching of $x^* \in R^n$ such that there is no $x \in R^n$ with $F(x) \neq F(x^*)$ and $f_i(x) \leq f_i(x^*)$ for all $i \in \{1, 2, \ldots, m\}$, i.e., if $f_j(x) < f_j(x^*)$, for some $j \in \{1, 2, \ldots, m\}$, then there exists $l \in \{1, 2, \ldots, m\}$ such that $f_l(x) > f_l(x^*)$. Problem (0-2) is called a multicriteria, multiobjective or Pareto optimization problem, and any point like $x^*$ is called a Pareto point. The multicriteria problem is a generalization of the scalar optimization problem. Indeed, by taking $m = 1$, in Problem (0-2) we have it reduces to Problem (0-1). Observe that

$$R^m_+ = \{y = (y_1, \ldots, y_m) \in R^m \mid y_1 \geq 0, \ldots, y_m \geq 0\}$$

is a convex, closed, pointed and non-empty cone, and the relations

$$u \preceq v \Leftrightarrow v - u \in R^m_+ \quad \text{and} \quad u \prec v \Leftrightarrow v - u \in int(R^m_+),$$

where

$$int(R^m_+) = \{y = (y_1, \ldots, y_m) \in R^m \mid y_1 > 0, \ldots, y_m > 0\},$$

define two partial orders in $R^m_+$. So, for convex, closed, pointed and non-empty cone $\mathcal{K} \subset R^m$, problem

$$\min_{\mathcal{K}} F(x) \tag{0-3}$$

is called a vector optimization problem and consists in the searching of $x^* \in R^n$ such that it does not exist $x \in R^n$ with $F(x) \neq F(x^*)$ and $F(x) \preceq_{\mathcal{K}} F(x^*)$, where

$\preceq_{\mathcal{K}}$ is the partial order defined by $\mathcal{K}$ in $R^m$ as

$$u \preceq_{\mathcal{K}} v \Leftrightarrow v - u \in \mathcal{K} \quad \text{and} \quad u \prec_{\mathcal{K}} v \quad if \quad v - u \in int(\mathcal{K}).$$

Points as $x^*$ is called $\mathcal{K} - Pareto\, minimizer$ or $Pareto\ \ efficient\ \ point$. The image of $x^*$ by $F$ is the $Pareto\ \ front$ in $\mathbb{R}^m$. Sometimes, it is useful to consider weakly efficient solutions of Problem (0-3). A point $\bar{x}$ is called *weakly efficient* when there is not $y$ such that $F(y) \prec_K F(x)$. See [20, 21, 27, 28, 34, 39, 40, 42].

The history of Optimization Problem (0-1) blends together with that of Mathematics itself and has developed over time. Euclid (300 bc) already considered the shortest distance between a point and a line and showed that a square have the largest area between the rectangles of the same perimeter. In the centuries 17 and 19, Newton and Gauss had already proposed iterative methods to move towards a minimum. In 1847 Cauchy presented the gradient method. The term "linear programming" was mentioned for some situations by George B. Dantzig, although the theory was introduced by Leonid Kantorovich in 1939. In 1947 Dantzig published the Simplex Algorithm and in the same year John von Neuman created the theory of duality. Over the course of time, several other methods were created and generalized to more general contexts.

Problem (0-2) emerged aiming to fill some gaps left by other sciences. Francis Y. Edgeworth (1845-1926) and Vilfredo Pareto (1848-1923) were the first to introduce the concept of "non-inferiority" in economics and since then multi-objective optimization has been developing and highlighting in areas such as engineering and design. In 1881 the concept of "optimum" for multicriteria economic decision making was defined, initially introduced at King's College (London) and later at Oxford, by Economics professor F.Y. Edgeworth. For them, "optimum" means a point such that in any direction that we take small steps, the objectives do not increase together, but one decreases while some others increase. In 1906 Pareto introduced his theory stating that "The optimum allocation of the resources of a society is not attained so long as it is possible to make at least one individual better off in his own estimation while keeping others as well off as before in their own estimation". And from there, the development of multi-objective methods in Applied Mathematics and Engineering has flourished, highlighting the contributions of Stadler 1979 and Steuer 1985.

According to [39], the vector optimization theory started with the studies of Edgeworth (1881) and Pareto (1906) about economic equilibrium and welfare theories and with mathematical backgrounds of ordered spaces of Cantor (1897) and Hausdorff (1906). Later also collaborated with the development of Problem (0-

3) the game theory of Borel (1921) and von Neumann (1926) and production theory of Koopmans (1951). But, only with the publication of Kuhn-Tucker's paper (1951) on the necessary and sufficient conditions for efficiency, and of Deubreu's paper (1954) on valuation equilibrium and Pareto optimum, that vector optimization had its deserved recognition.

Let $F$ of class $\mathcal{C}^1$, i.e., the first-order derivative of $F$ at $x$, the Jacobian of $F$ at $x$, denoted by $JF(x)$, is continuous. The image on $\mathbb{R}^m$ by $JF(x)$ is denoted by Image$(JF(x))$. A necessary condition for the $\mathcal{K}$-optimality of $x^*$ is

$$-\mathrm{int}(\mathcal{K}) \cap \mathrm{Image}(JF(x^*)) = \emptyset.$$

A point $x^*$ of $\mathbb{R}^n$ is called $\mathcal{K}$-critical for $F$ when it satisfies this condition. Therefore, if $x$ is not $\mathcal{K}$-critical, there exists $v \in \mathbb{R}^n$ such that $JF(x)v \in -\mathrm{int}(\mathcal{K})$.

In this thesis, we study two kinds of iterative methods for Vector Optimization: the class of steepest-descent methods and the class of conjugate gradient methods.

One of the oldest and simplest method for Problem 1 is the Cauchy Method or Gradient Method - see [5]. It consists of generating a sequence $\{x^k\}$ given by $x^k + \alpha_k d^k$ where $\alpha_k$ is the step length and $d^k$ is the direction. This procedure has been extended by Fliege and Svaiter for the multiobjective context in [21] and for a more general context of vector optimization by Graña-Drummond and Svaiter in [28].

In 1964 Fletcher and Reeves introduced in [18] the conjugate gradient methods for Problem (0-1). It consists of generating a sequence given by

$$x^{k+1} = x^k + \alpha_k d^k, \quad k = 0, 1, 2, \ldots,$$

where $\alpha_k > 0$ is the step length, and $d^k \in \mathbb{R}^n$ is the line search direction. The direction is defined by

$$d^k = \begin{cases} -\bigtriangledown f(x^k), & \text{if } k = 0, \\ -\bigtriangledown f(x^k) + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases}$$

where $\beta_k$ is a scalar algorithmic parameter. When $f$ is a nonquadratic function, the algorithm is known as *nonlinear conjugate gradient methods* and we have several formulas for $\beta_k$. Below we list some of them:

Fletcher-Reeves (FR) [18] : $\beta_k = \frac{\langle \bigtriangledown f(x^k), \bigtriangledown f(x^k)\rangle}{\langle \bigtriangledown f(x^{k-1}), \bigtriangledown f(x^{k-1})\rangle}$,

Conjugate descent (CD) [19]: $\beta_k = \frac{-\langle \bigtriangledown f(x^k), \bigtriangledown f(x^k)\rangle}{\langle d^{k-1}, \bigtriangledown f(x^{k-1})\rangle}$, Dai-Yuan (DY) [9]: $\beta_k = \frac{\langle \nabla f(x^k), \nabla f(x^k)\rangle}{\langle d^{k-1}, \nabla f(x^k) - \nabla f(x^{k-1})\rangle}$,

Polak-Ribière-Polyak (PRP) [46, 47]: $\beta_k = \frac{\langle \nabla f(x^k), \nabla f(x^k) - \nabla f(x^{k-1})\rangle}{\langle \nabla f(x^{k-1}), \nabla f(x^{k-1})\rangle}$,

Hestenes-Stiefel (HS) [30]: $\beta_k = \frac{\langle \nabla f(x^k), \nabla f(x^k) - \nabla f(x^{k-1})\rangle}{\langle d^{k-1}, \nabla f(x^k) - \nabla f(x^{k-1})\rangle}$.

The nonlinear conjugate gradient methods, as well as the respective beta's listed above, were extended to vector context by Lucambio and Prudente, see [40].

Throughout history, iterative methods have been developed to solve Problem (0-1), where in each iteration $x^k$ a line search $\alpha_k$ is made along one direction $d^k$. In addition to those already mentioned, second-order methods stand out for their rapid convergence, as Newton's method, which consists of generating a sequence $\{x^k\}$ with the direction given by $(-(\nabla^2 f(x^k))^{-1} \nabla f(x^k))$. We also remember the non-monotone line search methods, which produce a sequence for which not necessarily all objectives are decreasing. Between then, we must highlight the max-type and the average-type methods introduced in [29] e [54], respectively.

Such methods were generalized to the vector or multicriteria context as well as their respective directions and line searches. The direction of the Gradient (Steepest descent) has been extended together with the line search of the Armijo by Fliege and Svaiter in [21] and Drummond and Svaiter in [28]. The direction of the Newton in [20] by Fliege, Drummond and Svaiter. The direction of the conjugate gradient, the standard Wolfe conditions and strong Wolfe conditions have been extended in [40] by Lucambio and Prudente. In the same way, the max-type and the average-type non-monotone line searches were generalized in [17, 42, 49].

The object of study in this thesis is the Problem (0-3). We will present a practical way to calculate a $\sigma-$approximate $\mathcal{K}$-steepest descent direction, in addition to generalizing a new way to calculate step length using gradient information only, all this in the vector context.

In the first part of this work we will present a new way to calculate the step length $\alpha_k$. Introduced initially in 2010 by Yunda Dong, for the conjugate gradient method for the scalar problem - see [14]. In this paper, the author presented general convergence results and in 2012 the same author, in [15], presents specific convergence results for the Polak-Ribière method, see [46, 47]. We will extend this procedure for vector optimization problem. This new search is of great importance because works only with gradient information, so the method is quite efficient for functions that have a simpler gradient expression than itself.

The idea, for scalar problems, is the following. Assume that $d$ is descent direction for $F$ at $\bar{x}$, i.e., there is an interval of step-sizes, $(0,\tau)$, such that $0 < t < \tau$ implies that $F(\bar{x}+td) \leq F(\bar{x})$. Take $\rho > 0$ and compute $\nu > 0$, such that

$$\langle F'(\bar{x}+\rho d)\,,\, d\rangle \geq \ell(\rho) = -\frac{\nu\|d\|^2}{2}\rho + \delta\langle F'(\bar{x})\,,\, d\rangle,$$

where $\delta \in (0,1)$ is a given parameter. It is easy to find such $\nu$. Indeed, if $\langle F'(\bar{x}+\rho d)\,,\, d\rangle \geq \delta\langle F'(\bar{x})\,,\, d\rangle$ then, any $\nu \geq 0$ will fulfill the condition, otherwise

$$\nu = 2\frac{\langle \delta F'(\bar{x}) - F'(\bar{x}+\rho d)\,,\, d\rangle}{\rho\|d\|^2}. \tag{0-4}$$

Satisfies the required condition. Now, given $\omega \in (0,1)$, calculate

$$i = \min\{j \in \mathbb{N}\colon \langle F'(\bar{x}+\omega^j\rho d)\,,\, d\rangle < \ell(\omega^j\rho)\} \tag{0-5}$$

and set the step length $\alpha = \omega^i$. Observe that $\langle F'(\bar{x}+\omega^j d)\,,\, d\rangle \to \langle F'(\bar{x})\,,\, d\rangle$ and $\ell(\omega^j\rho) \to \ell(0) = \delta\langle F'(\bar{x})\,,\, d\rangle$ as $j \to \infty$ because $\omega \in (0,1)$ and $F'$ is continuous. Henceforth, since $\delta \in (0,1)$ and $\langle F'(\bar{x})\,,\, d\rangle < 0$, the existence of $i$ is assured. The following example shows that our line search procedure can be non-monotone.

**Example 0.1.** *Take $\rho = \omega = 0.9$, $\delta = 0.001$ and*

$$f(x) = -\frac{3}{10}x + \frac{1}{\pi^2}\sum_{m=1}^{70}\left[\frac{-3\sin(m\pi/5) + 5\sin(4m\pi/5)}{m^2}\right]\sin(m\pi x).$$

*At $x = 0$, we have $f'(0) = -1$. Then, $d = 1$ is descent direction for $f$ at $0$. By (0-4) $\nu = 4.5414$ and by (0-5) $i = 1$ because $f'(0,81) < \ell(0,81)$. Therefore, the new iterate would be $0,81$ and $f(0,81) = 0.081220 > f(0) = 0$.*

In the multi-objective setting, define

$$f(\bar{x}+\alpha d, d) = \max\left\{\langle F_r'(\bar{x}+\alpha d)\,,\, d\rangle \colon r = 1,2,\ldots,m\right\},$$

and the line search to consider is the following: compute $\nu > 0$, such that

$$f(\bar{x}+\rho d\,,\, d) \geq \ell(\rho) = -\frac{\nu\|d\|^2}{2}\rho + \delta f(\bar{x}\,,\, d),$$

where $\rho > 0$ is given. Then, using a fixed $\omega \in (0,1)$, compute

$$i = \min\{j \in \mathbb{N}\colon f(\bar{x}+\omega^j\rho d\,,\, d) < \ell(\omega^j\rho)\}.$$

Finally, set the step length $\alpha = \omega^i$. As we will see, function $f$ is continuous. Then,

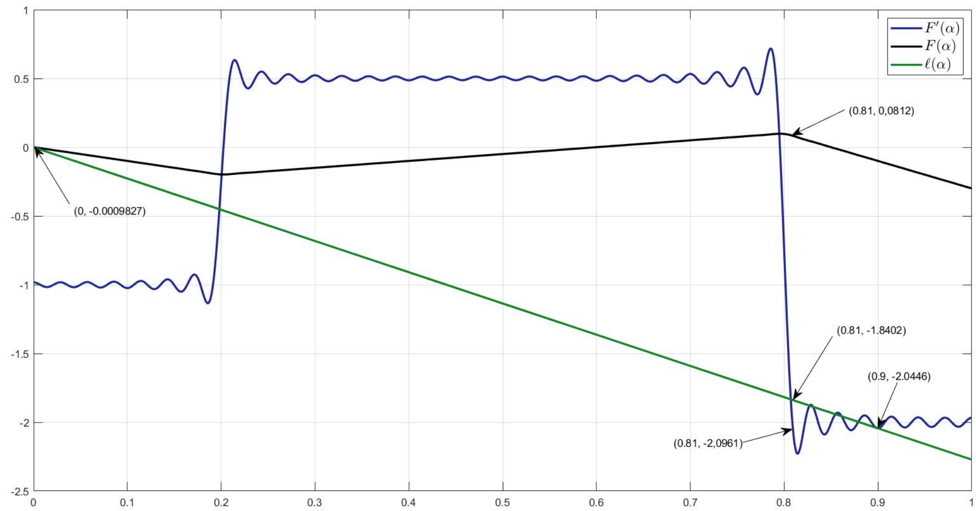



**Figure 1:** *The line search, case scalar.*

all the indicated computations are possible. The following example illustrates the behavior of our procedure.

**Example 0.2.** *As in Example 1, take $\rho = \omega = 0.9$, $\delta = 0.001$. Consider the bi-criteria optimization problem where the objective function $F\colon \mathbb{R} \to \mathbb{R}^2$ is defined as follows:*

$$F(x) = \begin{pmatrix} -\frac{3}{10}x + \frac{1}{\pi^2}\sum_{m=1}^{70}\left[\frac{-3\sin(m\pi/5)+5\sin(4m\pi/5)}{m^2}\right]\sin(m\pi x) \\ \\ \frac{-25}{3}x^3 + 10x^2 - 3x \end{pmatrix}.$$

*At $x = 0$, we have $JF(0) = (-0.9827\ \ -3)^T$, implying that $d = 1$ is descent direction for both objective. Observe that $f(0,1) = -0.9827$ and $f(0.81,1) = -2.0961$. Then $\nu = 4.5414$. Therefore, $f(0.81,1) < \ell(0.81)$ and the new iterate would be $x = 0.81$. Function-values $F(0) = (0\quad 0)^T$ and $F(0.81) = (0.0812\quad -0.2976)^T$ are non-comparable according to the order defined by the Pareto cone.*

We use this new line search in the vector context for the Conjugate Gradient method where we reproduce the same results presented by Lucambio and Prudente in [40]. In addition, we will replace the Armijo-type search in the steepest-descent algorithm with the new line search that we are proposing. We show some convergence results and present some numerical experiments testing the effectiveness of the resultant algorithm.

In the multi-objective scenario, the steepest descent direction $d$ is calculated

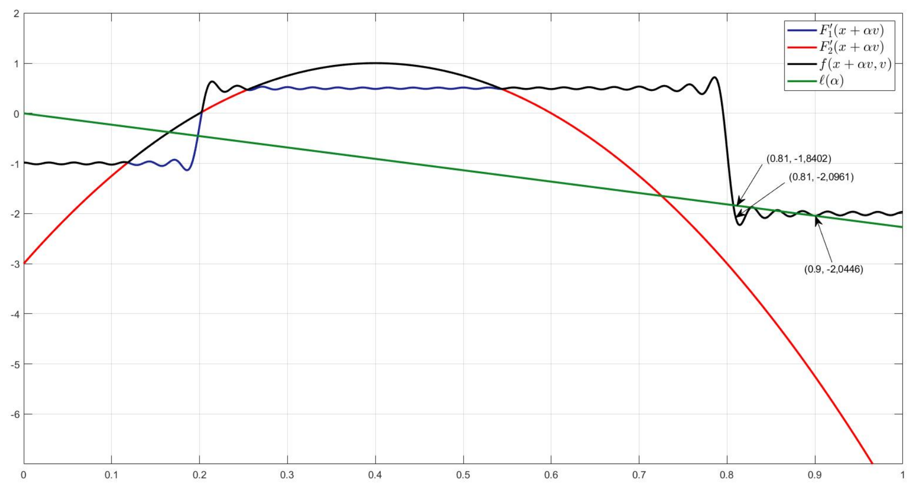



**Figure 2:** *The line search, case bi-criterion.*

solving the problem - see [21]

$$\begin{array}{ll} Minimize & f(x,d) + \frac{1}{2}\|d\|^2. \\ subject\ to & d \in \mathbb{R}^n \end{array} \tag{0-6}$$

Fliege and Svaiter, in this same article, suggested an approximate way to calculate the direction: $d$ is a approximate solution of (0-6) with tolerance $\sigma \in (0,1]$ if $f(x,d) + \frac{1}{2}\|d\|^2 \leq \sigma\theta(x)$, where $f(x,v) = max_i(JF(x)d)_i$ and $\theta(x)$ is the optimal value of Problem (0-6). Drummond and Iusem presented the same definition in [27] and showed that every accumulation point of the sequence generated by the projected gradient method is a stationary point, and when $F$ is convex the generated sequence converges to a weakly efficient solution, in addition to presenting results of limitation of direction $\sigma-$approximate. For Drummond and Svaiter, [28], $d$ is a $\sigma$-approximate $\mathcal{K}$-steepest descent direction at $x \in \mathbb{R}^n$ if $f(x,d) + \frac{1}{2}\|d\|^2 \leq (1-\sigma)\theta(x)$ with $\sigma \in [0.1)$, now in the vector context. Moreover, they introduced a succinct theory that served as the basis for us to present a practical way of calculating an approximate direction. Fukuda and Drummond in [24] presented an inexact projected Gradient Method for Vector Optimization Problems, in such work defining a direction $\sigma-$approximated in the same way as in [28], they also presented some properties for direction, obtaining convergence results similar to those in [27].

In the second part of this, we will present a practical way to calculate a $\sigma$-approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$, for vector Optimization problem. Some numerical experiments were carried out to verify the efficiency of this way of calculating direction.

## 0.1 Thesis outline

This work is basically divided into two parts, the first one introduces a manner to calculate the step length using only informations of first order, whereas the second part introduces a practical form of calculating a $\sigma-$approximate $\mathcal{K}-$steepest descent direction for $F$ at $x$.

In Chapter 1 are present some concepts and results which will used throughout this work.

In Chapter 2 we present a new way to calculate the step length using gradient information only. This is composed of six sections where the first three are dedicated to demonstrating the convergence of the method using the convexity hypothesis of the function, Jacobian $JF$ is Lipschitz continuos and for the general case, respectively. In the fourth section we discuss the convergence analysis for the specific $\beta$'s of Fletcher-Reeves, Conjugate Descent, Dai-Yuan, Polak-Rivière-Polyak and Hestenes-Stiefel. In the fifth section we present some results exploring the idea of complexity and finally in the last section we describe the numerical experiments.

In Chapter 3 we present a Gradient-type Algorithm with the new line search. Initially we present the search and then three algorithms, one using the convexity of the function, another that use the value of the Lipschitz constant. And finally, last one is for the general case, i.e., the objective are not convex and the Lipschitz constant is unknown. The last section of this chapter is dedicated to present numerical experiences with our algorithm for the convex case and for the general case.

In Chapter 4 we present a practical way to calculate a $\sigma-$approximate $\mathcal{K}-$steepest descent direction. In the first section, we present some results that will be used later. In the second section we explain how to calculate the direction $\sigma-$approximate. We show convergence results of a Gradient-type algorithm using the $\sigma-$approximate direction. Several experiments showing efficiency and robustness of our propose were performed and are presented at the closing section of the chapter.

CHAPTER 1

# Preliminary

In this chapter presents definitions and results known in the literature that will be used later as well as present the notation used throughout this work.

## 1.1 Basic Results

We denoted by $\mathbb{R}^n$ the n-dimensional Euclidean space. A vector $x \in \mathbb{R}^n$ has coordinates $x^1, \ldots, x^n$, that is $x = (x^1, \ldots, x^n)^T$, it will be considered as a column vector. Now, $\langle \cdot, \cdot \rangle$ stands for the usual inner product in $\mathbb{R}^n$ and $\| \cdot \|$ is the Euclidean norm.

Henceforth, we will follow with the definitions of convex sets and functions. The following results can be found in [6].

**Definition 1.1.** *A subset $\mathcal{C}$ of $\mathbb{R}^n$ is called convex if*

$$\alpha x + (1-\alpha) y \in \mathcal{C}, \qquad \forall \;\; x, y \in C, \;\; \forall \; \alpha \in [0,1].$$

We will denoted an arbitrary set in $\mathbb{R}^n$ by $X$ and if it is convex by $\mathcal{C}$.

**Definition 1.2.** *Let $\mathcal{C}$ be a convex subset of $\mathbb{R}^n$. A function $g : \mathcal{C} \to \mathbb{R}$ is called convex if*

$$g(\alpha x + (1-\alpha) y) \leq \alpha g(x) + (1-\alpha) g(y).$$

We will denote by $conv(\mathcal{X})$ the *convex hull* of a set $\mathcal{X}$, which is the intersection of all convex sets containing $\mathcal{X}$, and it is a convex set. The *conic hull* of a set $X$ denoted by $cone(X)$ is the intersection of all convex cone containing $X$.

Next, we will present the definition of the Lipschitz continuous gradient, which will help us in further results.

**Definition 1.3.** *([5]) A condition of the form*

$$\| \bigtriangledown g(x) - \bigtriangledown g(y) \| \leq L \| x - y \|, \qquad \forall \quad x, y \in \mathbb{R}^n,$$

*is called a Lipschitz continuity condition on $\bigtriangledown g$, with constant $L > 0$.*

Now, we introduce the well-known concept of quasi-Fejér convergent in Euclidean spaces. This definition was initially presented in [16], and it can be found in a more elaborate way in [7, 33].

**Definition 1.4.** *A sequence $\{y^k\} \in \mathbb{R}^n$ is said to be quasi-Fejér convergent to $V \subset \mathbb{R}^n$, $V \neq \emptyset$, if for each $v \in V$ there exists a sequence of real numbers $\varepsilon_k \geq 0$ such that $\sum_{k=0}^{\infty} \varepsilon_k < \infty$ and*

$$\| y^{k+1} - v \|^2 \leq \| y^k - v \|^2 + \varepsilon_k.$$

**Theorem 1.5.** *([33], Theorem 4.1) If $\{y^k\}$ is quasi-Fejér convergent to a nonempty set $V$, then $\{y^k\}$ is bounded. If $V$ contains a limit point of $\{y^k\}$ then $\{y^k\}$ converges.*

We close this section with the result extracted from [40], which will contribute to the results of Chapter 3.

**Lemma 1.6.** *For any scalar $a$, $b$, and $\alpha \neq 0$, we have*

*(a)* $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$
*(b)* $2ab \leq 2\alpha^2 a^2 + \frac{b^2}{2\alpha^2}$
*(c)* $(a+b)^2 \leq 2a^2 + 2b^2$
*(d)* $(a+b)^2 \leq (1+2\alpha^2)a^2 + (1+1/2\alpha^2)b^2$

*Proof.* See [40, Lemma 2.3] □

## 1.2 Vector Optimization

In this section we will present results and formalize the notation about vector optimization problem. Further discussion on the subject can be found in [34, 39]. Let $\mathcal{K} \subset \mathbb{R}^m$ be a closed, convex and pointed cone with non-empty interior. The pointed cone means that ($\mathcal{K} \cap (-\mathcal{K}) = \{0\}$), where $\mathcal{K}$ defines a partial order in $\mathbb{R}^m$, $\preceq_{\mathcal{K}}$. We say that $u \preceq_{\mathcal{K}} v$ if, and only if, $v - u \in \mathcal{K}$. Other partial order relations in $\mathbb{R}^m$ are defined by $\mathcal{K}$ analogously, for example, $u \prec_{\mathcal{K}} v$ says that $v - u$ lays in the interior of $\mathcal{K}$, ($v - u \in int(\mathcal{K})$).

The positive polar cone of $\mathcal{K}$ is the set $\mathcal{K}^* = \{w \in \mathbb{R}^m \mid \langle w, y\rangle \geq 0, \quad \forall\, y \in \mathcal{K}\}$. Since $\mathcal{K}$ is closed and convex, $\mathcal{K} = \mathcal{K}^{**}$, $-\mathcal{K} = \{y \in \mathbb{R}^m \mid \langle y, w\rangle \leq 0, \quad \forall\, w \in \mathcal{K}^*\}$ and $-int(\mathcal{K}) = \{y \in \mathbb{R}^m \mid \langle y, w\rangle < 0, \quad \forall\, w \in \mathcal{K}^* - \{0\}\}$. Let $G \subset \mathcal{K}^* - \{0\}$ be compact set such that

$$\mathcal{K}^* = cone(conv(G)),$$

i.e., $\mathcal{K}^*$ is the conic hull of the convex hull of $G$. For a generic $\mathcal{K}$, the set

$$G = \{w \in \mathcal{K}^* \mid \|w\| = 1\},$$

has the desired property. In classical optimization $\mathcal{K} = \mathbb{R}_+$, then $\mathcal{K}^* = \mathbb{R}_+$ and we can take $G = \{1\}$. For multiobjective optimization $\mathcal{K} = \mathbb{R}^m_+$, then $\mathcal{K}^* = \mathcal{K}$ and we may take $G$ as the canonical basis of $\mathbb{R}^m$. If $\mathcal{K}$ is a polyhedral cone, $\mathcal{K}^*$ is also polyhedral and $G$ can be taken as the finite set of extremal rays of $\mathcal{K}^*$.

Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$, the first-order derivative of $F$ at $x$, the Jacobian of $F$ at $x$, will be denoted by $JF(x)$, and the image on $\mathbb{R}^m$ by $JF(x)$ will be denoted by $Im(JF(x))$. $F$ is of the classes $\mathcal{C}^1$ if, the Jacobian of $F$ is continuous.

Function $F$ is called $\mathcal{K}$-convex when

$$F(\alpha x + (1-\alpha)y) \preceq_{\mathcal{K}} \alpha F(x) + (1-\alpha)F(y)$$

for all $\alpha \in [0,1]$. In the multi-criteria setting we have that $\mathbb{R}^m_+$-convexity of $F$ is equivalent to the convexity of all coordinate functions of $F$. When $F$ is convex and continuously differentiable,

$$F(y) \succeq_{\mathcal{K}} F(x) + JF(x)(y-x),$$

which is an extension of the classical gradient inequality for convex differential functions. When $F$ is $\mathcal{K}$-convex and $m > 1$, optimal solution's set may be non-convex, unlike what occurs in the scalar optimization case.

**Definition 1.7.** *A point $x^* \in \mathbb{R}^n$ is an efficient or $\mathcal{K}-$optimal solution for problem (0-3) if does not exist $x \in \mathbb{R}^n$ with $F(x) \neq F(x^*)$ and $F(x) \preceq_{\mathcal{K}} F(x^*)$. The images of $x^*$ by $F$ are called Pareto front in $\mathbb{R}^m$.*

When $\mathcal{K}$ is the Pareto cone, i.e., $\mathcal{K} = \mathbb{R}^m_+$, (0-3) is known a multicriteria optimization problem. In this case, optimal solutions are known as Pareto points. Observe that, if $x$ is Pareto, then it is impossible to improve one objective without another becoming worse.

Let us geometrically illustrate the definitions above. We want to minimize the functions $f_1(x) = x^2$ and $f_2(x) = (x-a)^2$, where $a > 0$, simultaneously, that is, to minimize the bi-function $F(x) = (f_1(x), f_2(x))$ considering the partial order induced by $\mathbb{R}^2_+$. As we can see in figure 1.1, the minimum of $f_1(x)$ is $x = 0$ whereas the minimum of $f_2(x)$ is $x = a$. Observe that for values of $x$ greater than "$a$" the two objectives increase, if we move to the right, so any values belonging to this interval would not be optimal or efficient solutions, the same is true for values of $x$ less than "0". On the other hand, for values of $x$ between "0" and "$a$", as we move from left to right, the values of function $f_1$ increases while those of $f_2$ decrease, the opposite occurring when we walk from right to left. Therefore, any value of $x$ between "0" and "$a$" is an efficient or optimal point.

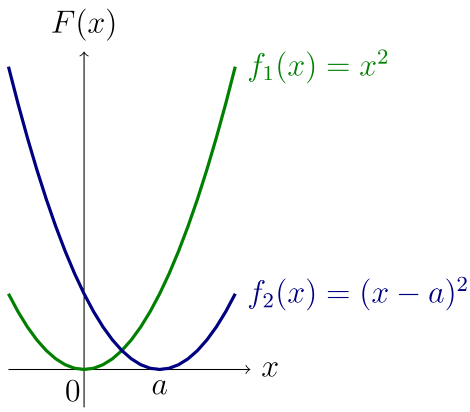



**Figure 1.1:** *Bi-criterion function.*

The graphic of the image of $F(x)$, Figure 1.2, shows that the functional values for $x$ greater than $a$ or less than 0 do not belong to the Pareto front, because in these intervals, the functional values increase or decrease simultaneously.

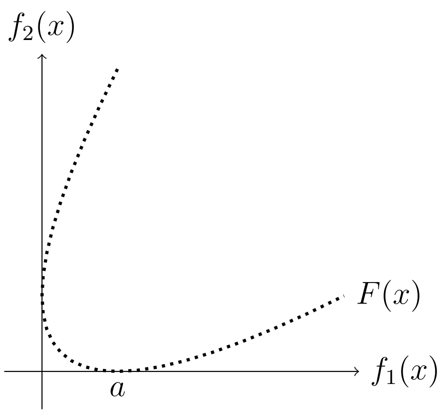



**Figure 1.2:** *Image of $F(x)$.*

Figure 1.3 illustrates the Pareto front of $F(x)$. Note that if we choose a value of $F(x)$ that does not belong in this range, there will always be a point belonging that is better in at least one of the objectives, and it is not worse on the other.

We need to present optimality conditions for Problem (0-3). So, a necessary condition for $\mathcal{K}$-optimality of $x^*$ is

$$-int(\mathcal{K}) \cap Image(JF(x^*)) = \emptyset.$$

A point $x^*$ of $\mathbb{R}^n$ is called $\mathcal{K}$-critical for $F$ when it satisfies this condition. Therefore, if $x$ is not $\mathcal{K}$-critical, there exists $v \in \mathbb{R}^n$ such that $JF(x)v \in -int(\mathcal{K})$. Every such

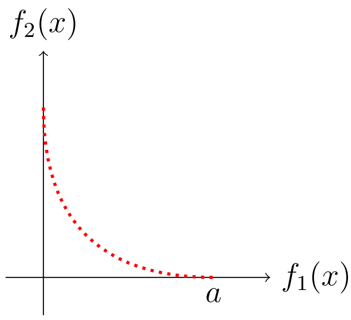



**Figure 1.3:** *Pareto front.*

vector $v$ is a $\mathcal{K}$-descent direction for $F$ at $x$, i.e., there exists $T > 0$ such that $0 < t < T$ implies that $F(x + tv) \prec_{\mathcal{K}} F(x)$, see [27].

The following definitions and results can be found in [28]. Define the function

$$\begin{aligned} \varphi\colon \mathbb{R}^m &\to \mathbb{R} \\ y &\longmapsto \varphi(y) = \sup\{\langle y, w\rangle \mid w \in G\}. \end{aligned}$$

In view of the compactness of $G$, $\varphi$ is well-defined. Function $\varphi$ has some useful properties.

**Lemma 1.8.** *Let $y$ and $y' \in \mathbb{R}^m$. Then:*

*(a) $\varphi(y + y') \le \varphi(y) + \varphi(y')$ and $\varphi(y) - \varphi(y') \le \varphi(y - y')$;*
*(b) If $y \preceq_{\mathcal{K}} y'$, then $\varphi(y) \le \varphi(y')$; if $y \prec_{\mathcal{K}} y'$, then $\varphi(y) < \varphi(y')$;*
*(c) $\varphi$ is Lipschitz continuous with constant 1.*

*Proof.* See [28, Lemma 3.1]. □

Function $\varphi$ gives characterizations of $-\mathcal{K}$ and $-int(\mathcal{K})$:

$$-\mathcal{K} = \{y \in \mathbb{R}^m \mid \varphi(y) \le 0\} \quad \text{and} \quad -int(\mathcal{K}) = \{y \in \mathbb{R}^m \mid \varphi(y) < 0\}.$$

Note that $\varphi(x) > 0$ does not imply that $x \in \mathcal{K}$, but $x \in \mathcal{K}$ implies that $\varphi(x) \ge 0$ and $x \in int(\mathcal{K})$ implies that $\varphi(x) > 0$.

Now define the function $f\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$f(x, d) = \varphi(JF(x)d) = \sup\{\langle JF(x)d, w\rangle \mid w \in G\}.$$

The function $f$ gives a characterization of $\mathcal{K}$-descent directions and of $\mathcal{K}$-critical points:

**Lemma 1.9.** *Let $x \in \mathbb{R}^n$, then*

*(a) $d$ is $\mathcal{K}$-descent direction for $F$ at $x$ if $f(x, d) < 0$;*

*(b) $x$ is $\mathcal{K}$-critical if and only if $f(x,d) \geq 0$ for all $d$.*

*Proof.* See [28]. □

The next function allows us to extend the notion of steepest descent direction to the vector case.

Define $v\colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$v(x) = \arg\min \left\{ f(x,d) + \frac{\|d\|^2}{2} \mid d \in \mathbb{R}^n \right\}, \tag{1-1}$$

and $\theta\colon \mathbb{R}^n \to \mathbb{R}$ by $\theta(x) = f(x, v(x)) + \|v(x)\|^2/2$. Since $f(x,\cdot)$ is a real closed convex function, $v(x)$ exists and is unique. Observe that in the scalar minimization case, where $F\colon \mathbb{R}^n \to \mathbb{R}$ and $\mathcal{K} = \mathbb{R}_+$, taking $G = \{1\}$, we obtain $f(x,d) = \langle \nabla F(x), d\rangle$, $v(x) = -\nabla F(x)$ and $\theta(x) = -\|\nabla F(x)\|^2/2$. The following lemma shows that $v(x)$ can be considered the vector extension of the steepest descent direction of the scalar case.

**Lemma 1.10.** *(a) If $x$ is $\mathcal{K}$-critical, then $v(x) = 0$ and $\theta(x) = 0$.*

*(b) If $x$ is not $\mathcal{K}$-critical $f(x, v(x)) \leq -\frac{\|v(x)\|^2}{2} < 0$, and $v(x)$ is a $\mathcal{K}$-descent direction for $F$ at $x$.*

*(c) The mappings $v$ and $\theta$ are continuous.*

*Proof.* See [28, Lemma 3.3]. □

Since $f(x,.)$ is positive homogeneous, it is easy to verify that

$$f(x, v(x)) = -\|v(x)\|^2. \tag{1-2}$$

For multiobjective optimization, where $\mathcal{K} = \mathbb{R}^m_+$, with $G$ given by the canonical basis of $\mathbb{R}^m$, $v(x)$ can be computed by solving

$$\begin{array}{ll} Minimize & \alpha + \frac{1}{2}\|d\|^2 \\ \text{subject to} & [JF(x)d]_i \leq \alpha, \quad i = 1, \dots, m. \end{array} \tag{1-3}$$

see [21]. The following result displays the Lipschitz constant for $f(x,d)$ and later we will show that this same function is monotone non-decreasing.

**Lemma 1.11.** *Function $f(x,\cdot)$ is continuous. If $F$ is differentiable, then $f$ is continuous. If $L$ is the Lipschitz constant of $JF$, then $L\|d\|$ is the Lipschitz constant of $f(\cdot\,, d)$.*

*Proof.* Immediate consequence of Lemma 1.8(c). □

Recall that $F\colon \mathbb{R}^n \to \mathbb{R}^m$ is called $\mathcal{K}$-convex when

$$F((1-\lambda)x + \lambda y) \preceq_{\mathcal{K}} (1-\lambda)F(x) + \lambda F(y)$$

for all $x$, $y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$ -see [34, 39]. In other words, $\lambda \in [0,1]$ implies that

$$F(x) + \lambda[F(y) - F(x)] - F(x + \lambda(y-x)) \in \mathcal{K}. \tag{1-4}$$

**Lemma 1.12.** *If $F \in \mathcal{C}^1$ is convex then $JF(x)$ is a subgradient of $F$ at $x$.*

*Proof.* Since $F$ is continuously differentiable, by (1-4) we get

$$\lim_{\lambda \to 0^+} \frac{F(x) - F(x + \lambda(y-x))}{\lambda} + F(y) - F(x) \in \mathcal{K}.$$

Hence,

$$F(y) \succeq_{\mathcal{K}} F(x) + JF(x)(y-x).$$

□

Above results can be fond at [34], Theorem 2.20. The last expression is the $\mathcal{K}$-vector version of the well known Jensen Inequality. Hence, analogous to the scalar case, $JF$ is $\mathcal{K}$-monotone operator, i.e.,

$$[JF(x) - JF(y)](x-y) \succeq_{\mathcal{K}} 0.$$

In the following lemma, we summarize two important facts, which we will use later on.

Observe that

$$f(x,d) = max\{\langle JF(x)d, w\rangle | w \in G\} = max\{\langle JF(x)d, w\rangle | w \in conv(G)\}.$$

Then,

$$\theta(x) = min\left\{ max\left\{\langle JF(x)d, w\rangle | w \in conv(G)\right\} + \frac{\|d\|^2}{2} |\ d \in \mathbb{R}^n \right\}.$$

The dual of this problem above is

$$max\left\{ mim\left\{\langle JF(x)d, w\rangle |\ d \in \mathbb{R}^n\right\} + \frac{\|d\|^2}{2}\ |w \in conv(G) \right\}$$

$$max\left\{ mim\left\{ \langle JF(x)d, w\rangle + \frac{\|d\|^2}{2} |\ d \in \mathbb{R}^n \right\} |w \in conv(G) \right\}$$

$$max\left\{-\frac{\|JF(x)^{\top}w\|^2}{2}|w \in conv(G)\right\}.$$

So, we achieve the following results.

**Theorem 1.13.** *The following two statements are true*

*(a) For all* $w \in conv(G)$*,*

$$\theta(x) \geq -\frac{\|JF(x)^{\top}w\|^2}{2}.$$

*(b) If* $\bar{w} \in \arg\min\left\{\|JF(x)^{\top}w\| \; |w \in conv(G)\right\}$ *then,* $v(x) = JF(x)^{\top}w$ *and* $\theta(x) = -\frac{\|JF(x)^{\top}w\|^2}{2}$.

*Proof.* Item (a) is true becuase the Weak Duality Theorem. Item (b) is true because the Strong Duality Theorem since there is not duality's gap between

$$\min\left\{f(x,d)+\frac{\|d\|^2}{2} \mid d\in\mathbb{R}^n\right\} \quad and \quad \max\left\{-\frac{\|JF(x)^{\top}w\|^2}{2}|w \in conv(G)\right\}.$$

□

**Definition 1.14.** *We say that* $\mathcal{D} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ *is a subgradient of* $F$ *at* $x$ *if*

$$F(y) \succeq_{\mathcal{K}} F(x) + \mathcal{D}(y-x), \qquad for\ any\ y \in \mathbb{R}^n.$$

**Lemma 1.15.** *Let* $F \colon \mathbb{R}^n \to \mathbb{R}^m$ *be convex and continuously differentiable.*

*(a) If* $F(y) \preceq_{\mathcal{K}} F(x)$*, then*

$$\langle v(x)\,,\, x-y\rangle \leq 0.$$

*(b) Fix* $x$ *and* $d \in \mathbb{R}^n$*. Function* $\gamma\colon \mathbb{R} \to \mathbb{R}$*, defined as*

$$\gamma(t) = f(x+td, d),$$

*is monotone non-decreasing.*

*Proof.* By Definition 1.14 and Lemma 1.12, we have that

$$0 \succeq_{\mathcal{K}} JF(x)(y-x)$$

Take $\bar{w} \in conv(G)$ such that $v(x) = -JF(x)^{\top}\bar{w}$. Theorem 1.13 says that such $\bar{w}$ exists. By Caratheodory's Theorem - see Proposition B.6 in [5]- there are $m+1$ elements of $G$ $w_0, w_1, \dots, w_m$, and $m+1$ non-negative scalars, $\lambda_0, \lambda_1, \dots, \lambda_m$, such that $\bar{w} = \sum_{i=0}^{m} \lambda_i w_i$ and $\sum_{i=0}^{m} \lambda_i = 1$. Then

$$\langle v(x), y-x\rangle = \langle -JF(x)^\top \bar{w}, y-x\rangle = -\sum_{i=0}^{m} \lambda_i \langle JF(x)^\top w_i, y-x\rangle \tag{1-5}$$

$$= -\sum_{i=0}^{m} \lambda_i \langle JF(x)^\top (y-x), w_i\rangle \geq 0. \tag{1-6}$$

To show item (b) take $t \in \mathbb{R}$ and $w \in G$ such that $f(x+td, d) = \langle w\,,\, JF(x+td)d\rangle$. Then, $\hat{t} > t$ implies

$$f(x+td, d) = \langle w\,,\, JF(x+td)d\rangle \leq \langle w\,,\, JF(x+\hat{t}d)d\rangle \leq f(x+\hat{t}d, d)$$

where in the first inequality we use that $JF$ is $\mathcal{K}$-monotone, and in the second one, we use the definition of $f$. □

CHAPTER 2

# Conjugate gradient methods with a new line search

In this chapter we will generalize to the vector context the line search introduced in 2010 by Yunda Dong for the scalar conjugate gradient method, see [14, 15]. It line search is characterizes by does not use of functional values, so, it excels in problems where evaluating its gradients is simpler than the function itself. This does not make use of functional values, so it has great advantages for problems where evaluating its gradients is more simpler than the function itself. By extending the Nonlinear conjugate gradient methods for vector optimization, in [40], Lucambio and Prudente used as line search the standard Wolfe and strong Wolfe conditions. We will rewrite the algorithm presented in [40] with the new line search proposed here for vector optimization, we will call this method Nonlinear Conjugate Gradient Methods with New Line search. We will show the well definition of the search and convergence results, initially using the convexity hypotheses of the functional values and after the Lipschitz gradient, and ultimately in the general case. We show too convergence results for the specific beta's of Fletcher-Reeves, conjugate descent, Dai-Yuan, Polak-Ribière-Polyak and Hestenes-Stiefel. Numerical experiments will also be presented showing the efficiency of the new search in the conjugate gradient algorithms to solve vector problems.

## 2.1 Convex case

We start this chapter by introducing the Nonlinear conjugate gradient methods for vector optimization with a new line search for cases in which the objective $F$ is convex. Initially two basic hypotheses are needed.

Under the hypothesis of $\mathcal{K}$-convexity of $F$, we have the algorithm.

**Algorithm 2.1.** *Given constants $\rho > 0$, $\omega$, $\delta \in (0,1)$, $\nu > 1$ and $e \in Int(\mathcal{K})$ such that $0 < \langle e, w\rangle$, for all $w \in G$. Conjugate gradient algorithm for convex case is defined as follows.*

**0. Initialization**: *Take $x^0 \in \mathbb{R}^n$. Compute $v(x^0)$ and initialize $k \leftarrow 0$.*

**1. Stopping criterium**: *If $v(x^k) = 0$, then STOP.*

**2. Direction**: *Define*

$$d^k = \begin{cases} v(x^k), & \text{if } k = 0 \\ v(x^k) + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{2-1}$$

*where $\beta_k$ is an algorithmic parameter.*

**3. Line search**: *Compute*

$$j_k = \min \left\{ j \geq 1 \;\middle|\; f(x^k + \rho\omega d^k, d^k) + \frac{\nu^j \rho\omega \|d^k\|^2}{2} \geq \delta f(x^k, d^k) \right\}, \tag{2-2}$$

*and*

$$i_k = \min \left\{ i \geq 1 \middle| \begin{array}{lcl} f(x^k + \rho\omega^i d^k, d^k) + \frac{\nu_k \rho\omega^i \|d^k\|^2}{2} & \leq & \delta f(x^k, d^k) \\ JF(x^k + \rho\omega^i d^k)d^k + \frac{\nu_k \rho\omega^i \|d^k\|^2}{2} e & \preceq_{\mathcal{K}} & \delta JF(x^k) d^k \end{array} \right\}. \tag{2-3}$$

**4. Iteration step**: *Define*

$$\alpha_k = \rho\omega^{i_k}, \tag{2-4}$$

*and*

$$x^{k+1} = x^k + \alpha_k d^k. \tag{2-5}$$

*Compute $v(x^{k+1})$, set $k \leftarrow k+1$, and go to Step 1.*

The choice of updating $\beta_k$ remains deliberately open. In later sections we will consider several choices of $\beta_k$ that result in globally convergent methods.

The following Lemma assures us that if there is a $\hat{t}$ such that

$$JF(x^k + \hat{t}d^k)d^k + \frac{\nu_k \hat{t}\|d^k\|^2}{2} e \preceq_{\mathcal{K}} \delta JF(x^k)d^k \quad \text{for all} \quad k = 0, 1, \ldots,$$

then the above inequality holds for all $t \in [0, \hat{t}]$.

**Lemma 2.2.** *Assume that $F$ is $\mathcal{K}$-convex, $e \in Int(\mathcal{K})$ and there exists $\hat{t} > 0$ such that*

$$JF(x^k + \hat{t}d^k)d^k + \frac{\nu_k \hat{t}\|d^k\|^2}{2} e \preceq_{\mathcal{K}} \delta JF(x^k)d^k \quad \text{for all} \quad k = 0, 1, \ldots. \tag{2-6}$$

*Then, $t \in [0, \hat{t}]$ implies*

$$JF(x^k + td^k)d^k + \frac{\nu_k t\|d^k\|^2}{2} e \preceq_{\mathcal{K}} \delta JF(x^k)d^k \quad \text{for all} \quad k = 0, 1, \ldots. \tag{2-7}$$

*Proof.* Take $t \in [0, \hat{t}]$. Since $JF$ is $\mathcal{K}$-monotone and $\frac{\nu_k(\hat{t}-t)\|d^k\|^2}{2}e \succeq_{\mathcal{K}} 0$, we get

$$[JF(x^k + \hat{t}d^k) - JF(x^k + td^k)]d^k + \frac{\nu_k(\hat{t}-t)\|d^k\|^2}{2}e \succeq_{\mathcal{K}} 0.$$

Then,

$$JF(x^k + td^k)d^k + \frac{\nu_k t\|d^k\|^2}{2}e \preceq_{\mathcal{K}} JF(x^k + \hat{t}d^k)d^k + \frac{\nu_k \hat{t}\|d^k\|^2}{2}e \preceq_{\mathcal{K}} \delta JF(x^k)d^k.$$

□

Since $\rho > 0$, $\omega \in (0, 1)$ and $\nu > 1$, $j_k$ will be the smallest positive integer that fulfills (2-2). If $i_k$ is computable, then Algorithm 2.1 is well-defined. Remember that $f$ is continuous with respect to its first argument because $F$ is continuously differentiable - see Lemma 1.11. Next, we prove that $i_k$ is computable.

**Lemma 2.3.** *For any $k \geq 0$ there exists integer $i_k \geq 1$ fulfilling (2-3).*

*Proof.* For $i = 1$

$$f(x^k + \rho\omega d^k, d^k) + \frac{\nu_k \rho\omega\|d^k\|^2}{2} \geq \delta f(x^k, d^k)$$

by the definition of $\nu_k$. Assume that

$$f(x^k + \rho\omega^i d^k, d^k) + \frac{\nu_k \rho\omega^i\|d^k\|^2}{2} > \delta f(x^k, d^k),$$

for all positive integer $i$. Then, taking limits as $i$ goes to $\infty$, we obtain

$$f(x^k, d^k) \geq \delta f(x^k, d^k)$$

because $f(\cdot, d^k)$ is continuous and $\omega \in (0, 1)$. That is a contradiction because $f(x^k, d^k) < 0$ and $\delta \in (0, 1)$. □

The following result is of extreme importance in the demonstration of several others.

**Lemma 2.4.** *For $k = 0, 1, \ldots$ it holds*

$$f(x^k, d^{k-1}) < \delta f(x^{k-1}, d^{k-1}). \tag{2-8}$$

*Proof.* Observe that, by (2-5) and (2-3),

$$\begin{aligned} f(x^k, d^{k-1}) &= f(x^{k-1} + \alpha_{k-1} d^{k-1}, d^{k-1}) \\ &\leq \delta f(x^{k-1}, d^{k-1}) - \nu_{k-1} \rho \omega^{i_{k-1}} \frac{\|d^{k-1}\|^2}{2} \\ &< \delta f(x^{k-1}, d^{k-1}). \end{aligned}$$

□

Let us now show that $d^k$ is descent direction for $F$ at $x^k$.

**Lemma 2.5.** *For $k = 0, 1, \ldots$, $\beta_k \geq 0$ and Take any $c \in (0, 1)$, it holds*

*(a)* $f(x^k, d^k) \leq -\dfrac{\|v(x^k)\|^2}{2}$,

*(b)* $f(x^k, d^k) \leq c f(x^k, v(x^k))$.

*Proof.* (a) This proof is given by induction. For $k = 0$, we have $d^0 = v(x^0)$, then, by Lemma 1.10(b), $f(x^0, d^0) < -\|v(x^0)\|^2/2$. For some $k \geq 1$, assume that $f(x^{k-1}, d^{k-1}) \leq -\|v(x^{k-1})\|^2/2 < 0$. Then, using definitions of $d^k$ and $f$, Lemma 1.8(a), non-negativeness of $\beta_k$, (2-8) and Lemma 1.10(b), we get

$$\begin{aligned} f(x^k, d^k) = f(x^k, v(x^k) + \beta_k d^{k-1}) &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \\ &\leq \frac{-\|v(x^k)\|^2}{2} + \beta_k \delta f(x^{k-1}, d^{k-1}) \\ &< \frac{-\|v(x^k)\|^2}{2} < 0 \end{aligned}$$

because, by assumption, $f(x^{k-1}, d^{k-1}) \leq -\|v(x^{k-1})\|^2/2 < 0$.

(b) When $k = 0$, $d^0 = v(x^0)$. Then, $f(x^0, d^0) = f(x^0, v(x^0)) < c f(x^0, v(x^0))$ , by Lemma 1.10 (b). For $k \geq 1$, we have

$$\begin{aligned} f(x^k, d^k) &= f(x^k, v(x^k) + \beta_k d^{k-1}) \\ &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \end{aligned}$$

by definition of $d^k$ and positiviness of $\beta_k$. By (2-8) and Lemma 1.9, $f(x^k, d^{k-1}) < 0$. Lemma 1.10 (b) states that $f(x^k, v(x^k)) \leq 0$. Then,

$$f(x^k, d^k) \leq f(x^k, v(x^k)) < c f(x^k, v(x^k)).$$

□

Actually a stronger result holds: $d^k$ satisfies the *sufficient descent condition*, that is, $f(x^k, d^k) \leq c f(x^k, v(x^k))$ for all $k = 0, 1, \ldots$ and $c \in (0, 1)$.

Next, we prove that if $F$ is convex, then Algorithm 2.1 generates a monotone $\mathcal{K}$-decreasing sequence $\{F(x^k)\}$.

**Lemma 2.6.** *Assume that $F$ is convex. Then,*

$$F(x^{k+1}) \preceq_{\mathcal{K}} F(x^k) + \delta\alpha_k f(x^k, d^k)e, \tag{2-9}$$

*for all $k \geq 0$.*

*Proof.* Observe that, for $k = 0, 1, 2, \ldots,$

$$F(x^{k+1}) = F(x^k) + \int_0^{\alpha_k} JF(x_k + td^k)d^k dt$$

Then, by (2-3) and Lemma 2.2

$$\begin{aligned} F(x^{k+1}) &\preceq_{\mathcal{K}} F(x^k) + \int_0^{\alpha_k} \left(\delta JF(x^k)d^k - \frac{1}{2}\nu_k t\|d^k\|^2 e\right) dt \\ &= F(x^k) + \alpha_k\delta JF(x_k)d^k - \frac{1}{4}\nu_k\alpha_k^2\|d^k\|^2 e \\ &\preceq_{\mathcal{K}} F(x^k) + \alpha_k\delta JF(x_k)d^k. \end{aligned} \tag{2-10}$$

For all $w \in G$, it is true that

$$\begin{aligned} \langle \omega, f(x^k, d^k)e - JF(x^k)d^k \rangle &= \langle \omega, f(x^k, d^k)e \rangle - \langle \omega, JF(x^k)d^k \rangle \\ &= f(x^k, d^k)\langle \omega, e \rangle - \langle \omega, JF(x^k)d^k \rangle \\ &\geq f(x^k, d^k) - \langle \omega, JF(x^k)d^k \rangle \end{aligned}$$

because $0 < \langle e, w \rangle < 1$. By the definition of $f$,

$$f(x^k, d^k) - \langle \omega, JF(x^k)d^k \rangle \geq 0.$$

Then,

$$f(x^k, d^k)e \succeq_{\mathcal{K}} JF(x^k)d^k.$$

Hence, by (2-10), we have

$$F(x^{k+1}) \preceq_{\mathcal{K}} F(x^k) + \alpha_k\delta f(x^k, d^k)e.$$

□

**Corollary 2.7.** *If $F$ is convex, then $\{F(x^k)\}_{k\geq 0}$ is strictly monotone $\mathcal{K}$-decreasing, i.e., $F(x^{k+1}) \prec_{\mathcal{K}} F(x^k)$, $k = 0, 1, \ldots$.*

*Proof.* Immediately. □

**Assumptions:**

**A1.** Function $F$ is bounded below on $\mathcal{L} = \{x \in \mathbb{R}^n \colon F(x) \preceq_{\mathcal{K}} F(x^0)\}$, i.e., for any sequence $\{y^k\} \subset \mathcal{L}$ with $F(y^k) \succeq_{\mathcal{K}} F(y^{k+1})$, for all $k$, there exists $\mathcal{F} \in \mathbb{R}^m$ such that $F(y^k) \succeq_{\mathcal{K}} \mathcal{F}$, for all $k$. There exists an open $\mathcal{N}$ such that $\mathcal{L} \subset \mathcal{N}$ and $JF$ is Lipschitz-continuous with constant $L$ on $\mathcal{N}$.

**A2.** Sequence $\{\nu_k\}$, generated by Algorithm A, is bounded, i.e., there exists $\bar{\nu}$ such that $0 < \nu_k < \bar{\nu}$ for all $k = 0, 1, \ldots$.

The following results are the basis for proving the convergence theorem.

**Lemma 2.8.** *If* **A1** *holds and* $F$ *is convex, then* $\sum_{k\geq 0} \alpha_k f(x^k, d^k)$ *is convergent.*

*Proof.* Observe that under these hypothesis and by (2-9) we get that there exists $\mathcal{F}$ such that

$$F(x^0) - \mathcal{F} \succcurlyeq_{\mathcal{K}} F(x^0) - F(x^{k+1}) \succcurlyeq_{\mathcal{K}} \sum_{\ell=0}^{k} -\alpha_\ell \delta f(x^\ell, d^\ell) e$$

because **A1** holds also. Then,

$$\frac{\langle w, F(x^0) - \mathcal{F}\rangle}{\delta \langle w, e\rangle} \geq \sum_{\ell=0}^{k} -\alpha_\ell f(x^\ell, d^\ell) > 0$$

for all $w \in G$. □

The Zoutendijk condition for vector optimization problems,

$$\sum_{k\geq 0} \frac{f^2(x^k, d^k)}{\|d^k\|^2} < \infty,$$

was introduced in [40]. With the next lemma, we show that a Zoutendijk's like condition is fulfilled by Algorithm 2.1

Assumptions **A1** are minimal, i.e., every results on this work need both to hold. Such hypotheses were already necessary for the algorithm in the scalar case.

**Lemma 2.9.** *Assume that* **A1** *holds and* $F$ *is convex. Then,*

$$\sum_{k\geq 0} \frac{1}{L + \frac{\nu_k}{2}} \frac{f^2(x^k, d^k)}{\|d^k\|^2} < \infty. \tag{2-11}$$

*Proof.* Observe that

$$\begin{aligned}\left(L+\frac{\nu_k}{2}\right)\alpha_k\omega^{-1}\|d^k\|^2 &= L\alpha_k\omega^{-1}\|d^k\|^2+\frac{\nu_k\alpha_k\omega^{-1}\|d^k\|^2}{2}\\ &= L\|d^k\|\|x^k+\alpha_k\omega^{-1}d^k-x^k\|+\frac{\nu_k\alpha_k\omega^{-1}\|d^k\|^2}{2}\\ &\geq \left|f(x^k+\alpha_k\omega^{-1}d^k,d^k)-f(x^k,d^k)\right|+\frac{\nu_k\alpha_k\omega^{-1}\|d^k\|^2}{2}\\ &\geq f(x^k+\alpha_k\omega^{-1}d^k,d^k)-f(x^k,d^k)+\frac{\nu_k\alpha_k\omega^{-1}\|d^k\|^2}{2}.\end{aligned}$$

Taking in account (2-3), we get

$$\begin{aligned}\left(L+\frac{\nu_k}{2}\right)\alpha_k\omega^{-1}\|d^k\|^2 &\geq f(x^k+\alpha_k\omega^{-1}d^k,d^k)-f(x^k,d^k)+\frac{\nu_k\alpha_k\omega^{-1}\|d^k\|^2}{2}\\ &> (\delta-1)f(x^k,d^k)>0.\end{aligned} \tag{2-12}$$

From these last inequalities we get that

$$0<\frac{1}{L+\frac{\nu_k}{2}}\frac{f^2(x^k,d^k)}{\|d^k\|^2}<\frac{1}{\omega(\delta-1)}\alpha_k f(x^k,d^k).$$

Since, by Lemma 2.8, $\sum_{k\geq 0}\alpha_k f(x^k,d^k)$ is convergent, we conclude that

$$\sum_{k\geq 0}\frac{1}{L+\frac{\nu_k}{2}}\frac{f^2(x^k,d^k)}{\|d^k\|^2}<\infty.$$

□

**Corollary 2.10.** *Assume that* **A1**, **A2** *hold and* $F$ *is convex. Then,*

$$\sum_{k\geq 0}\frac{f^2(x^k,d^k)}{\|d^k\|^2}<\infty.$$

*Proof.* Immediately. □

**Theorem 2.11.** *Assume that* **A1** *and* **A2** *hold and* $F$ *is convex. If*

$$\sum_{k\geq 0}\frac{1}{\|d^k\|^2}=\infty, \tag{2-13}$$

*then,*

$$\liminf\|v(x^k)\|=0.$$

*Proof.* Let us assume that there is $\gamma>0$ such that $\|v(x^k)\|>\gamma$ for all $k\geq 0$. Then, using Lemma 2.5, we have

$$0 < \frac{\gamma^2}{2} < \frac{\|v(x^k)\|^2}{2} < -f(x^k, d^k).$$

Hence,

$$\frac{\gamma^4}{4\left(L+\frac{\nu_k}{2}\right)\|d^k\|^2} < \frac{\|v(x^k)\|^4}{4\left(L+\frac{\nu_k}{2}\right)\|d^k\|^2} < \frac{1}{L+\frac{\nu_k}{2}}\frac{f^2(x^k, d^k)}{\|d^k\|^2}$$

for all $k \geq 0$. Since (2-11) holds,

$$\frac{\gamma^4}{4(L+\bar{\nu})}\sum_{k\geq 0}\frac{1}{\|d^k\|^2} < \sum_{k\geq 0}\frac{1}{L+\frac{\nu_k}{2}}\frac{f(x^k, d^k)^2}{\|d^k\|^2} < \infty,$$

in contradiction to our hypothesis, concluding that

$$\liminf \|v(x^k)\| = 0.$$

□

Therefore, using the function convexity hypothesis we were able to demonstrate the standard convergence result of the conjugate gradient method.

## 2.2 Non-convex case

In this section we consider non-convex $F$ with $JF$ Lipschitz continuous. This, we can present a new Algorithm of the Conjugate Gradient with a new line search.

**Algorithm 2.12.** *Let constants: $\rho > 0$, $\omega, \delta \in (0,1)$ and $\nu > 1$.*

**0. Initialization**: *Take $x^0 \in \mathbb{R}^n$. Compute $v(x^0)$ and initialize $k \leftarrow 0$.*

**1. Stopping criterium**: *If $v(x^k) = 0$, then STOP.*

**2. Direction**: *Define*

$$d^k = \begin{cases} v(x^k), & \text{if } k = 0 \\ v(x^k) + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{2-14}$$

*where $\beta_k$ is an algorithmic parameter.*

**3. Line search**: *Compute positive integers*

$$j_k = \min\left\{ j \geq 1 \;\middle|\; f(x^k + \rho\omega d^k, d^k) + \frac{\nu^j \rho\omega\|d^k\|^2}{2} \geq \delta f(x^k, d^k) \right\} \tag{2-15}$$

*and*

$$i_k = \min\left\{ i \geq 1 | \rho\omega^i < \frac{(\delta - 1)}{\left(L + \frac{\nu_k}{2}\right) \|d^k\|^2} f(x^k, d^k) \right\} \tag{2-16}$$

*where* $\nu_k = \nu^{j_k}$.

**4. Iteration step**: *Define*

$$\alpha_k = \rho\omega^{i_k} \tag{2-17}$$

*and*

$$x^{k+1} = x^k + \alpha_k d^k. \tag{2-18}$$

*Compute* $v(x^{k+1})$*, set* $k \leftarrow k + 1$*, and go to Step 1.*

The following Lemma shows that if $JF$ is $L-$Lipschitz continuous, then we can bound the step length.

**Lemma 2.13.** *Suppose that* $JF$ *is* $L-$*Lipschitz continuous and* $\hat{t} > 0$ *is such that*

$$JF(x^k + \hat{t}d^k)d^k + \frac{\nu_k \hat{t} \|d^k\|^2}{2} e \succ_K \delta JF(x^k)d^k \quad \textit{for all} \quad k = 0, 1, \ldots. \tag{2-19}$$

*Then,*

$$\hat{t} > \frac{(\delta - 1)}{\left(L + \frac{\nu_k}{2}\right) \|d^k\|^2} f(x^k, d^k).$$

*Proof.* Define $l(t) = \langle [JF(x^k + td^k) - \delta JF(x^k)]d^k, w\rangle + \frac{\nu_k t \|d^k\|^2}{2} \langle e, w\rangle$. Note that, $l(0) = (1 - \delta)\langle JF(x^k)d^k, w\rangle$ and, by (2-19), $l(\hat{t}) > 0$.

$$\begin{aligned} l(\hat{t}) - l(0) &= \langle [JF(x^k + \hat{t}d^k) - \delta JF(x^k)]d^k, w\rangle + \frac{\nu_k \hat{t} \|d^k\|^2}{2} \langle e, w\rangle \\ &- (1 - \delta)\langle JF(x^k)d^k, w\rangle \\ &= \langle [JF(x^k + \hat{t}d^k) - JF(x^k)]d^k, w\rangle + \frac{\nu_k \hat{t} \|d^k\|^2}{2} \langle e, w\rangle. \end{aligned}$$

By hypothesis, $JF$ is $L-$Lipschitz continuous, so

$$\begin{aligned} \langle [JF(x^k + \hat{t}d^k) - JF(x^k)]d^k, w\rangle &\leq \|[JF(x^k + \hat{t}d^k) - JF(x^k)]d^k\| \|w\| \\ &\leq L\hat{t}\|d^k\|^2. \end{aligned}$$

Since $l(\hat{t}) \geq 0$, we have

$$-l(0) < l(\hat{t}) - l(0) \leq L\hat{t}\|d^k\|^2 + \frac{\nu_k \hat{t}\|d^k\|^2}{2}\langle e, w\rangle$$
$$-l(0) < \left(L + \frac{\nu_k}{2}\right)\hat{t}\|d^k\|^2$$
$$\hat{t} > \frac{(\delta-1)\langle JF(x^k)d^k, w\rangle}{\left(L + \frac{\nu_k}{2}\right)\|d^k\|^2}.$$

Implying

$$\hat{t} > \frac{(\delta-1)}{\left(L + \frac{\nu_k}{2}\right)\|d^k\|^2}\max\langle JF(x^k)d^k, w\rangle = \frac{(\delta-1)}{\left(L + \frac{\nu_k}{2}\right)\|d^k\|^2} f(x^k, d^k).$$

□

Lemma 2.3 and Lemma 2.13 guarantee us the well definition of Algorithm 2.12.

Condition (2-16) of Algorithm 2.12 guarantees that $\alpha_k \leq \frac{(\delta-1)}{\left(L + \frac{\nu_k}{2}\right)\|d^k\|^2} f(x^k, d^k)$ and therefore by Lemma 2.13, we have $JF(x^k + \hat{t}d^k)d^k + \frac{\nu_k \hat{t}\|d^k\|^2}{2} e \preceq_{\mathcal{K}} \delta JF(x^k)d^k$ for all $k = 0, 1, \ldots$. So we can rewrite Lemma 2.6 and Corollary 2.7, whose respective proofs are identical, showing that function $F(x^k)$ will be monotonous descending.

**Lemma 2.14.** *Assume $JF$ $L-$Lipschitz continuous. Then,*

$$F(x^{k+1}) \preceq_{\mathcal{K}} F(x^k) + \delta\alpha_k f(x^k, d^k)e, \tag{2-20}$$

*for all $k \geq 0$.*

**Corollary 2.15.** *If $JF$ $L-$Lipschitz continuous, then $\{F(x^k)\}_{k\geq 0}$ is strictly monotone $\mathcal{K}$-decreasing, i.e., $F(x^{k+1}) \prec_{\mathcal{K}} F(x^k)$, $k = 0, 1, \ldots$.*

Following the same idea of replacing the convexity hypothesis of function $F$ with that of $JF$ $L-$Lipschitz continuous, the results, Lemma 2.8, Lemma 2.9, Corollary 2.10, and Theorem 2.11 can be reproduced and demonstrated in the same way. Thus, convergence is assured for this case.

## 2.3 General Case

In this section, we present an algorithm applicable to any continuously differentiable function $F$. This procedure does not require knowledge about the Lipschitz constant for the Jacobian of $F$.

**Algorithm 2.16.** *Let be four exogeneous positive constants: $\delta$ and $\omega < 1$, $\nu > 1$ and $\rho$.*

**0. Initialization**: *Take $x^0 \in \mathbb{R}^n$. Compute $v(x^0)$ and initialize $k \leftarrow 0$.*

**1. Stopping criterium**: *If $v(x^k) = 0$, then STOP.*

**2. Direction**: *Define*

$$d^k = \begin{cases} v(x^k), & \text{if } k = 0 \\ v(x^k) + \beta_k d^{k-1}, & \text{if } k \geq 1, \end{cases} \tag{2-21}$$

*where $\beta_k$ is an algorithmic parameter.*

**3. Line search**: *Compute positive integers*

$$j_k = \min\left\{ j \geq 1 \;\middle|\; f(x^k + \rho\omega d^k, d^k) + \frac{\nu^j \rho\omega \|d^k\|^2}{2} \geq \delta f(x^k, d^k) \right\} \tag{2-22}$$

*and*

$$i_k = \min\left\{ i \geq 1 \;\middle|\; f(x^k + \rho\omega^i d^k, d^k) + \frac{\nu_k \rho\omega^i \|d^k\|^2}{2} \leq \delta f(x^k, d^k) \right\} \tag{2-23}$$

*where $\nu_k = \nu^{j_k}$.*

**4. Iteration step**: *Define*

$$\alpha_k = \rho\omega^{i_k} \tag{2-24}$$

*and*

$$x^{k+1} = x^k + \alpha_k d^k. \tag{2-25}$$

*Compute $v(x^{k+1})$, set $k \leftarrow k+1$, and go to Step 1.*

The choice of updating $\beta_k$ remains deliberately open. In the next section, we will consider several choices of $\beta_k$ that result in globally convergent methods.

Well definiteness of Algorithm 2.16 follows from Lemma 2.3.

### 2.3.1 Convergence Analysis

Algorithm 2.16 successfully stops if a $\mathcal{K}$-critical point of $F$ is found. Hence, from now on, let us consider that $v(x^k) \neq 0$ for all $k \geq 0$.

From now on, we will need some additional hypotheses on the problem and/or Algorithm 2.16.

**Assumptions**

**A3.** Suppose that $0 < \gamma \leq \| v(x^k) \| \leq \bar{\gamma}$, and there exist constants $b > 1$ and $\lambda > 0$ such that, for all $k$,

$$\beta_k \leq b$$

and

$$\| s^{k-1} \| \leq \lambda \Rightarrow |\beta_k| \leq \frac{1}{2b},$$

where $s^{k-1} = x^k - x^{k-1}$.

Gilbert and Nocedal introduced in [25] a property to show the convergence of the Conjugated Gradient Algorithm in the scalar context for the betas of PRP and HS, this property has been extended to vector minimization in [40] and is reproduced in this work as hypothesis A3. The following results assure us that changes in line search directions are not too sudden.

**Lemma 2.17.** *Consider Algorithm 2.16 with $\beta_k \geq 0$ and $d^k$ is a $\mathcal{K}$-descent direction of $F$ at $x^k$. Assume that assumptions* **A1** *and* **A2** *hold. Then,*

*(i)* $\displaystyle\sum_{k\geq 0} \frac{\| v(x^k) \|^4}{\| d^k \|^2} < \infty,$

*(ii)* $\displaystyle\sum_{k\geq 1} \| u^k - u^{k-1} \|^2 < \infty$, *where* $u^k = d^k / \| d^k \|$.

*Proof.* Once we have item (i), the proof of item (ii) would be quite similar to the proof of Lemma 5.8(ii) in [40]. Let us proof item (i).

Since $d^k$ is a descent direction of $F$ at $x^k$, it implies that $d^k \neq 0$. Hence, $\| v(x^k) \|^4 / \| d^k \|^2$ and $u^k$ are well defined. For (2-21) $-\beta_k d^{k-1} = -d^k + v(x^k)$, so

$$\| -\beta_k d^{k-1} \|^2 = \| -d^k + v(x^k) \|^2 .$$

By Lemma 1.6 (c)

$$\beta_k^2 \| d^{k-1} \|^2 \leq [\| d^k \| + \| v(x^k) \|]^2 \leq 2 \| d^k \|^2 + 2 \| v(x^k) \|^2$$

$$\frac{\| d^k \|^2}{\| d^{k-1} \|^2} \geq \frac{\beta_k^2}{2} - \frac{\| v(x^k) \|^2}{\| d^{k-1} \|^2}. \tag{2-26}$$

On the other hand, using (2-21) and (2-8)

$$0 < -f(x^k, v(x^k)) \leq -f(x^k, d^k) + \beta_k f(x^k, d^{k-1}) \leq -f(x^k, d^k) + \delta\beta_k f(x^{k-1}, d^{k-1}).$$

From the previous inequality and Lemma 1.6 (b) with $\alpha = 1$, we obtain

$$\begin{aligned} f^2(x^k, v(x^k)) &\leq (-f(x^k, d^k) + \delta\beta_k f(x^{k-1}, d^{k-1}))^2 \\ &= f^2(x^k, d^k) + \delta^2\beta_k^2 f^2(x^{k-1}, d^{k-1}) - 2f(x^k, d^k)\delta\beta_k f(x^{k-1}, d^{k-1}) \\ &\leq f^2(x^k, d^k) + \delta^2\beta_k^2 f^2(x^{k-1}, d^{k-1}) + 2f^2(x^k, d^k)\delta^2 + \beta_k^2 f^2(x^{k-1}, d^{k-1})/2 \\ &= (1 + 2\delta^2)(f^2(x^k, d^k) + \beta_k^2 f^2(x^{k-1}, d^{k-1})/2). \end{aligned}$$

Since $d^k$ is a descent direction for $F$ at $x^k$, $f(x^k, v(x^k)) \leq - \| v(x^k) \|^2 /2$, then $f^2(x^k, v(x^k)) \geq \| v(x^k) \|^4 /4$. Hence,

$$f^2(x^k, d^k) + \frac{\beta_k^2}{2} f^2(x^{k-1}, d^{k-1}) \geq \frac{1}{1+2\delta^2} f^2(x^k, v(x^k)) \geq \frac{1}{4(1+2\delta^2)} \| v(x^k) \|^4 . \tag{2-27}$$

Note that, by (2-26),

$$\begin{aligned}
&\frac{f^2(x^k, d^k)}{\| d^k \|^2} + \frac{f^2(x^{k-1}, d^{k-1})}{\| d^{k-1} \|^2} = \frac{1}{\| d^k \|^2} \left[ f^2(x^k, d^k) + \frac{\| d^k \|^2}{\| d^{k-1} \|^2} f^2(x^{k-1}, d^{k-1}) \right] \\
&\geq \frac{1}{\| d^k \|^2} \left[ f^2(x^k, d^k) + \left( \frac{\beta_k^2}{2} - \frac{\| v(x^k) \|^2}{\| d^{k-1} \|^2} \right) f^2(x^{k-1}, d^{k-1}) \right] \\
&= \frac{1}{\| d^k \|^2} \left[ f^2(x^k, d^k) + \frac{\beta_k^2}{2} f^2(x^{k-1}, d^{k-1}) - \frac{\| v(x^k) \|^2}{\| d^{k-1} \|^2} f^2(x^{k-1}, d^{k-1}) \right]
\end{aligned}$$

Using (2-27),

$$\begin{aligned}
\frac{f^2(x^k, d^k)}{\| d^k \|^2} + \frac{f^2(x^{k-1}, d^{k-1})}{\| d^{k-1} \|^2} &\geq \frac{1}{\| d^k \|^2} \left[ \frac{\| v(x^k) \|^4}{4(1+2\delta^2)} - \frac{\| v(x^k) \|^2}{\| d^{k-1} \|^2} f^2(x^{k-1}, d^{k-1}) \right] \\
&= \frac{\| v(x^k) \|^2}{\| d^k \|^2} \left[ \frac{\| v(x^k) \|^2}{4(1+2\delta^2)} - \frac{f^2(x^{k-1}, d^{k-1})}{\| d^{k-1} \|^2} \right].
\end{aligned}$$

The Zoutendijk condition holds under the hypotheses, and it implies that $f^2(x^k, d^k)/ \| d^k \|^2$ tends to zero, so we have

$$\frac{f^2(x^k, d^k)}{\| d^k \|^2} + \frac{f^2(x^{k-1}, d^{k-1})}{\| d^{k-1} \|^2} \geq \frac{\| v(x^k) \|^4}{8(1+2\delta^2) \| d^k \|^2}$$

for all sufficiently large $k$. Using Zoutendijk the proof is complete. □

For $\lambda > 0$ and a positive integer $\Delta$, define

$$\mathcal{M}_{k,\Delta}^{\lambda} = \{ i \in \mathbb{N} | k \leq i \leq k + \Delta - 1, \| s^{k-1} \| > \lambda \}$$

and denote by $|\mathcal{M}_{k,\Delta}^{\lambda}|$ the number of elements of $\mathcal{M}_{k,\Delta}^{\lambda}$.

Now we will show that the step size can not be too short.

**Lemma 2.18.** *Consider Algorithm 2.16. Assume that* **A1**, **A2** *and* **A3** *hold.* $d^k$ *is descent direction of* $F$ *at* $x^k$*. If there exists* $\gamma > 0$ *such that* $\| v(x^k) \| \geq \gamma$*, for all* $k > 0$*, then there exists* $\lambda > 0$ *such that, for any* $\Delta \in \mathcal{N}$ *and any index* $k_0$*, there is a greater index* $k \geq k_0$ *such that*

$$|\mathcal{M}_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}.$$

*Proof.* See Lemma 5.9 of [40]. □

**Theorem 2.19.** *Assume that level set* $\mathcal{L} = \{x | F(x) \leq F(x_0)\}$ *is bounded,* **A1** *and* **A3** *hold. Consider Algorithm 2.16 where* $\beta_k \geq 0$, $d^k$ *is a descent direction of* $F$ *at* $x^k$. *Then,*

$$\liminf \|v(x^k)\| = 0.$$

*Proof.* See Theorem 5.10 of [40]. □

## 2.4 Analysis of convergence for specific $\beta_k$'s

In this section we will present the convergence analysis using the line search introduced above, for the $\beta_k$'s specifics of the Fletcher-Reeves (FR), Conjugate Descent (CD), Dai-Yuan (DY), Polak-Ribière-Polyak (PRP) and Hestenes-Stiefel (HS). These $\beta_k$'s, as well as their convergence analysis were taken to the vector context in [40], using as line search the standard Wolfe conditions or the strong Wolfe conditions.

The parameter, originally proposed by Fletcher and Reeves in [18], was modified as

$$\beta_k^{FR} = \frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))}. \tag{2-28}$$

The called conjugate descent parameter, proposed by Fletcher in [19], was modified as

$$\beta_k^{CD} = \frac{f(x^k, v(x^k))}{f(x^{k-1}, d^{k-1})}. \tag{2-29}$$

Dai and Yuan in [9] proposed this parameter modified as

$$\tilde{\beta}_k^{DY} = \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})}.$$

Lemma 1.9 guarantees positiveness of $\beta_k^{FR}$ and $\beta_k^{CD}$. In [40] the positiveness of $\tilde{\beta}_k^{DY}$ is a consequence of the Wolfe-like line search. In our case, Algorithm 2.16 does not guarantee that $f(x^k, d^{k-1}) > f(x^{k-1}, d^{k-1})$. Therefore, we redefine the Dai-Yuan parameter as

$$\beta_k^{DY} = \begin{cases} \tilde{\beta}_k^{DY}, & \text{if} \quad f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1}) > 0 \\ 0, & \text{if} \quad f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1}) \leq 0 \end{cases}. \tag{2-30}$$

Gilbert and Nocedal in [25] proved that global convergence can be obtained for $\beta_k = max\{\beta_k^{PRP}, 0\}$ and $\beta_k = max\{\beta_k^{HS}, 0\}$, for scalar minimization case. In vector optimization context the $PRP$ and $HS$ parameters are given by

$$\tilde{\beta}_k^{PRP} = \frac{-f(x^k, v(x^k)) + f(x^{k-1}, v(x^k))}{-f(x^{k-1}, v(x^{k-1}))}, \quad where \quad \beta_k^{PRP} = max\{0, \tilde{\beta}_k^{PRP}\}, \tag{2-31}$$

and

$$\tilde{\beta}_k^{HS} = \frac{-f(x^k, v(x^k)) + f(x^{k-1}, v(x^k))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})}, \quad where \quad \beta_k^{HS} = max\{0, \tilde{\beta}_k^{HS}\}. \tag{2-32}$$

Before we begin the analysis for specific $\beta_k$'s, let us do an important observation.

**Lemma 2.20.** *Consider Algorithm 2.16, if A1 hold and and* $\Sigma = \{k \geq 0 \colon \beta_k = 0\}$ *is infinite, then,*

$$\liminf_k \|v(x^k)\| = 0.$$

*Proof.* We claim that

$$\lim_{k\to\infty,\ k\in\Sigma} \|v(x^k)\| = 0.$$

Observe that Lemma 2.8 implies $\lim_{k\to\infty} \alpha_k f(x^k, v(x^k)) = \lim_{k\to\infty} \alpha_k \|v(x^k)\| = 0$. Then,

$$\lim_{k\to\infty,\ k\in\Sigma} \alpha_k f(x^k, v(x^k)) = \lim_{k\to\infty, k\in\Sigma} \alpha_k \|v(x^k)\| = 0.$$

Now, we have two cases to analyse.

- Case 1: $\liminf \alpha_k > 0$. In this case $\lim_{k\to\infty,\ k\in\Sigma} \|v(x^k)\| = 0$.
- Case 2: $\liminf \alpha_k = 0$. In this case, there exists infinite $\Sigma_0 \subset \Sigma$ such that $\lim_{k\to\infty,\ k\in\Sigma_0} \alpha_k = 0$. By (2-23) and (2-24) ,

$$f(x^k + \frac{\alpha_k}{\omega} v(x^k), v(x^k)) + \frac{\nu_k \frac{\alpha_k}{\omega} \|v(x^k)\|^2}{2} > \delta f(x^k, v(x^k)), \quad \text{for all} \quad k \in \Sigma_0.$$

Then, for all $\varepsilon > 0$ there exists $\hat{k}(\varepsilon)$ such that

$$\begin{aligned} 0 &\leq (\delta - 1) f(x^k, v(x^k)) \\ &\leq \left| f(x^k + \alpha_k v(x^k)/\omega, v(x^k)) - f(x^k, v(x^k)) \right| + \frac{\nu_k \alpha_k \|v(x^k)\|^2}{2\omega} < \varepsilon \end{aligned}$$

for all $k \in \Sigma_0$ and $k \geq \hat{k}(\varepsilon)$, because $f(\cdot, v(x^k))$ is continuos. In other words,

$$0 = \lim_{k\to\infty,\ k\in\Sigma_0} f(x^k, v(x^k)) = \lim_{k\to\infty,\ k\in\Sigma_0} \|v(x^k)\|.$$

□

The next paragraphs are dedicated to the convergence studies when the $\beta_k$'s are specified as one between the presented before.

### 2.4.1 Fletcher-Reeves

The following theorem shows us that under some hypotheses Algorithm 2.16, using $\beta_k^{FR}$ converges. This result will serve as a basis for demonstrating the same convergence's results for some of the $\beta_k$'s already mentioned.

**Theorem 2.21.** *Consider Algorithm 2.16 with* $0 \leq \beta_k < \eta\beta_k^{FR}$*, where* $0 < \eta < 1$ *and assume that* **A1** *hold. Then,*

$$\liminf_k \|v(x^k)\| = 0.$$

*Proof.* know that, $d^k$ satisfies the sufficient descent condition. Assume, by contradiction, that there exists $\gamma$ such that $0 < \gamma \leq \|v(x^k)\|$ for all $k \geq 0$. By Lemma 1.10 (b), remember that

$$\|v(x^k)\|^2 \leq -2f(x^k, v(x^k)).$$

Then,

$$\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} \leq \frac{4}{\|v(x^k)\|^2} \leq \frac{4}{\gamma^2}. \tag{2-33}$$

Observe that for any $a$ and $b \in \mathbb{R}$, taking $\alpha = \eta/\sqrt{2(1-\eta^2)}$, using Lemma 1.6 (d),

$$(a+b)^2 \leq (1+2\alpha^2)a^2 + \left(1 + \frac{1}{2\alpha^2}\right)b^2 = \frac{a^2}{1-\eta^2} + \frac{b^2}{\eta^2}$$

is true. Then, using (2-21), the Triangle Inequality and equation above we get

$$\begin{aligned}\|d^k\|^2 = \|v(x^k) + \beta_k d^{k-1}\|^2 &\leq \left(\|v(x^k)\| + \beta_k\|d^{k-1}\|\right)^2 \\ &\leq \frac{\|v(x^k)\|^2}{1-\eta^2} + \frac{\beta_k^2\|d^{k-1}\|^2}{\eta^2}.\end{aligned}$$

Now, dividing by $f^2(x^k, v(x^k))$ and using hypothesis,

$$\begin{aligned}\frac{\|d^k\|^2}{f^2(x^k, v(x^k))} &\leq \frac{1}{1-\eta^2}\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} + \frac{\beta_k^2}{\eta^2}\frac{\|d^{k-1}\|^2}{f^2(x^k, v(x^k))} \\ &\leq \frac{1}{1-\eta^2}\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} + (\beta_k^{FR})^2\frac{\|d^{k-1}\|^2}{f^2(x^k, v(x^k))} \\ &= \frac{1}{1-\eta^2}\frac{\|v(x^k)\|^2}{f^2(x^k, v(x^k))} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, v(x^{k-1}))}.\end{aligned}$$

Therefore, by (2-33),

$$\begin{aligned}
\frac{\|d^k\|^2}{f^2(x^k, v(x^k))} &\leq \frac{4}{(1-\eta^2)\gamma^2} + \frac{\|d^{k-1}\|^2}{f^2(x^{k-1}, v(x^{k-1}))} \\
&\vdots \\
&\leq \frac{4}{(1-\eta^2)\gamma^2}k + \frac{\|d^0\|^2}{f^2(x^0, v(x^0))} \\
&\leq \frac{4}{(1-\eta^2)\gamma^2}k + \frac{4}{\gamma^2} \\
&= \frac{4}{\gamma^2}\left(\frac{k}{1-\eta^2} + 1\right) = \frac{4}{\gamma^2}\left(\frac{k+1-\eta^2}{1-\eta^2}\right).
\end{aligned}$$

Concluding

$$\frac{f^2(x^k, v(x^k))}{\|d^k\|^2} \geq \frac{\gamma^2(1-\eta^2)}{4(k+1-\eta^2)} \geq \frac{\gamma^2(1-\eta^2)}{4}\frac{1}{k+1}.$$

Henceforth, by sufficient descent condition and inequality above,

$$\sum_k \frac{f^2(x^k, d^k)}{\|d^k\|^2} \geq \sum_k c^2 \frac{f^2(x^k, v(x^k))}{\|d^k\|^2} \geq \frac{c^2\gamma^2(1-\eta^2)}{4}\sum_k \frac{1}{k+1} = \infty,$$

in contradiction with Zoutendijk's condition. Thus, theorem is demonstrated. □

### 2.4.2 Conjugate Descent

Now we will show the convergence of Algorithm 2.16 using the $\beta_k^{CD}$. Convergence is guaranteed by showing that $CD$ is less than a multiple of $FR$.

**Lemma 2.22.** *Consider Algorithm 2.16 with $0 \leq \beta_k \leq c\beta_k^{CD}$. Then, $d^k$ satisfies the sufficient descent condition, with constant $c = 1-\delta$.*

*Proof.* By (2-21) , (2-8), definition $\beta_k^{CD}$

$$\begin{aligned}
f(x^k, d^k) &= f(x^k, v(x^k) + \beta_k d^{k-1}) \\
&\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \\
&\leq f(x^k, v(x^k)) + \beta_k \delta f(x^{k-1}, d^{k-1}) \\
&\leq f(x^k, v(x^k)) + \beta_k^{CD} \delta f(x^{k-1}, d^{k-1}) \\
&\leq f(x^k, v(x^k)) + \delta f(x^k, v(x^k)) \\
&\leq (1+\delta) f(x^k, v(x^k)) \\
&\leq (1-\delta) f(x^k, v(x^k)).
\end{aligned}$$

□

The following Lemma compares $\beta_k^{CD}$ and $\beta_k^{FR}$. So we can use Theorem 2.21 to demonstrate the convergence of Algorithm 2.16 using the $\beta_k^{CD}$.

**Lemma 2.23.** *Consider Algorithm 2.16, $\beta_k^{FR}$ defined in (2-28) and $\beta_k^{CD}$ defined in (2-29). Then,*

$$\beta_k^{CD} \leq \frac{1}{1-\delta}\beta_k^{FR}.$$

*Proof.* By Lemma 2.22 $d^k$ satisfies the sufficient descent condition, then

$$\begin{aligned}\beta_k^{CD} &= \frac{f(x^k, v(x^k))}{f(x^{k-1}, d^{k-1})} = \frac{1-\delta}{1-\delta}\frac{f(x^k, v(x^k))}{f(x^{k-1}, d^{k-1})} \\ &\leq \frac{1}{1-\delta}\frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))} \leq \frac{1}{1-\delta}\beta_k^{FR}\end{aligned}$$

□

**Theorem 2.24.** *Consider Algorithm 2.16 where $0 < \beta_k = \eta\beta_k^{CD}$ and $0 \leq \eta \leq 1-\delta$. Assume that* **A1** *holds. Then,*

$$\liminf_{k \longrightarrow \infty} \|v(x^k)\| = 0.$$

*Proof.* It follows from Lemma 2.22 that $d^k$ satisfies the sufficient descent condition, with $c = 1-\delta$ for all $k$. Therefore,

$$0 \leq \beta_k \leq \eta\beta_k^{CD} \leq \frac{\eta}{1-\delta}\beta_k^{FR}.$$

As $0 \leq \dfrac{\eta}{1-\delta} \leq 1$, this proof follows as the one of Theorem 2.21. □

### 2.4.3 Dai-Yuan

In the same way as in CD, the convergence of Algorithm 2.16 using the $\beta_k^{DY}$ will be shown using Theorem 2.21, of the convergence of $FR$.

**Lemma 2.25.** *Consider Algorithm 2.16, with $0 \leq \beta_k \leq \beta_k^{DY}$. Then $d^k$ satisfies the condition of sufficient descent with $c = 1/(1+\delta)$, that is,*

$$f(x^k, d^k) \leq cf(x^k, v(x^k)).$$

*Proof.* If $f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1}) \leq 0$, then for (2-21),

$$\begin{aligned}f(x^k, d^k) &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \leq f(x^k, v(x^k)) + \beta_k^{DY} f(x^k, d^{k-1}) \\ &= f(x^k, v(x^k)) \leq cf(x^k, v(x^k)).\end{aligned}$$

If $f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1}) > 0$, for (2-21),

$$\begin{aligned}
f(x^k, d^k) &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \\
&\leq f(x^k, v(x^k)) + \beta_k^{DY} f(x^k, d^{k-1}) \\
&= f(x^k, v(x^k)) + \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} f(x^k, d^{k-1}) \\
&= f(x^k, v(x^k)) \left( \frac{-f(x^{k-1}, d^{k-1})}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \right) \\
&= f(x^k, v(x^k)) \left( \frac{f(x^{k-1}, d^{k-1})}{f(x^{k-1}, d^{k-1}) - f(x^k, d^{k-1})} \right).
\end{aligned}$$

Using (2-8) we have, $-f(x^k, d^{k-1}) \geq -\delta f(x^{k-1}, d^{k-1})$, so

$$\begin{aligned}
f(x^k, d^k) &\leq f(x^k, v(x^k)) \left( \frac{f(x^{k-1}, d^{k-1})}{f(x^{k-1}, d^{k-1}) - \delta f(x^{k-1}, d^{k-1})} \right) \\
&= f(x^k, v(x^k)) \left( \frac{1}{1-\delta} \right).
\end{aligned}$$

As $1 - \delta < 1 + \delta$ and $f(x^k, v(x^k)) < 0$,

$$f(x^k, d^k) \leq \frac{1}{1+\delta} f(x^k, v(x^k)).$$

□

Let the set $K = \{k \in \mathbb{N} | \ \beta_k^{DY} > 0\}$. Assume that $K$ is infinite and $k \in K$.

**Theorem 2.26.** *Let assumption in* **A1** *hold. Consider Algorithm 2.16 with* $\beta_k = \eta \beta_k^{DY}$ *where* $0 < \eta \leq \dfrac{1-\sigma}{1+\sigma}$. *If* $K$ *is infinite, then*

$$\liminf_{k \longrightarrow \infty} \|v(x^k)\| = 0.$$

*Proof.* Using (2-8) and Lemma 2.25,

$$\begin{aligned}
f(x^{k-1}, d^{k-1}) - f(x^k, d^{k-1}) &\geq f(x^{k-1}, d^{k-1}) - \delta f(x^{k-1}, d^{k-1}) \\
&= (1-\delta) f(x^{k-1}, d^{k-1}) \\
&= -(\delta - 1) f(x^{k-1}, d^{k-1}) \\
&\geq (\delta - 1) \left( \frac{-1}{1+\delta} \right) f(x^{k-1}, v(x^{k-1})) \\
&= \frac{1-\delta}{1+\delta} f(x^{k-1}, v(x^{k-1}))
\end{aligned}$$

Then,

$$\frac{1+\delta}{1-\delta} \geq \frac{f(x^{k-1}, v(x^{k-1}))}{f(x^{k-1}, d^{k-1}) - f(x^k, d^{k-1})} \Rightarrow \frac{-f(x^{k-1}, v(x^{k-1}))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \leq \frac{1+\delta}{1-\delta}.$$

Define $\sigma = \dfrac{\eta(1+\delta)}{1-\delta}$. By definition of $\beta_k$, we have

$$\begin{aligned}
\beta_k &= \eta \beta_k^{DY} \\
&= \frac{\sigma(1-\delta)}{1+\delta} \left( \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \right) \frac{-f(x^{k-1}, v(x^{k-1}))}{-f(x^{k-1}, v(x^{k-1}))} \\
&= \frac{\sigma(1-\delta)}{1+\delta} \left( \frac{-f(x^k, v(x^k))}{-f(x^{k-1}, v(x^{k-1}))} \right) \left( \frac{-f(x^{k-1}, v(x^{k-1}))}{f(x^k, d^{k-1}) - f(x^{k-1}, d^{k-1})} \right) \\
&\leq \frac{\sigma(1-\delta)}{1+\delta} \left( \frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))} \right) \frac{1+\delta}{1-\delta} \\
&= \sigma \frac{f(x^k, v(x^k))}{f(x^{k-1}, v(x^{k-1}))} \\
&= \sigma \beta_k^{FR}.
\end{aligned}$$

As $0 \leq \sigma < 1$, by Theorem 2.21, the result follows. □

Following the same idea presented in [40], we can modify the $\beta_k^{DY}$ and get the convergence of Algorithm 2.16. The modified parameter of the Dai-Yuan will be defined by

$$\tilde{\beta}_k^{mDY} = \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1})}$$

and

$$\beta_k^{mDY} = \begin{cases} \tilde{\beta}_k^{mDY}, & \text{if} \quad f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1}) > 0 \\ 0, & \text{if} \quad f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1}) \leq 0 \end{cases}, \quad Whith \ \tau > 1. \tag{2-34}$$

Similarly to Lemma 2.25, we show that $d^k$ satisfies the sufficient descent condition, with $c = \tau/(\tau + \delta)$.

**Lemma 2.27.** *Consider Algorithm 2.16, with $0 \leq \beta_k \leq \beta_k^{mDY}$. Then $d^k$ satisfies the condition of sufficient descent with $c = \tau/(\tau+\delta)$, this is,*

$$f(x^k, d^k) \leq c f(x^k, v(x^k)).$$

*Proof.* If $f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1}) \leq 0$, then for (2-21),

$$\begin{aligned} f(x^k, d^k) &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \leq f(x^k, v(x^k)) + \beta_k^{mDY} f(x^k, d^{k-1}) \\ &= f(x^k, v(x^k)) \leq c f(x^k, v(x^k)). \end{aligned}$$

If $f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1}) > 0$, for (2-21),

$$\begin{aligned} f(x^k, d^k) &\leq f(x^k, v(x^k)) + \beta_k f(x^k, d^{k-1}) \\ &\leq f(x^k, v(x^k)) + \beta_k^{mDY} f(x^k, d^{k-1}) \\ &= f(x^k, v(x^k)) + \frac{-f(x^k, v(x^k))}{f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1})} f(x^k, d^{k-1}) \\ &= \tau f(x^k, v(x^k)) \left( \frac{-f(x^{k-1}, d^{k-1})}{f(x^k, d^{k-1}) - \tau f(x^{k-1}, d^{k-1})} \right) \\ &= \tau f(x^k, v(x^k)) \left( \frac{f(x^{k-1}, d^{k-1})}{\tau f(x^{k-1}, d^{k-1}) - f(x^k, d^{k-1})} \right). \end{aligned}$$

Using (2-8), we have $-f(x^k, d^{k-1}) \geq -\delta f(x^{k-1}, d^{k-1})$, soon

$$\begin{aligned} f(x^k, d^k) &\leq \tau f(x^k, v(x^k)) \left( \frac{f(x^{k-1}, d^{k-1})}{\tau f(x^{k-1}, d^{k-1}) - \delta f(x^{k-1}, d^{k-1})} \right) \\ &= f(x^k, v(x^k)) \left( \frac{\tau}{\tau - \delta} \right). \end{aligned}$$

As $\tau - \delta < \tau + \delta$ and $f(x^k, v(x^k)) < 0$,

$$\begin{aligned} f(x^k, d^k) &\leq \frac{\tau}{\tau - \delta} f(x^k, v(x^k)) \\ &\leq \frac{\tau}{\tau + \delta} f(x^k, v(x^k)) \end{aligned}$$

□

**Theorem 2.28.** *Let it be as assumed in* **A1**. *Consider Algorithm 2.16 with* $\beta_k = \beta_k^{mDY}$. *Then,*

$$\liminf_{k \longrightarrow \infty} \|v(x^k)\| = 0.$$

*Proof.* See proof of Theorem 5.7 in [40]. □

### 2.4.4 Polak-Rivière-Polyak e Hestenes-Stiefel

Lastly, we will demonstrate now the convergence of Algorithm 2.16 for $\beta_k^{PRP}$ and $\beta_k^{HS}$.

**Theorem 2.29.** *Assume that the level set* $\mathcal{L} = \{x | F(x) \leq F(x_0)\}$ *is bounded, there exists an open set* $\mathcal{N}$ *such that* $\mathbb{L} = \{x | F(x) \leq F(x_0)\} \subset \mathcal{N}$ *and the Jacobian* $JF$

*is Lipschitz continuos on $\mathcal{N}$ with constant $L > 0$. Consider Algorithm 2.16 with $\beta_k = max\{\beta_k^{PRP}\}$ or $\beta_k = max\{\beta_k^{HS}, 0\}$, $d^k$ is a descent direction of $F$ at $x^k$. Then,*

$$\liminf \|v(x^k)\| = 0.$$

*Proof.* See proof of Theorem 5.11 in [40]. □

## 2.5 Complexity

In this section we will extend the results presented in [14, 15] to the vector context. These are about the complexity of Algorithm 2.16.

**Lemma 2.30.** *Consider Algorithm 2.16. Assume that $0 \le \alpha_k \le \rho\|v(x^k)\|^2/\|d^k\|^2$, (**A1**) and (**A2**). If $v(x^k) \neq 0$ for all $k$. Then,*

*(a) $\alpha_k \ge \tau \parallel v(x^k) \parallel^2 / \parallel d^k \parallel^2$ for some $\tau > 0$.*

*(b) Consider $\beta_k^{PRP}$, $\parallel d^k \parallel \le (1 + L\rho) \parallel v(x^k) \parallel$.*

*Proof.* First, we will prove (a). By Lemma 2.3, $i_k$ is computed and for $(3-14)$ $\alpha_k$ can be found after a finite number of trials. Then, $\omega^{-1}\alpha_k$ does not satisfy (3-13). This is,

$$f(x^k + \omega^{-1}\alpha_k d^k, d^k) + \frac{\nu_k\omega^{-1}\alpha_k\|d^k\|^2}{2} > \delta f(x^k, d^k).$$

Adding $-f(x^k, d^k)$ to both sides it yields

$$f(x^k + \omega^{-1}\alpha_k d^k, d^k) - f(x^k, d^k) + \frac{\nu_k\omega^{-1}\alpha_k\|d^k\|^2}{2} > -(1-\delta)f(x^k, d^k).$$

From Lemma 1.11,

$$\begin{aligned} L\alpha_k\omega^{-1} \parallel d^k \parallel^2 + \frac{\nu_k\omega^{-1}\alpha_k\|d^k\|^2}{2} &> -(1-\delta)f(x^k, d^k) \\ \left(L + \frac{\nu_k}{2}\right)\alpha_k\omega^{-1} \parallel d^k \parallel^2 &> -(1-\delta)f(x^k, d^k). \end{aligned}$$

for Lemma 2.5 and (**A2**),

$$\begin{aligned} \left(L + \frac{\nu_k}{2}\right)\alpha_k\omega^{-1} \parallel d^k \parallel^2 &> (1-\delta)\frac{\parallel v(x^k) \parallel^2}{2} \\ \alpha_k &> \frac{(1-\delta)\omega}{L+\nu}\frac{\parallel v(x^k) \parallel^2}{\parallel d^k \parallel^2} \\ \alpha_k &> \tau\frac{\parallel v(x^k) \parallel^2}{\parallel d^k \parallel^2}, \end{aligned}$$

where $\tau = \dfrac{(1-\delta)\omega}{L+\nu}$. Item (a) is shown.

To prove part (b), if $k=0$, then $\| d^0 \| \leq (1+L\rho) \| v(x^0) \|$.

When $k \geq 1$, it follows from (2-21), (2-31) and Lemma 1.11 that

$$\begin{aligned}
\|d^{k+1}\| &= \|v(x^{k+1}) + \beta_{k+1}^{PRP} d^k\| \\
&\leq \|v(x^{k+1})\| + \frac{\|-f(x^{k+1}, v(x^{k+1})) + f(x^k, v(x^{k+1}))\|}{\|-f(x^k, v(x^k))\|}\|d^k\| \\
&= \|v(x^{k+1})\| \left(1 + \frac{L\alpha_k\|d^k\|^2}{\|f(x^k, v(x^k))\|}\right).
\end{aligned}$$

Using Lemma 2.5 and hypothesis,

$$\begin{aligned}
\|d^{k+1}\| &\leq \|v(x^{k+1})\| \left(1 + \frac{2L\alpha_k\|d^k\|^2}{\|v(x^k)\|^2}\right) \\
&\leq \|v(x^{k+1})\|(1+2L\rho).
\end{aligned}$$

□

**Theorem 2.31.** *Consider Algorithm 2.16 with* $\beta_k^{PRP}$. *Assume that* $0 \leq \alpha \leq \rho\|v(x^k)\|^2/\|d^k\|^2$, (**A1**) *and* (**A2**). *If* $v(x^k) \neq 0$ *for all* $k$, *there exist positive numbers* $\alpha$, $\beta$, $\gamma$ *such that the following holds*

*(a)* $\alpha_k \geq \alpha > 0$;

*(b)* $\sum_{k=0}^{\infty} \|v(x^k)\|^2 \leq \beta(F(x^0) - F^*)$;

*(c)* $\sum_{k=0}^{\infty} \|d^k\|^2 \leq \gamma(f(x^0) - F^*)$, *where* $F^*$ *is the limit of the decreasing and lower bounded sequence* $\{F(x^k)\}$.

*Proof.*

(a) Follow from Lemma 2.30,

$$\alpha_k \geq \tau\frac{\|v(x^k)\|^2}{\|d^k\|^2} = \frac{\tau}{(1+L\rho)^2} := \alpha > 0.$$

(b) From Lemma 2.14

$$\begin{aligned}
F(x^{k+1}) = F(x^k + \alpha_k d^k) &\preceq_{\mathcal{K}} F(x^k) + \delta\alpha_k f(x^k, d^k)e \\
\alpha_k(-f(x^k, d^k)e) &\preceq_{\mathcal{K}} \delta^{-1}(F(x^k) - F(x^{k+1})).
\end{aligned}$$

By Lemma 2.5 and summing,

$$\sum_{i=0}^{k} \alpha_i(-f(x^i,d^i)e) \preceq_{\mathcal{K}} \delta^{-1}(F(x^0)-F(x^{k+1}))$$
$$\sum_{i=0}^{k} \alpha_i\|v(x^i)\|^2 \leq 2\delta^{-1}(F(x^0)-F(x^{k+1})).$$

Using (a)

$$\sum_{i=0}^{k} \|v(x^i)\|^2 \leq 2(k\alpha\delta)^{-1}(F(x^0)-F(x^{k+1}))$$
$$\sum_{i=0}^{\infty} \|v(x^i)\|^2 \leq \beta(F(x^0)-F^*).$$

(c) Combining the item (b) with (b) of Lemma 2.30,

$$\left(\frac{\|d^k\|}{\|1+L\rho\|}\right)^2 \leq \|v(x^k)\|^2$$
$$\frac{1}{(1+L\rho)^2}\sum_{i=1}^{k} \|d^i\|^2 \leq \sum_{i=1}^{k} \|v(x^i)\|^2 \leq \beta(F(x^0)-F^*)$$
$$\sum_{i=1}^{k} \|d^i\|^2 \leq \beta(1+L\rho)^2(F(x^0)-F^*)$$
$$\sum_{i=1}^{k} \|d^i\|^2 \leq \gamma(F(x^0)-F^*).$$

□

## 2.6 Computational experiments

We will now present some numerical experiments to verify the applicability of the proposed conjugate gradient with the new line search. Check effectiveness of the method developed with all the betas presented in section 2.4, in addition to testing the constants that present best performance of the method. The sets of examples are divided in two groups, a convex and a non-convex group. All problems presented in this section are multiobjective, this is, $\mathcal{K} = R^m_+$.

The experiences were done using MATLAB R2020 on a computer with CPU Intel Core i7 2GHz and 8GB of memory. We stopped the execution at $x^k$ declaring convergence if $\theta(x^k) \geq -5 \times eps^{1/2}$, where $eps$ denotes the machine precision, in our case, $eps = 2^{-52} \approx 2.22 \times 10^{-16}$. This is the convergence criterion considered in the numerical tests of [20]. Since, by Lemma 1.10, $v(x) = 0$ if and only if $\theta(x) = 0$, this is a reasonable stopping criterion. The maximum number of allowed iterations was set to 10000. If iteration 10000 is achieved, the algorithms stop and declare failure.

To calculate the steepest descent direction $v(x^k)$, we solved problem 1-3 using the function "quadprog", a Matlab subroutine that solves quadratic problems with linear constraints.

### 2.6.1 Constants

Let us start the numerical experiments by verifying the influence that constants $\eta$ and $\tau$ have in the performance of NCGMNL. A similar study was done in [40]. Remember that in the section 2.4 the convergence of the Algorithm 2.16 was shown with $\beta_k \leq \eta\beta_k^{FR}$ where $0 < \eta < 1$ see Theorem 2.21 , $\beta_k \leq \eta\beta_k^{CD}$ where $0 < \eta < 1 - \delta$ see Theorem 2.24, $\beta_k = \eta\beta_k^{DY}$ where $0 < \eta < (1-\delta)/(1+\delta)$ see Theorem 2.26, $\beta_k = \beta_k^{mDY}$ see Theorem 2.28.

To verify the influence of these constants, let us consider the problem $SLC2$, see [50]. This example is convex and not much complicated to solve, given by $F : \mathbb{R}^n \to \mathbb{R}^2$

$$F_1(x) = (x_1 - 1)^4 + \sum_{i=2}^{n}(x_i - 1)^2,$$

$$F_2(x) = (x_2 - 1)^4 + \sum_{i=1, i\neq 2}^{n}(x_i + 1)^2.$$

We vary the values of constants $\eta$ and $\tau$ to verify the improvement in method performance. The problem was compiled 200 times with a number of variables equal to 100 and the starting point randomly taken in the range of $[-50, 50]$. The percentage of times the problem was successfully resolved was recorded, that is, the algorithm stopped at a critical point. This information is presented in the tables bellow.

| FR | |
|---|---|
| $\eta$ | % |
| 1 | 68.00 |
| 0.99 | 98.00 |
| 0.98 | 100.00 |

| CD | |
|---|---|
| $\eta$ | % |
| 1 | 98.00 |
| 0.99 | 100.00 |
| $1-\delta$ | 100.00 |

| DY | |
|---|---|
| $\eta$ | % |
| 1 | 80.00 |
| 0.99 | 100.00 |
| $\frac{1-\delta}{1+\delta}$ | 100.00 |

| DYm | |
|---|---|
| $\tau$ | % |
| 1 | 82.00 |
| 1.01 | 100.00 |

**Table 2.1:** *Constants for betas*

The first one of tables 2.1 provides us with information about beta $FR$ and tells us that 68,00% of the problems stopped at a critical point when $\eta = 1$, 98,00% the times the problem was compiled with $\eta = 0.99$ ended successfully and all problems were resolved with $\eta = 0.98$. About $CD$, we reached 100% of success with $\eta = 0.99$ and $\eta = 1 - \delta = 0.999$, similar results were obtained for $DY$ with 100% from success to $\eta = 0.99$ and $\eta = (1-\delta)/(1+\delta) = 0,998$. In turn $mDY$ solved 82% of the problems with $\tau = 1$ and 100% with $\tau = 1.01$.

Using the results presented in Tables 2.1 along with the Convergence Theorems, the numerical experiments were performed with the respective betas presented in the following way, $\beta_k = 0.98\beta_k^{FR}$ for the beta of $FR$, $\beta_k = 0.98\beta_k^{CD}$ for the beta $CD$, $\beta_k = 0.98\beta_k^{DY}$ for the beta of $DY$. For $PRP+$ and $HS+$ we took the betas $\beta_k = \max\{\beta_k^{PRP+}, 0\}$ $\beta_k = \max\{\beta_k^{HS+}, 0\}$ respectively.

### 2.6.2 Numerical Results

The tables below transcribe the information about the problems and their performance against the respective betas. The examples in Table 2.2 are all convex and the ones in Table 2.3 are non-convex. All problems were compiled 300 times, with the starting point taken randomly inside the specified range. The tables are presented in blocks with four lines each. The first column gives us problem information, an acronym to identify it, the bibliographic reference where the problem was found, number of variables, number of objective functions and interval where the starting point was taken randomly. The second column refers to the performance of the problem during numerical tests, the first line informs the percentage of solved problems (%), the second the average of iterations necessary for the algorithm to find a critical point (it), the third presents the average gradient evaluation of the problems where the algorithm reached a critical point (evalg) and the fourth line shows the average time needed for each problem to be solved (time). From the third to the ninth column are presented the performance results in relation to the respective betas presented in section 2.4.

| | | $FR$ | $CD$ | $DY$ | $mDY$ | $PRP+$ | $HS+$ |
|---|---|---|---|---|---|---|---|
| AP1 , [1] | % | 72.33 | 95.33 | 64.00 | 79.67 | 95.00 | 95.00 |
| n = 2 | it | 1241.29 | 1324.12 | 1801.33 | 1492.69 | 669.11 | 669.11 |
| m = 3 | evalg | 3829.96 | 5533.63 | 7400.43 | 6079.48 | 2682.72 | 2682.72 |
| $x^0 \in [-10, 10]$ | time | 2.49 | 2.70 | 3.72 | 3.03 | 1.37 | 1.38 |
| AP2 , [1] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 1 | it | 15.88 | 18.96 | 15.03 | 15.66 | 12.90 | 12.90 |
| m = 2 | evalg | 50.63 | 77.82 | 62.12 | 64.64 | 53.61 | 53.61 |
| $x^0 \in [-100, 100]$ | time | 0.04 | 0.04 | 0.03 | 0.04 | 0.03 | 0.03 |
| AP4 , [1] | % | 69.67 | 92.00 | 59.33 | 76.33 | 95.33 | 95.33 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| n = 3 | it | 1574.85 | 1488.65 | 3073.84 | 1661.93 | 891.77 | 891.77 |
| m = 3 | evalg | 4885.39 | 6234.23 | 12343.30 | 6698.33 | 3574.15 | 3574.15 |
| $x^0 \in [-10, 10]$ | time | 3.11 | 2.98 | 6.12 | 3.35 | 1.81 | 1.81 |
| BK1 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 12.83 | 15.29 | 12.20 | 12.60 | 10.46 | 10.46 |
| m = 2 | evalg | 41.49 | 63.17 | 50.80 | 52.41 | 43.85 | 43.85 |
| $x^0 \in [-5, 10]$ | time | 0.03 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 |
| DGO2 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 1 | it | 60.21 | 97.81 | 69.21 | 70.32 | 93.36 | 93.36 |
| m = 2 | evalg | 182.73 | 392.32 | 279.41 | 283.80 | 374.58 | 374.58 |
| $x^0 \in [-9, 9]$ | time | 0.12 | 0.19 | 0.14 | 0.14 | 0.18 | 0.18 |
| FDS , [20] | % | 100.00 | 99.67 | 85.67 | 100.00 | 100.00 | 100.00 |
| n = 5 | it | 1877.92 | 382.17 | 5314.93 | 1285.97 | 156.41 | 156.41 |
| m = 3 | evalg | 5999.97 | 1618.18 | 21264.70 | 5148.43 | 628.38 | 628.38 |
| $x^0 \in [-2, 2]$ | time | 4.02 | 0.82 | 11.55 | 2.81 | 0.34 | 0.35 |
| IKK1 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 11.43 | 13.57 | 10.89 | 11.34 | 9.57 | 9.57 |
| m = 3 | evalg | 36.87 | 56.07 | 45.35 | 47.14 | 40.07 | 40.07 |
| $x^0 \in [-50, 50]$ | time | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| JOS1 , [35] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 100 | it | 467.90 | 416.07 | 508.19 | 459.64 | 406.87 | 406.87 |
| m = 2 | evalg | 1404.71 | 1664.27 | 2037.77 | 1842.55 | 1627.49 | 1627.49 |
| $x^0 \in [-100, 100]$ | time | 1.75 | 1.56 | 1.94 | 1.76 | 1.56 | 1.58 |
| Lov1 , [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 13.53 | 16.04 | 12.84 | 13.23 | 10.93 | 10.93 |
| m = 2 | evalg | 43.60 | 66.15 | 53.37 | 54.93 | 45.71 | 45.71 |
| $x^0 \in [-10, 10]$ | time | 0.03 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 |
| MGH33 , [43] | % | 46.67 | 46.33 | 45.33 | 47.33 | 46.33 | 46.33 |
| n = 10 | it | 5.52 | 5.50 | 8.27 | 5.51 | 6.12 | 6.12 |
| m = 10 | evalg | 28.09 | 32.33 | 44.06 | 33.16 | 34.80 | 34.80 |
| $x^0 \in [-1, 1]$ | time | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 |
| MHHM2 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 10.06 | 11.79 | 9.44 | 9.79 | 8.19 | 8.19 |
| m = 3 | evalg | 33.18 | 49.15 | 39.74 | 41.15 | 34.77 | 34.77 |
| $x^0 \in [0, 1]$ | time | 0.03 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 |
| MOP7 , [32] | % | 94.00 | 94.33 | 92.67 | 90.67 | 95.33 | 95.33 |
| n = 2 | it | 195.62 | 173.87 | 195.72 | 232.67 | 160.28 | 160.28 |
| m = 3 | evalg | 589.23 | 697.14 | 796.46 | 938.81 | 642.79 | 642.79 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x^0 \in [-400, 400]$ | time | 0.41 | 0.36 | 0.41 | 0.49 | 0.33 | 0.34 |
| PNR , [48] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 158.43 | 47.23 | 433.59 | 99.13 | 22.25 | 22.25 |
| m = 2 | evalg | 516.05 | 199.09 | 1738.16 | 400.11 | 92.15 | 92.15 |
| $x^0 \in [-2, 2]$ | time | 0.32 | 0.10 | 0.86 | 0.20 | 0.05 | 0.05 |
| SD , [52] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 4 | it | 68.56 | 37.11 | 159.38 | 72.24 | 32.84 | 32.84 |
| m = 2 | evalg | 212.13 | 150.37 | 639.62 | 290.96 | 133.12 | 133.12 |
| $x^0 \in [1, 3]$ | time | 0.14 | 0.08 | 0.33 | 0.15 | 0.07 | 0.07 |
| SLC2 , [50] | % | 100.00 | 100.00 | 98.00 | 100.00 | 100.00 | 100.00 |
| n = 100 | it | 625.33 | 80.28 | 814.27 | 564.13 | 55.34 | 54.80 |
| m = 2 | evalg | 2070.22 | 352.46 | 3282.35 | 2281.43 | 242.53 | 239.55 |
| $x^0 \in [-50, 50]$ | time | 2.47 | 0.45 | 3.20 | 2.29 | 0.36 | 0.36 |
| SLCDT2 , [51] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 10 | it | 35.73 | 19.48 | 70.59 | 33.32 | 13.85 | 13.85 |
| m = 3 | evalg | 114.37 | 82.83 | 286.51 | 137.46 | 60.81 | 60.81 |
| $x^0 \in [-1, 1]$ | time | 0.08 | 0.04 | 0.15 | 0.07 | 0.03 | 0.03 |
| SP1 , [32] | % | 98.00 | 97.67 | 98.00 | 98.00 | 98.67 | 98.67 |
| n = 2 | it | 29.83 | 25.81 | 29.01 | 27.73 | 27.44 | 25.68 |
| m = 2 | evalg | 94.55 | 108.07 | 122.13 | 115.86 | 116.38 | 110.29 |
| $x^0 \in [-100, 100]$ | time | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |
| Toi4 , [53] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 4 | it | 16.74 | 19.89 | 16.34 | 16.64 | 13.65 | 13.65 |
| m = 2 | evalg | 53.23 | 81.57 | 67.38 | 68.56 | 56.60 | 56.60 |
| $x^0 \in [-100, 100]$ | time | 0.04 | 0.05 | 0.04 | 0.04 | 0.03 | 0.03 |
| Toi8 , [53] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 9.80 | 9.96 | 12.53 | 12.80 | 10.93 | 10.93 |
| m = 2 | evalg | 34.30 | 43.74 | 54.20 | 55.23 | 47.59 | 47.59 |
| $x^0 \in [-10, 10]$ | time | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| ZLT1 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 10 | it | 20.87 | 24.95 | 19.76 | 20.65 | 16.95 | 16.95 |
| m = 5 | evalg | 65.61 | 101.80 | 81.03 | 84.59 | 69.81 | 69.81 |
| $x^0 \in [-10^2, 10^2]$ | time | 0.06 | 0.06 | 0.05 | 0.05 | 0.04 | 0.05 |

**Table 2.2:** *Convex Problem.*

| | | $FR$ | $CD$ | $DY$ | $mDY$ | $PRP+$ | $HS+$ |
|---|---|---|---|---|---|---|---|
| AP3, [1] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| n = 2 | it | 184.55 | 79.75 | 734.03 | 175.07 | 68.36 | 68.78 |
| m = 2 | evalg | 2148.46 | 607.70 | 12893.24 | 2112.16 | 405.76 | 416.84 |
| $x^0 \in [-10, 10]$ | time | 0.39 | 0.16 | 1.65 | 0.37 | 0.14 | 0.14 |
| DD1 , [10] | % | 100.00 | 100.00 | 76.33 | 100.00 | 100.00 | 100.00 |
| n = 5 | it | 1444.38 | 285.87 | 3680.58 | 1040.55 | 122.37 | 122.62 |
| m = 2 | evalg | 4522.28 | 1208.89 | 14727.48 | 4167.16 | 492.45 | 493.13 |
| $x^0 \in [-20, 20]$ | time | 3.03 | 0.60 | 7.90 | 2.20 | 0.26 | 0.27 |
| DGO1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 1 | it | 11.85 | 11.50 | 13.08 | 13.27 | 16.38 | 16.32 |
| m = 2 | evalg | 70.52 | 71.52 | 92.77 | 92.64 | 97.89 | 97.06 |
| $x^0 \in [-10, 13]$ | time | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 | 0.04 |
| Far1 , [32] | % | 100.00 | 100.00 | 90.33 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 1048.74 | 259.65 | 1909.40 | 760.25 | 131.93 | 139.65 |
| m = 2 | evalg | 14235.73 | 2558.03 | 35520.68 | 10127.15 | 935.19 | 1010.09 |
| $x^0 \in [-1, 1]$ | time | 2.27 | 0.54 | 4.36 | 1.63 | 0.27 | 0.29 |
| FF1 , [32] | % | 100.00 | 100.00 | 99.33 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 496.03 | 112.89 | 1764.70 | 414.83 | 73.80 | 74.09 |
| m = 2 | evalg | 5927.49 | 871.91 | 30065.74 | 5089.50 | 416.62 | 419.68 |
| $x^0 \in [-1, 1]$ | time | 1.04 | 0.23 | 3.92 | 0.87 | 0.15 | 0.15 |
| Hil1 , [31] | % | 100.00 | 100.00 | 99.33 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 290.03 | 71.24 | 697.88 | 166.55 | 33.60 | 41.10 |
| m = 2 | evalg | 4213.39 | 785.99 | 12781.50 | 2290.19 | 287.30 | 359.35 |
| $x^0 \in [0, 1]$ | time | 0.63 | 0.15 | 1.56 | 0.36 | 0.07 | 0.09 |
| KW2 , [36] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 269.41 | 111.22 | 604.98 | 186.61 | 88.70 | 93.23 |
| m = 2 | evalg | 3335.39 | 895.17 | 10861.19 | 2298.04 | 512.92 | 557.53 |
| $x^0 \in [-3, 3]$ | time | 0.57 | 0.23 | 1.36 | 0.40 | 0.18 | 0.19 |
| LE1 , [32] | % | 94.00 | 100.00 | 98.33 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 980.38 | 182.02 | 958.36 | 329.74 | 106.00 | 98.37 |
| m = 2 | evalg | 20652.71 | 2463.15 | 19781.84 | 5597.37 | 953.98 | 771.77 |
| $x^0 \in [-5, 10]$ | time | 2.24 | 0.39 | 2.20 | 0.74 | 0.22 | 0.20 |
| Lov3 , [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 18.31 | 19.48 | 22.64 | 17.65 | 13.24 | 13.30 |
| m = 2 | evalg | 121.36 | 135.96 | 220.88 | 130.50 | 79.52 | 80.09 |
| $x^0 \in [-10^2, 10^2]$ | time | 0.04 | 0.04 | 0.05 | 0.04 | 0.03 | 0.03 |
| Lov4 , [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 19.92 | 19.05 | 25.27 | 19.35 | 13.67 | 13.69 |
| m = 2 | evalg | 139.14 | 134.16 | 247.84 | 150.52 | 83.15 | 83.35 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x^0 \in [-20, 20]$ | time | 0.05 | 0.04 | 0.06 | 0.04 | 0.03 | 0.03 |
| Lov5 , [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 3 | it | 113.04 | 82.42 | 107.45 | 115.25 | 107.16 | 106.88 |
| m = 2 | evalg | 826.89 | 444.94 | 1020.82 | 1024.04 | 524.48 | 521.05 |
| $x^0 \in [-2, 2]$ | time | 0.25 | 0.17 | 0.25 | 0.26 | 0.23 | 0.23 |
| MGH16 , [43] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 4 | it | 100.64 | 67.97 | 152.46 | 111.81 | 89.42 | 83.34 |
| m = 5 | evalg | 730.66 | 378.16 | 2130.58 | 1082.20 | 455.81 | 422.40 |
| $x^0 \in [-10, 10]$ | time | 0.24 | 0.15 | 0.41 | 0.27 | 0.20 | 0.19 |
| MGH26 , [43] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 4 | it | 26.74 | 17.48 | 57.34 | 24.56 | 17.52 | 17.52 |
| m = 4 | evalg | 232.31 | 114.31 | 836.59 | 221.97 | 101.23 | 101.26 |
| $x^0 \in [-1, 1]$ | time | 0.07 | 0.04 | 0.15 | 0.06 | 0.04 | 0.04 |
| MGH33 , [43] | % | 45.33 | 44.33 | 43.67 | 44.67 | 45.00 | 45.00 |
| n = 10 | it | 5.74 | 5.52 | 7.79 | 5.24 | 6.24 | 6.24 |
| m = 10 | evalg | 94.19 | 93.81 | 127.02 | 88.67 | 101.84 | 101.84 |
| $x^0 \in [-1, 1]$ | time | 0.03 | 0.03 | 0.04 | 0.03 | 0.03 | 0.03 |
| MHHM2 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 9.96 | 11.66 | 9.32 | 9.72 | 8.10 | 8.10 |
| m = 3 | evalg | 59.74 | 80.65 | 64.25 | 67.02 | 48.62 | 48.62 |
| $x^0 \in [0, 1]$ | time | 0.03 | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 |
| MLF1 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 1 | it | 7.13 | 8.28 | 6.94 | 7.20 | 7.01 | 7.03 |
| m = 2 | evalg | 47.62 | 58.62 | 50.71 | 52.95 | 43.88 | 43.45 |
| $x^0 \in [0, 20]$ | time | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| MMR1 , [41] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 54.65 | 23.43 | 102.48 | 39.16 | 14.81 | 14.78 |
| m = 2 | evalg | 668.62 | 234.42 | 1619.62 | 471.45 | 120.65 | 120.55 |
| $x^0 \in [0, 1]$ | time | 0.12 | 0.05 | 0.23 | 0.09 | 0.04 | 0.04 |
| MMR3 , [41] | % | 86.00 | 87.67 | 62.67 | 71.00 | 87.67 | 87.67 |
| n = 2 | it | 267.06 | 160.43 | 127.70 | 207.81 | 72.32 | 176.71 |
| m = 2 | evalg | 3257.10 | 1531.05 | 1845.95 | 2967.39 | 403.60 | 1545.29 |
| $x^0 \in [-1, 1]$ | time | 0.57 | 0.33 | 0.28 | 0.45 | 0.15 | 0.37 |
| MOP2 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 76.48 | 30.61 | 146.16 | 56.79 | 26.93 | 27.62 |
| m = 2 | evalg | 797.44 | 181.07 | 2106.53 | 576.62 | 122.69 | 126.32 |
| $x^0 \in [-4, 4]$ | time | 0.16 | 0.06 | 0.31 | 0.12 | 0.06 | 0.06 |
| MOP3 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| n = 2 | it | 44.58 | 23.85 | 81.95 | 36.80 | 19.44 | 20.04 |
| m = 2 | evalg | 457.07 | 199.32 | 1124.60 | 376.75 | 149.67 | 155.63 |
| $x^0 \in [-\pi, \pi]$ | time | 0.11 | 0.06 | 0.20 | 0.09 | 0.05 | 0.05 |
| MOP5 , [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 35.92 | 36.11 | 47.51 | 35.21 | 33.14 | 33.16 |
| m = 3 | evalg | 233.99 | 215.47 | 459.28 | 271.94 | 165.55 | 165.92 |
| $x^0 \in [-1, 1]$ | time | 0.08 | 0.08 | 0.11 | 0.08 | 0.07 | 0.07 |
| SK2 , [1] | % | 100.00 | 100.00 | 99.33 | 100.00 | 100.00 | 100.00 |
| n = 4 | it | 490.05 | 111.26 | 1727.52 | 374.82 | 55.77 | 56.66 |
| m = 2 | evalg | 6029.86 | 919.33 | 30039.56 | 4654.78 | 321.27 | 328.56 |
| $x^0 \in [-10, 10]$ | time | 1.06 | 0.23 | 3.93 | 0.81 | 0.12 | 0.12 |
| SLC1 , [50] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 14.72 | 16.26 | 14.71 | 13.83 | 12.97 | 13.00 |
| m = 2 | evalg | 91.86 | 110.09 | 112.50 | 102.02 | 75.33 | 75.30 |
| $x^0 \in [-5, 5]$ | time | 0.04 | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 |
| SLC2 , [50] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 100 | it | 207.85 | 77.02 | 433.88 | 141.81 | 57.22 | 55.16 |
| m = 2 | evalg | 2442.78 | 605.24 | 7101.08 | 1630.58 | 339.87 | 327.89 |
| $x^0 \in [-50, 50]$ | time | 0.98 | 0.42 | 2.02 | 0.71 | 0.35 | 0.35 |
| SLCDT1 , [51] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 2 | it | 14.76 | 16.32 | 14.15 | 13.88 | 13.10 | 13.12 |
| m = 2 | evalg | 91.80 | 110.24 | 107.49 | 102.69 | 76.00 | 75.96 |
| $x^0 \in [-5, 5]$ | time | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 |
| SLCDT2 , [51] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 10 | it | 37.98 | 20.01 | 78.27 | 35.13 | 14.12 | 14.12 |
| m = 3 | evalg | 349.90 | 156.31 | 1016.48 | 340.23 | 93.29 | 93.29 |
| $x^0 \in [-1, 1]$ | time | 0.10 | 0.05 | 0.21 | 0.09 | 0.04 | 0.04 |
| Toi9 , [53] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| n = 4 | it | 179.62 | 54.54 | 431.79 | 118.67 | 29.49 | 29.25 |
| m = 4 | evalg | 2484.20 | 567.32 | 7850.90 | 1557.57 | 229.34 | 228.15 |
| $x^0 \in [-1, 1]$ | time | 0.44 | 0.13 | 1.10 | 0.29 | 0.07 | 0.07 |
| Toi10 , [53] | % | 57.33 | 93.00 | 46.67 | 69.00 | 100.00 | 99.67 |
| n = 4 | it | 1617.16 | 2447.99 | 939.89 | 1887.43 | 1028.16 | 1051.93 |
| m = 3 | evalg | 28641.15 | 37622.29 | 18628.08 | 33336.35 | 12923.74 | 13206.96 |
| $x^0 \in [-2, 2]$ | time | 3.71 | 5.42 | 2.20 | 4.31 | 2.23 | 2.30 |
| VU1 , [32] | % | 42.33 | 81.67 | 18.00 | 61.00 | 100.00 | 100.00 |
| n = 2 | it | 4606.58 | 2804.37 | 5065.57 | 4139.88 | 2096.12 | 2099.95 |
| m = 2 | evalg | 15537.09 | 11870.94 | 20267.76 | 16598.69 | 8393.74 | 8405.55 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x^0 \in [-3, 3]$ | time | 9.17 | 5.60 | 10.29 | 8.34 | 4.23 | 4.28 |

**Table 2.3:** *Non-Convex Problem.*

We built the Performance Profile [13] in relation to time and iteration. Compared the NCGMNL for each beta for both the convex examples, figure 2.1, as for the non-convex, figure 2.2.

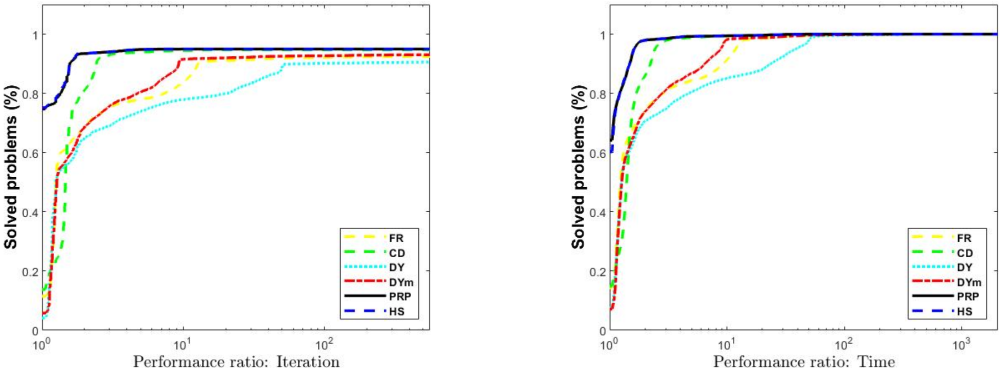



**Figure 2.1:** *Performance Profile-Convex problem*

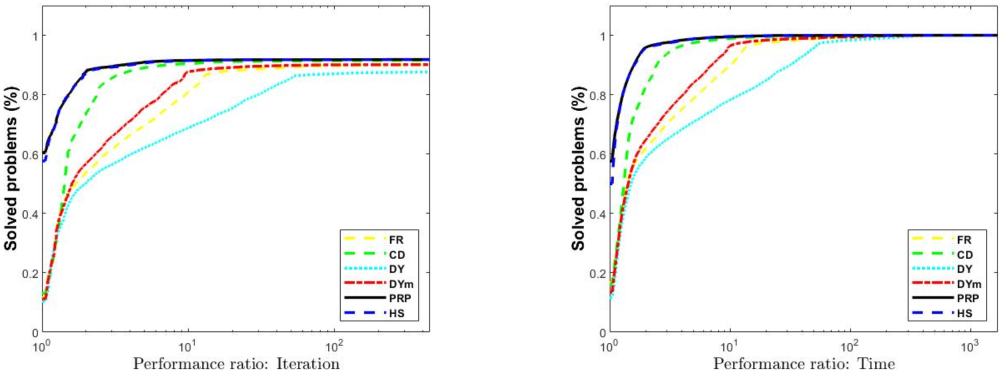



**Figure 2.2:** *Performance Profile-Non-Convex problem*

The graphics of Figures 2.1 and 2.2 show that the Method performs better with PRP and HS betas, both in terms of iteration and time, for both sets of convex and non-convex problems.

From the examples presented in Tables 2.2 and 2.3, we have selected five convex and five non-convex problems to graphically display their respective Pareto Frontier for each beta studied in section 2.4. In Figures 2.3 up to 2.7 the convex examples are shown: BK1, Lov1, PNR, SD and Toi8. From 2.8 to 2.12 we have the non-convex: Far1, Hil1, KW2, Lov4, MOP3.

Each figure presents information about a problem, specified above or in their respective caption, being composed of nine graphics each. The first ones were obtained by discretizing the boxes where the starting point is taken, corresponding through a fine grid and plotting all the points of the image. This figure gives us the representation of the image of F and, in turn, a geometric idea of the Pareto frontiers.

Second graphics were obtained by running Algorithm 2.16 for each problem 300 times, using randomly generated starting points belonging to the corresponding boxes. The starting point image is represented by the blue dot on this graph and the respective image of the critical point obtained by the algorithm is the black point, the gray line connects these two points.

The third graph was obtained similarly to the second, however running the problem only 20 times. The starting point image is represented by a blue asterisk, and the image of each iteration is represented by a blue dot, with each iteration being linked to the next by a gray line segment. The solution image is represented by a black dot.

The last six graphs are images of the critical points obtained by NCGMNL using the respective betas FR, CD, DY, DYm, PRP and HS. That is, the respective Pareto Fronts.

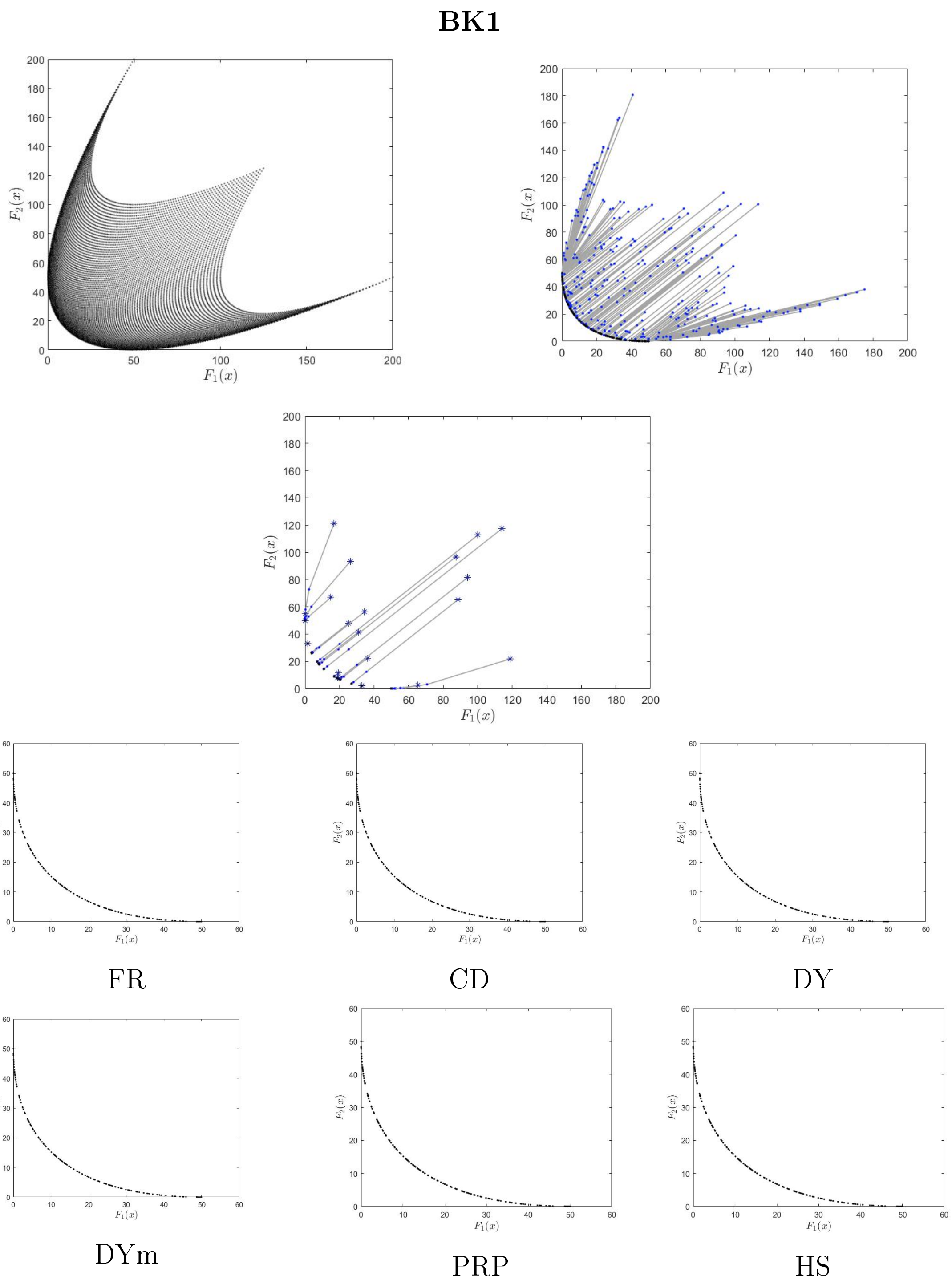



**Figure 2.3:** *Convex problem-BK1*

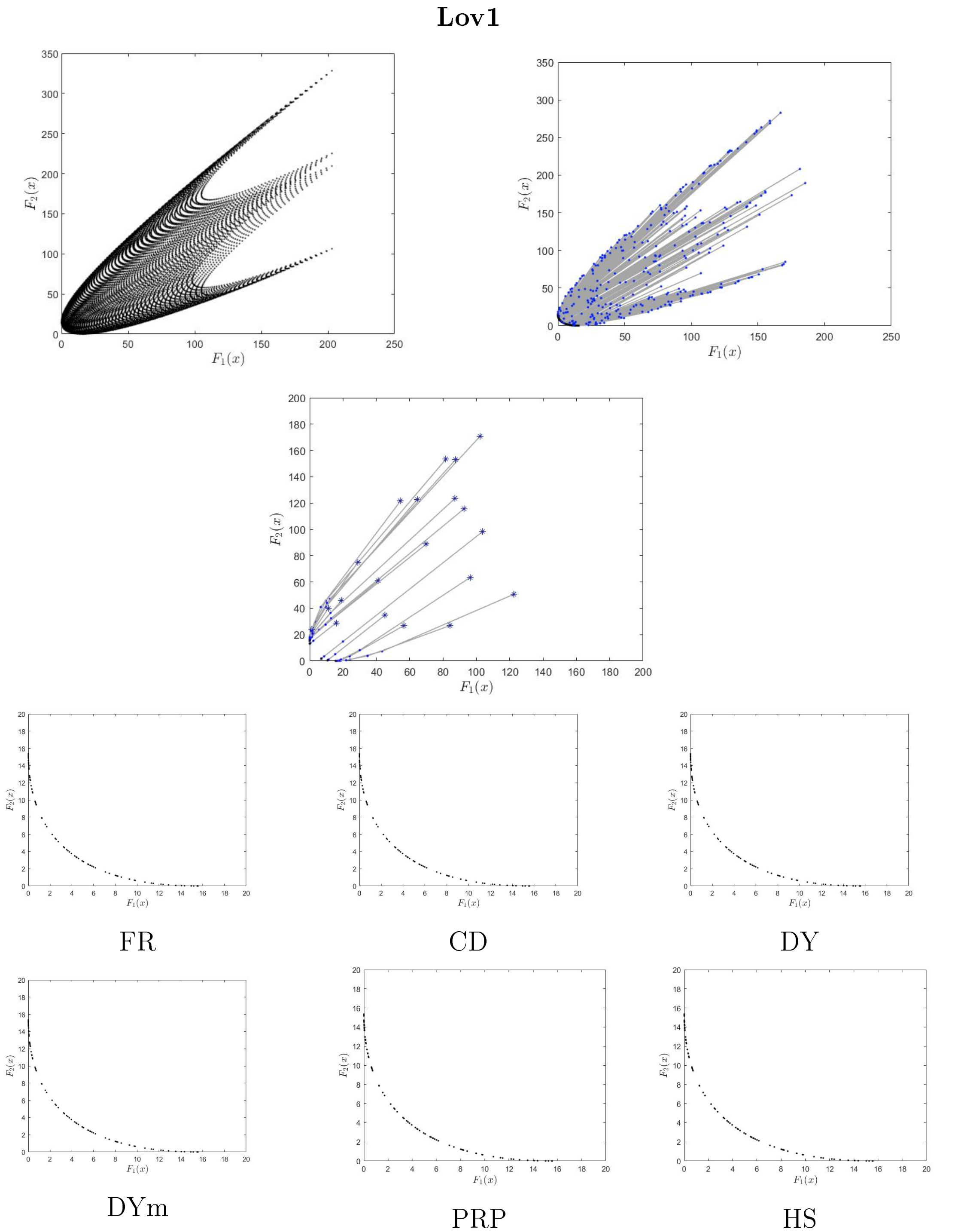



**Figure 2.4:** *Convex problem-Lov1*

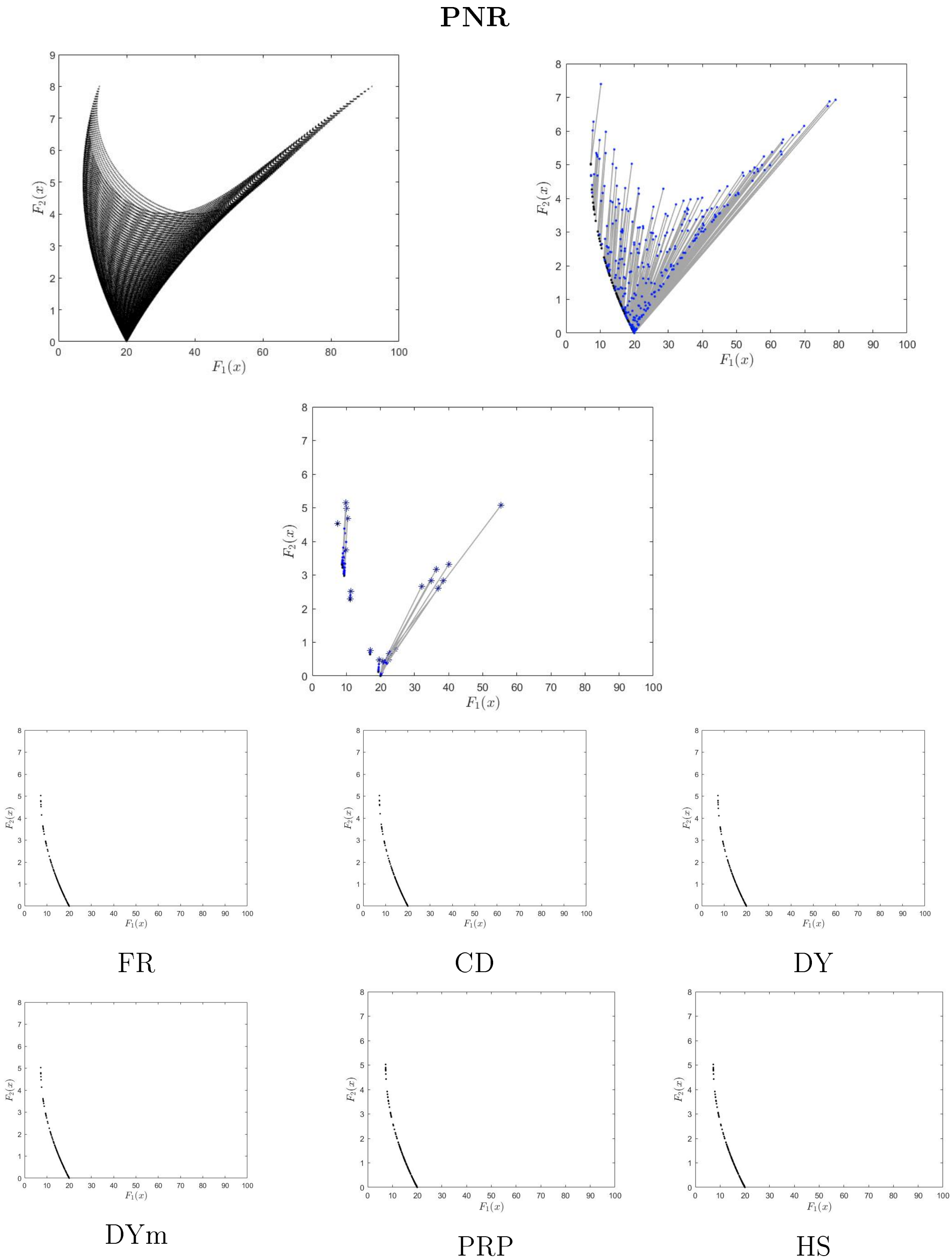



**Figure 2.5:** *Convex problem-PNR*

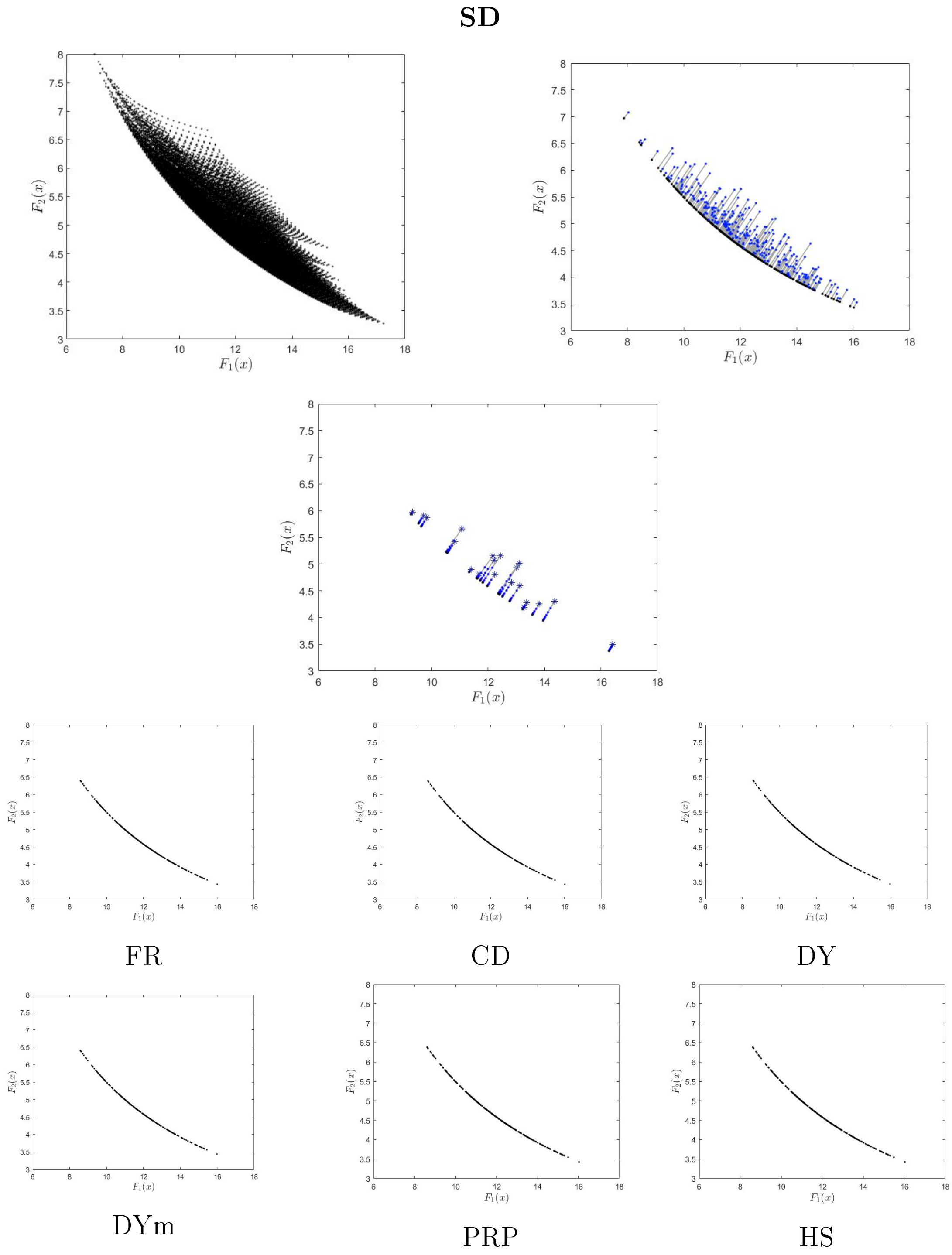



**Figure 2.6:** *Convex problem-SD*

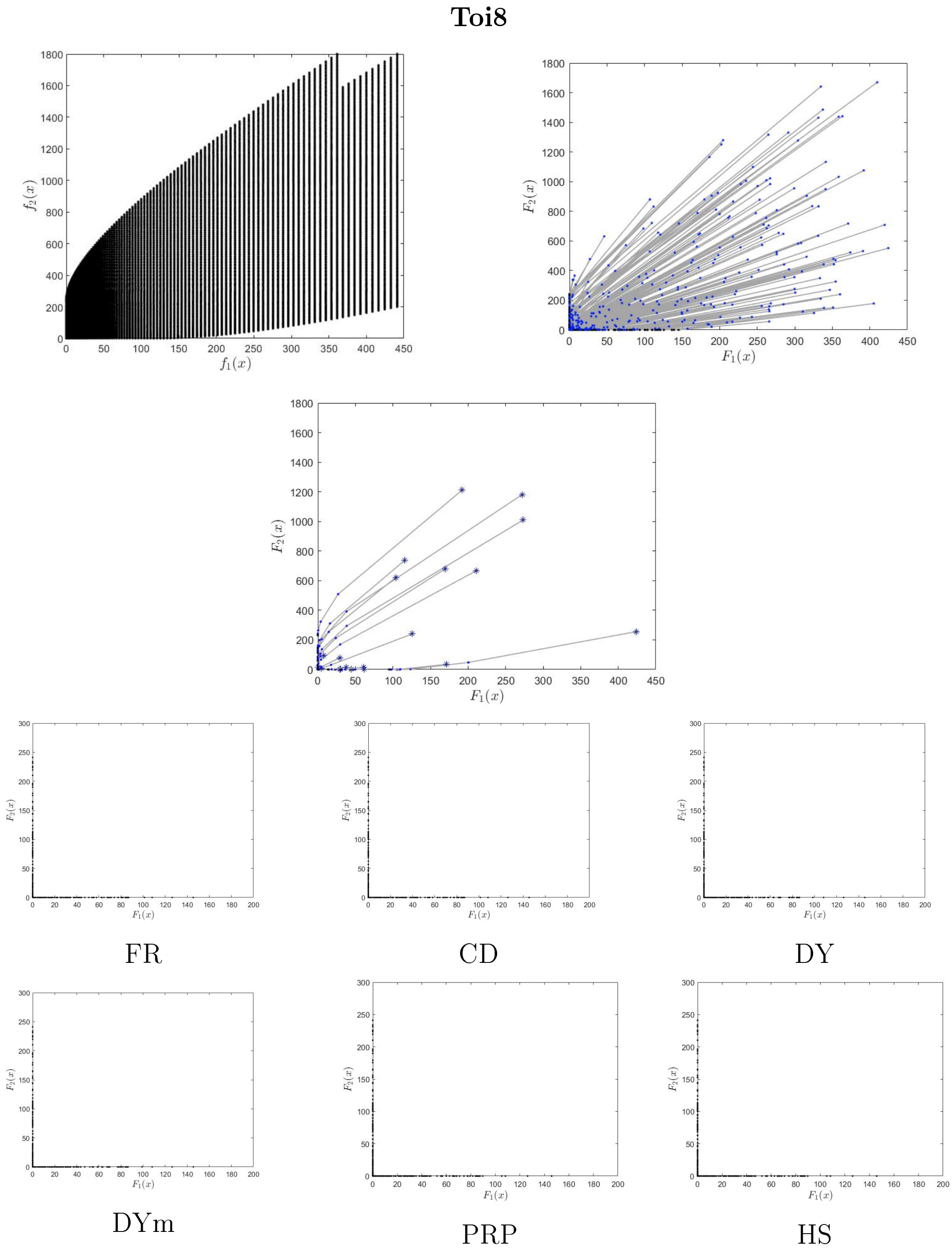



**Figure 2.7:** *Convex problem-Toi8*

**Far1**

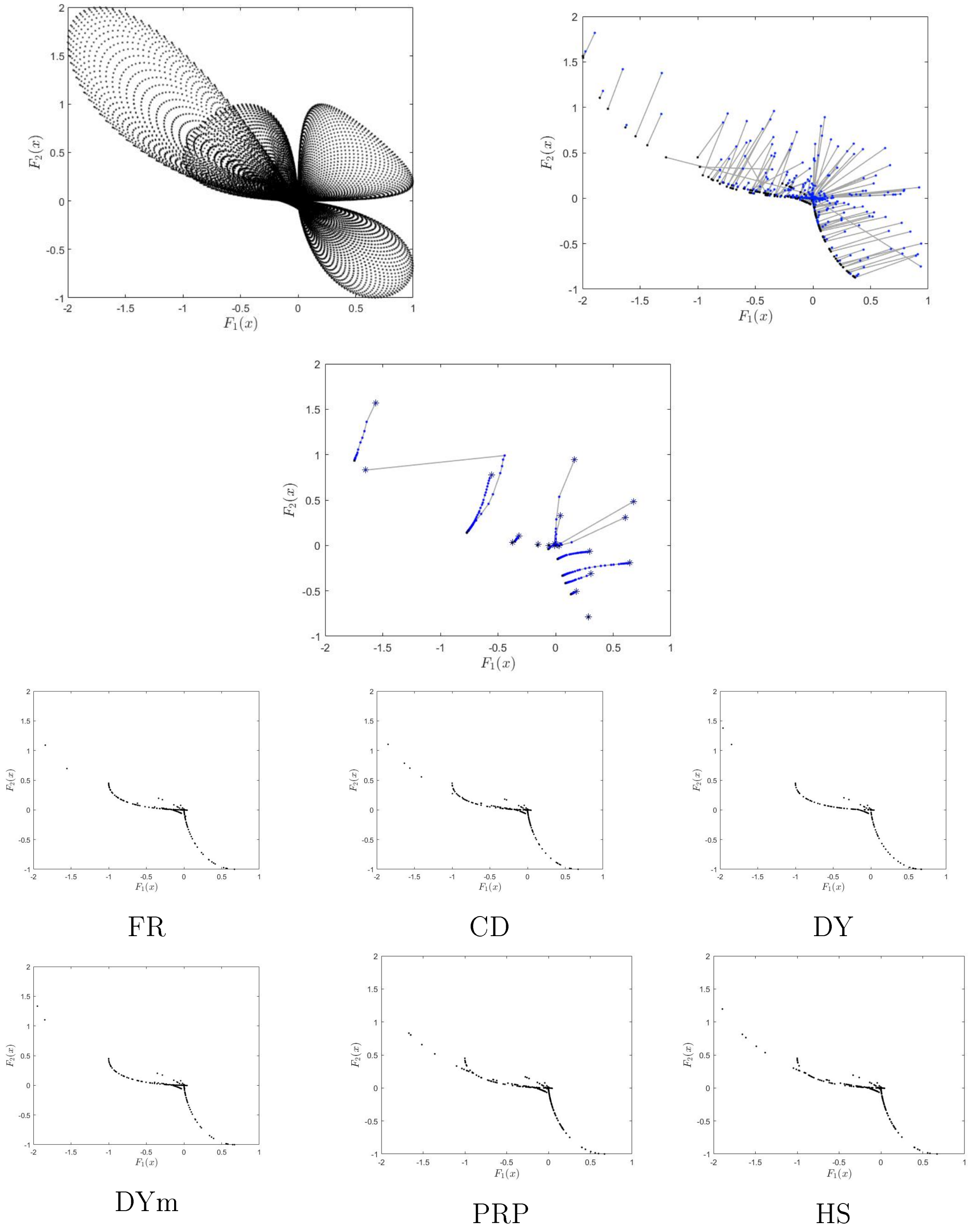



**Figure 2.8:** *Non-Convex problem-Far1*

**Hil1**

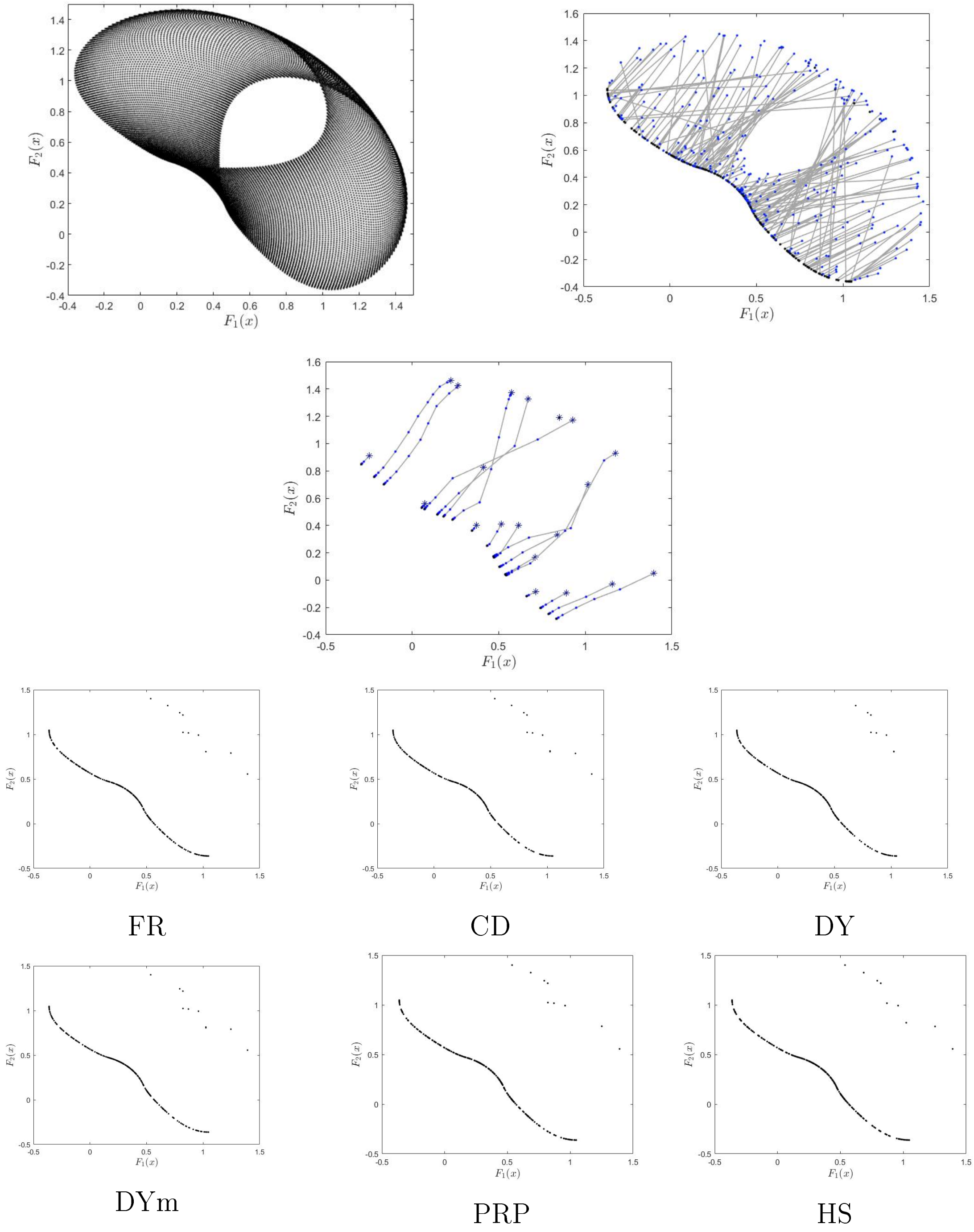



**Figure 2.9:** *Non-Convex problem-Hil1*

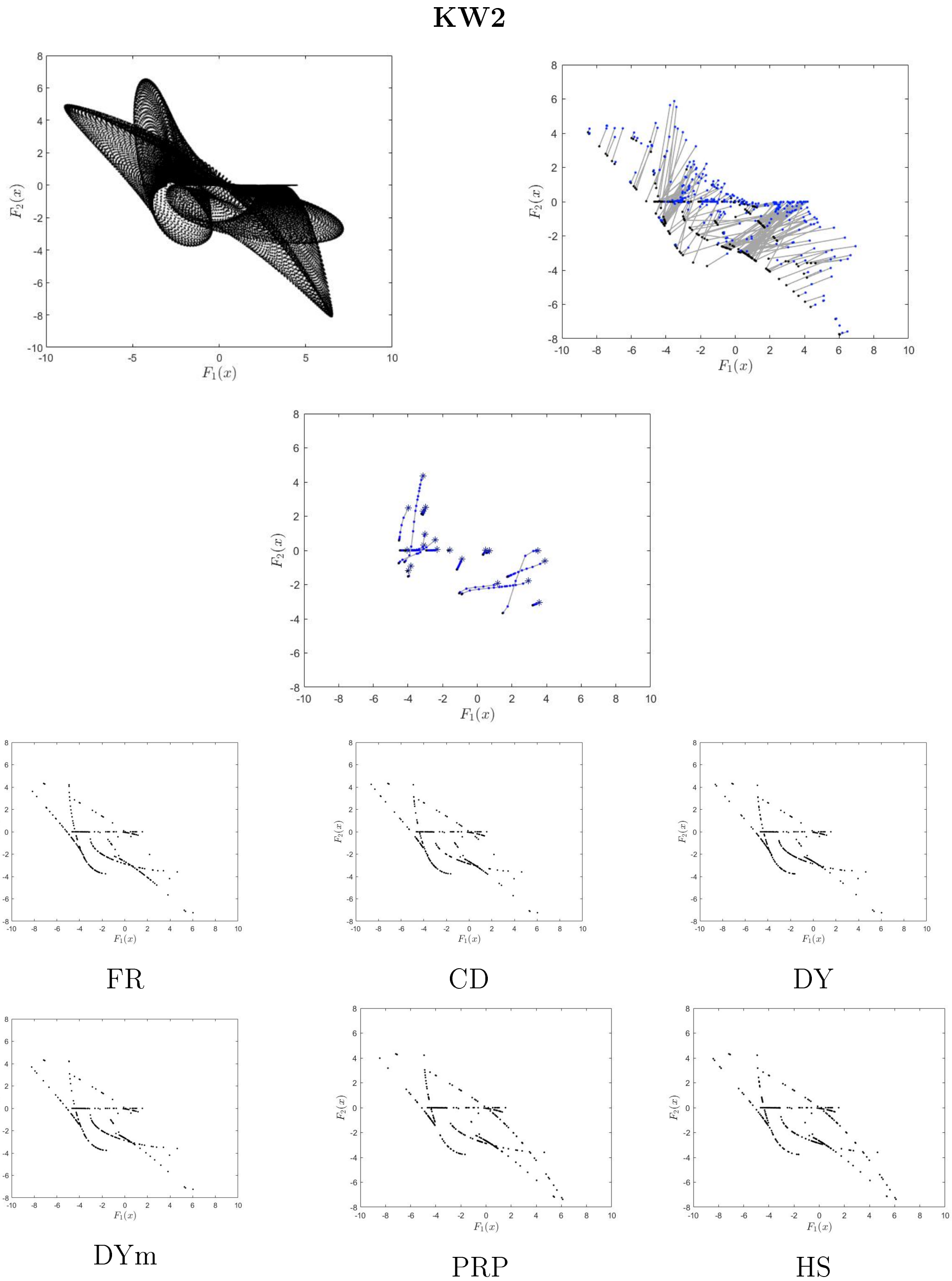



**Figure 2.10:** *Non-Convex problem-KW2*

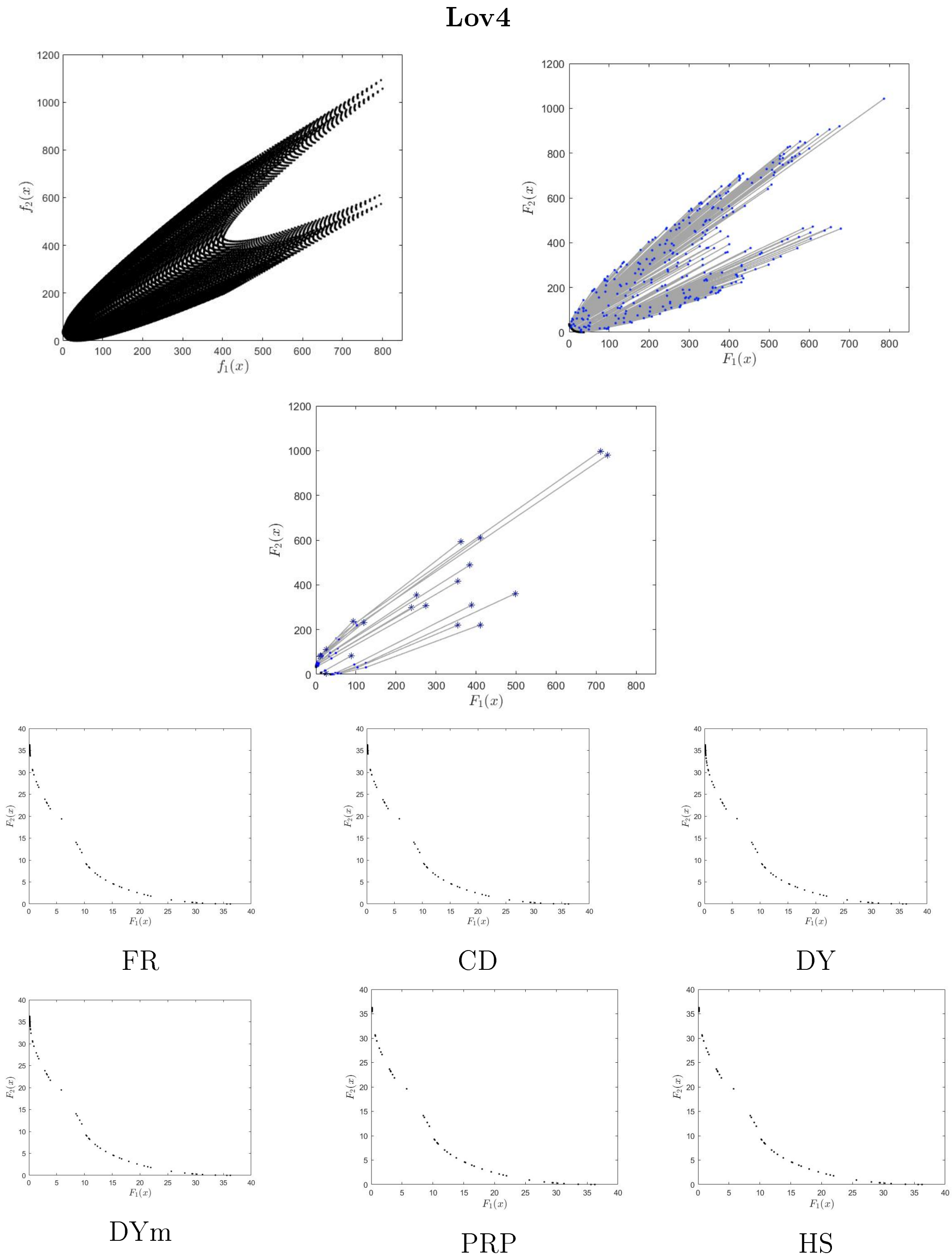



**Figure 2.11:** *Non-Convex problem-Lov4*

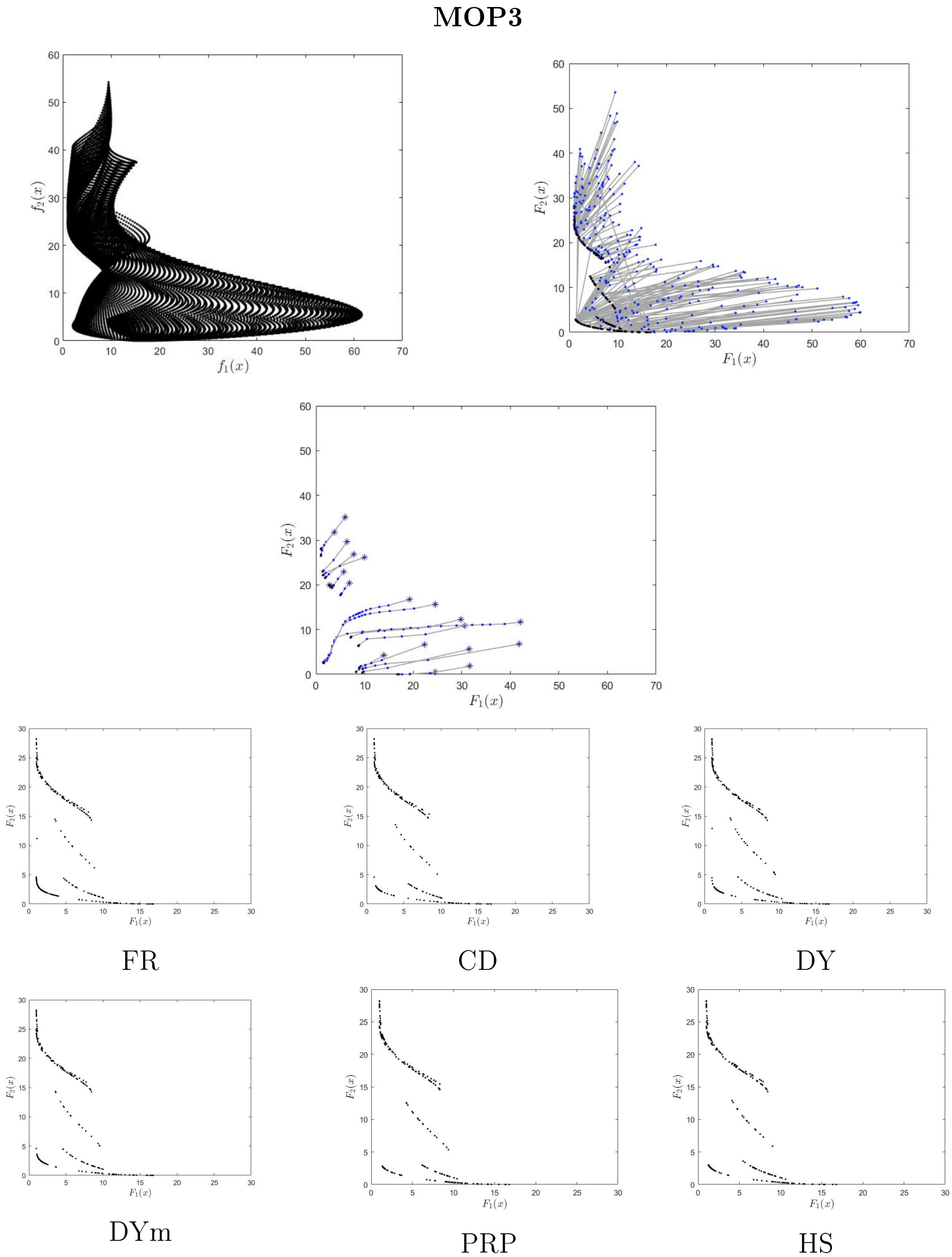



**Figure 2.12:** *Non-Convex problem-MOP3*

CHAPTER 3

# Steepest descent method with a new line search

In this chapter, we will study the gradient method when the Armijo search is replaced by the new line search introduced early. A family of conjugate gradient methods, with this new line search, was studied in Chapter 2. Theoretical results of convergence and numerical performance were satisfactory when compared with the results presented in [39]. Therefore, we decided to proceed with a similar study for the steepest descent algorithm. As was commented before, the new line search does not make use of function values. So, it may have good performance when the objectives are functions for which the is simpler to compute the gradient.

## 3.1 Modifying the new line search

The new line search was presented in Chapter 2. In our studies was necessary to introduce some hypotheses to guarantee that the sequence $\nu_k$ remains bounded. Now, we make a modification that assures that $\nu_k$ is bounded.

Let $\nu\colon \mathbb{R}^n \times \mathbb{R}^n \setminus \{0\} \times (0,1) \times (0,\infty) \to [0,\infty)$ and $i\colon \mathbb{R}^n \times \mathbb{R}^n \setminus \{0\} \times \mathbb{R}_+ \times (0,1) \times (0,\infty) \times (0,1) \to \mathbb{Z}_+$ be defined as

$$\nu(x,d,\delta,\rho) = \max\left\{0\,,\, 2\frac{\delta f(x,d) - f(x+\rho d, d)}{\rho\|d\|^2}\right\} \tag{3-1}$$

and

$$i(x,d,\nu,\delta,\rho,\omega) = \min\left\{i \geq 0 \;\mid\; f(x+\rho\omega^i d, d) + \frac{\nu\rho\omega^i\|d\|^2}{2} < \delta f(x,d)\right\}, \tag{3-2}$$

respectively. For some combinations of $x$ and $d$, which are of our interest, $i$ is well-defined.

**Lemma 3.1.** *If $d$ is $\mathcal{K}$-descent direction for $F$ at $x$ then, $i$ is well defined.*

*Proof.* The proof is by contradiction. Remember that $\nu$ and $\rho > 0$. Assume that

$$f(x + \rho\omega^i d, d) + \frac{\nu\rho\omega^i\|d\|^2}{2} \geq \delta f(x, d),$$

for all positive integer $i$. Then, considering the limit as $i \to \infty$, we get

$$f(x, d) \geq \delta f(x, d)$$

since $\omega \in (0, 1)$ and $f(\cdot, d)$ is continuous. The inequality above is impossible because $f(x, d) < 0$ and $\delta \in (0, 1)$. Thus, minimum at (3-2) must exist. □

**Lemma 3.2.** *Assume $L$ is Lipschitz constant of the Jacobian JF. Then,* $\frac{\nu(x, v(x), \delta, \rho)}{2} < L + \frac{1-\delta}{\rho}$.

*Proof.* As observed in (1-2),

$$f(x, v(x)) + \|v(x)\|^2 = 0.$$

Let $x$ be not $\mathcal{K}$-critical. By definition (3-1) and Lemma 1.11

$$\begin{aligned}
\frac{\nu(x, v(x), \delta, \rho)}{2} &\leq \frac{\delta f(x, v(x)) - f(x + \rho v(x), v(x))}{\rho\|v(x)\|^2} \\
&= \frac{(\delta - 1)f(x, v(x)) + [f(x, v(x)) - f(x + \rho v(x), v(x))]}{\rho\|v(x)\|^2} \\
&\leq \frac{\delta - 1}{\rho}\frac{f(x, v(x))}{\|v(x)\|^2} + L \\
&= \frac{1 - \delta}{\rho} + L.
\end{aligned}$$

□

Next, we will propose a procedure for computing a critical point of Problem 0-3. This method is based on the steepest descent algorithm proposed in [21, 28].

## 3.2 Convex case

For $\mathcal{K}$-convex and continuously differentiable $F$, our algorithm is the following.

**Algorithm 3.3.** *Consider three exogenous constants:* $0 < \rho, \omega, \delta < 1$

*0.* **Initialization***: Let it be* $x^0 \in \mathbb{R}^n$*. Compute* $v(x^0)$*, and initialize* $k \leftarrow 0$*.*

1. **Stopping criterium**: *If* $v(x^k) = 0$, *then STOP.*
2. **Line search**: *Compute*

$$\imath_k = \min\{\imath \geq 1 \mid f(x^k + \rho\omega^\imath v(x^k), v(x^k)) \leq \delta f(x^k, v(x^k))\}. \tag{3-3}$$

3. **Iteration step**: *Define*

$$\alpha_k = \rho\omega^{\imath_k} \tag{3-4}$$

*and*

$$x^{k+1} = x^k + \alpha_k v(x^k). \tag{3-5}$$

*Compute* $v(x^{k+1})$, *set* $k \leftarrow k+1$, *and go to Step 1.*

Herafter, $\{x^k\}$ refers to the sequence generated by Algorithm 3.3. Again, if there exists $k$ with $v(x^k) = 0$, our procedure stops successfully. Let us then assume $v(x^k) \neq 0$ for all $k$. Henceforth, $f(x^k, v(x^k)) < 0$ for all $k$.

**Lemma 3.4.** *The sequence* $\{F(x^k)\}$ *is strictly* $\mathcal{K}$*-decreasing, i.e.,* $F(x^{k+1}) < F(x^k)$ *for all* $k$.

*Proof.* Observe that

$$F(x^{k+1}) = F(x^k) + \int_0^{\alpha_k} JF(x^k + tv(x^k))v(x^k)dt.$$

Then, considering $w \in G$,

$$\begin{aligned}
\langle F(x^{k+1})\,,\, w\rangle &= \langle F(x^k)\,,\, w\rangle + \langle \int_0^{\alpha_k} JF(x^k + tv(x^k))v(x^k)dt\,,\, w\rangle \\
&= \langle F(x^k)\,,\, w\rangle + \int_0^{\alpha_k} \langle JF(x^k + tv(x^k))v(x^k)\,,\, w\rangle dt \\
&\leq \langle F(x^k)\,,\, w\rangle + \int_0^{\alpha_k} f(x^k + tv(x^k), v(x^k))dt \\
&\leq \langle F(x^k)\,,\, w\rangle + \int_0^{\alpha_k} f(x^k + \alpha_k v(x^k), v(x^k))dt \\
&= \langle F(x^k)\,,\, w\rangle + \alpha_k f(x^k + \alpha_k v(x^k), v(x^k)) \\
&\leq \langle F(x^k)\,,\, w\rangle + \alpha_k \delta f(x^k, v(x^k)) \\
&< \langle F(x^k)\,,\, w\rangle.
\end{aligned}$$

The first inequality above is validated by $f$'s definition, the second is a consequence of $f(x^k + tv(x^k), v(x^k))$'s monotonicity, Lemma 1.15 item (b), the third is true by (3-3), and $f(x^k, v(x^k)) < 0$ implies the fourth. Then,

$$F(x^{k+1}) < F(x^k).$$

□

**Lemma 3.5.** *If there exists* $\mathcal{F} \preceq_{\mathcal{K}} F(x^k)$ *for all* $k$*, then* $\lim_{k\to\infty} \alpha_k f(x^k, v(x^k)) = 0$.

*Proof.* Consider $e \in \mathcal{K}$ such that $0 < \langle e\,,\, w\rangle \le 1$ for all $w \in C$. In the proof of the previous lemma we showed that

$$F(x^{k+1}) \preceq_K F(x^k) + \alpha_k \delta f(x^k, v(x^k))e.$$

Therefore,

$$\mathcal{F} \preceq_K F(x^{k+1}) \preceq_K F(x^k) + \alpha_k \delta f(x^k, v(x^k))e \preceq_K F(x^0) + \delta[\sum_{s=0}^{k} \alpha_s f(x^s, v(x^s))]e, \tag{3-6}$$

for all $k$, series $\sum \alpha_k f(x^k, v(x^k))$ is summable, and $\lim_{k\to\infty} \alpha_k f(x^k, v(x^k)) = 0$. □

**Lemma 3.6.** *If*

$$\mathcal{T} = \{x \in \mathbb{R}^n \mid F(x) \preceq_K F(x^k) \;\; \textit{for all } k\} \neq \emptyset,$$

*then it exists* $x^* \in \mathcal{T}$ *such that* $\lim_{k\to\infty} x^k = x^*$.

*Proof.* The previous Lemma remains holds. Then, $\sum \alpha_k f(x^k, v(x^k))$ is summable. By Lemma 1.10 (b), $f(x^k, v(x^k)) < -\|v(x^k)\|^2/2$. Therefore, $\sum \alpha_k \|v(x^k)\|^2 < \infty$. Take $\hat{x} \in \mathcal{T}$. By Lemma 1.15 (a), it is true that $\langle x^k - \hat{x}, v(x^k)\rangle \le 0$, hence

$$\|x^k + \alpha_k v(x^k) - \hat{x}\| = \|x^{k+1} - \hat{x}\|^2 \le \|x^k - \hat{x}\|^2 + \alpha_k \|v(x^k)\|^2$$

because $0 < \alpha_k < 1$. Observe that $\mathcal{T}$ is convex, then $\{x^k\}$ is quase-Fèjer convergent to $\mathcal{T}$. Therefore, $\{x^k\}$ is bounded. Consider $x^*$, a accumulation point of $\{x^k\}$. Since $\{F(x^k)\}$ is monotone decreasing (see Lemma 3.4), $x^* \in \mathcal{T}$. Then,

$$\lim_{k\to\infty} x^k = x^*.$$

□

**Theorem 3.7.** *Assume* $\mathcal{T} \neq \emptyset$*. The sequence generated by Algorithm 3.3,* $\{x^k\}$*, is convergent to a weak-*$\mathcal{K}$*-minimum.*

*Proof.* The last two lemmas hold. Then, $\{x^k\}$ is convergent. Consequently, $\alpha_k \|v(x^k)\| \to 0$ as $k \to \infty$. Assume that $x^* = \lim_{k\to\infty} x^k$. We have two cases to analyze because $\{\alpha_k\} \subset (0\,,\,\rho]$.

- First, assume that $\lim_{k\to\infty} \alpha_k = 0$. In this case, given $k$, there exists $\tilde{\alpha}_k \in [\alpha_k\,,\, \omega^{-1}\alpha_k)$ such that

$$f(x^k + \tilde{\alpha_k} v(x^k)\,,\, v(x^k)) = \delta f(x^k, v(x^k)).$$

Considering the limit as $k \to \infty$, we get $f(x^*, v(x^*)) = \delta f(x^*, v(x^*))$. Then, $f(x^*, v(x^*)) = 0$ because $\delta \neq 1$.

- Now, assume that $\liminf \alpha_k = 2\alpha > 0$. Then, for all given $\varepsilon > 0$, there exists $\kappa$ such that $k > \kappa$ implies

$$\varepsilon > \alpha_k \|v(x^k)\| > \alpha \|v(x^k)\|.$$

Hence, $v(x^*) = 0$.

In both cases, $x^*$ is $\mathcal{K}$-critical. Then, by the $\mathcal{K}$-convexity of $F$, $x^*$ is a weak-$\mathcal{K}$-minimum for function $F$. □

### 3.2.1 Rate of convergence

In this section we will derive a convergence rate for Algorithm 3.3. Let it be $\lim_{k\to\infty} x^k = x^*$. We assume that $L$ is the Lipschitz constant of the Jacobian $JF$ and that the parameters of the algorithm are such that $\frac{\rho}{1-\delta} < \frac{1}{2L}$.

**Lemma 3.8.** *Assume that*

$$\frac{\rho}{1-\delta} < \frac{1}{2L}.$$

*Then, the following statement is true for any* $\omega \in (0,1)$*:*

$$\alpha_k = \omega\rho, \quad k = 0, 1, 2, \ldots,$$

*and, consequently* $\alpha_k < \frac{1}{2L}$, $\quad k = 0, 1, 2, \ldots.$

*Proof.* Assume that there exists $k$ such that $\alpha_k < \omega\rho$. Then,

$$f(x^k + w^{-1}\alpha_k v(x^k), v(x^k)) > \delta f(x^k, v(x^k)).$$

Then, by Lemma 1.11 and the above inequality,

$$\begin{aligned} L\omega^{-1}\alpha_k \|v(x^k)\|^2 &\geq f(x^k + w^{-1}\alpha_k v(x^k), v(x^k)) - f(x^k, v(x^k)) \\ &> (\delta - 1) f(x^k, v(x^k)). \end{aligned}$$

But, by Lemma 1.10 (b),

$$\frac{f(x^k, v(x^k))}{\|v(x^k)\|^2} < -\frac{1}{2},$$

implying

$$\frac{\omega(1-\delta)}{2L} < \frac{\omega(\delta-1)}{L}\frac{f(x^k, v(x^k))}{\|v(x^k)\|^2} < \alpha_k < \omega\rho,$$

contradicting the hypothesis. Therefore $\alpha_k = \omega\rho$, for all $k$.

Hence,

$$\alpha_k = \omega\rho < \rho < \frac{1-\delta}{2L} < \frac{1}{2L}.$$

□

Set $f_n^* = \min_{0\le k\le n} |f(x^k, v(x^k))|$. Following the ideas from Section 1.2.3 in [43], we get

$$(n+1)\delta\alpha_k f_n^* \le \delta\alpha_k \sum_{k=0}^{n} |f(x^k, v(x^k))| = -\delta \sum_{k=0}^{n} \alpha_k f(x^k, v(x^k)).$$

Take $e \in \operatorname{int}(\mathcal{K})$ such that $\max\{\langle e, w\rangle : w \in C\} = 1$. Substituting in (3-6), we get

$$\langle F(x^0) - F(x^*)\,,\, w\rangle \ge -\delta[\sum_{k=0}^{n} \alpha_k f(x^k, v(x^k))]\langle e, w\rangle \ge (n+1)\delta\rho\omega f_n^*\langle e, w\rangle$$

for all $w \in C$. Remember that $\|\omega\| = 1$. Therefore, using the Cauchy-Schwarz Inequality, we obtain the inequality

$$f_n^* \le \frac{1}{n+1}\frac{1}{\delta\rho\omega}\|F(x^0) - F(x^*)\|,$$

which describes the convergence rate of $\{f_n^*\}$ to zero.

## 3.3 Lipschitz case

In this section, we will show some properties the line-search has when the Jacobian is Lipschitz continuous. We start with the following auxiliary lemma.

**Lemma 3.9.** *Let $d$ be $K$-descent direction for $F$ at $x$, $L > 0$ Lipschitz constant of the Jacobian JF, $\delta \in (0,1)$ and $\rho > 0$.*

$$\textit{If} \quad L < \frac{(\delta-1)f(x,d)}{\rho\|d\|^2}, \quad \textit{then} \quad f(x+\alpha d, d) \le \delta f(x,d) \quad \textit{for all} \quad 0 \le \alpha \le \rho.$$

*Proof.* The thesis is true at $\alpha = 0$ because $f(x,d) < 0$. Assume that there exists $0 < \tilde{\alpha} \le \rho$ such that $f(x+\tilde{\alpha}d, d) > \delta f(x,d)$. Then, taking in account Lemma 1.11, we get

$$(\delta - 1)f(x,d) < f(x+\tilde{\alpha}d, d) - f(x,d) \leq L\tilde{\alpha}\|d\|^2 \leq L\rho\|d\|^2,$$

in contradiction with our hypothese. □

**Corollary 3.10.** *Assume that the hypothesis of Lemma 3.9 holds and let it be $x' = x + \alpha d$, where $\alpha \in (0, \rho)$. Then,*

$$F(x') \preceq_K F(x) + \delta\alpha f(x,d)e,$$

*for any $e \in int(\mathcal{K})$.*

*Proof.* By the Fundamental Theorem of Calculus,

$$F(x') = F(x) + \int_0^\alpha JF(x+td)d\ dt.$$

Then, taking any $w \in C$, it holds

$$\begin{aligned}
\langle F(x')\,,\,w\rangle &= \langle F(x)\,,\,w\rangle + \langle \int_0^\alpha JF(x+td)d\ dt\,,\,w\rangle \\
&= \langle F(x)\,,\,w\rangle + \int_0^\alpha \langle JF(x+td)d\,,\,w\rangle dt \\
&\leq \langle F(x)\,,\,w\rangle + \int_0^\alpha f(x+td,d)dt \\
&\leq \langle F(x)\,,\,w\rangle + \delta\int_0^\alpha f(x,d)dt \\
&= \langle F(x)\,,\,w\rangle + \alpha\delta f(x,d).
\end{aligned}$$

The first inequality above is validated by $f$'s definition, the second is a consequence of Lemma 3.9 because $\alpha < \rho$. Hence,

$$F(x') \preceq_\mathcal{K} F(x) + \delta\alpha f(x,d)e. \tag{3-7}$$

□

Observe that (3-7) implies in $F(x') \prec_\mathcal{K} F(x)$ because $f(x,d) < 0$.

For the sake of simplification, we introduce function LS: $\mathbb{R}^n \times \mathbb{R}^n \setminus \{0\} \times (0,1) \times (0,\infty) \times (0,1) \to \mathbb{R}^n$, which is defined as follows:

$$\mathrm{LS}(x,d,\delta,\rho,\omega) = x + \rho\omega^{\imath(x,d,\nu(x,d,\delta,\rho),\delta,\rho,\omega)}d, \tag{3-8}$$

where functions $\nu$ and $\imath$ were defined by expressions (3-1) and (3-2), respectively. The steepest descent algorithm with the line-search LS is the following.

**Algorithm 3.11.** *Let it be $\delta, \omega \in (0,1)$, $\rho > 0$ and $x^0 \in \mathbb{R}^n$.*

0. **Initialization**: *Compute $v(x^0)$, and initialize $k \leftarrow 0$.*
1. **Stopping criterium**: *If $v(x^k) = 0$, then STOP, else go to Step 2.*
2. **Iteration step**: *Compute $x^{k+1} = LS(x^k, v(x^k), \delta, \rho, \omega)$ and $v(x^{k+1})$. Set $k \leftarrow k+1$, and go to Step 1.*

If $v(x^k) \neq 0$, then $v(x^k)$ is a descent direction for $F$ at $x^k$ and, therefore, Step 2 will return a new iterate $x^{k+1}$ - see Lemma 3.1. Algorithm 3.11 stops at iteration $k$ only if $v(x^k) = 0$. In such a case, $\{x^k\}$ is finite and the last iterate is $K$-critical. Let us consider the case when $\{x^k\}$ is infinite, i.e., $v(x^k) \neq 0$ for all $k$. In what follows, numbers $\nu_k$ and $\alpha_k$ will denote $\nu(x^k, v(x^k), \delta, \rho)$ and $\rho\omega^{i(x^k, v(x^k), \nu(x^k, v(x^k), \delta, \rho), \delta, \rho, \omega)}$, respectively.

As observed before,

$$\frac{f(x, v(x))}{\|v(x)\|^2} < -\frac{1}{2}$$

when $x$ is not $K$-critical point for $F$. Then, in such a case,

$$\frac{\delta - 1}{\rho} \frac{f(x, v(x))}{\|v(x)\|^2} > -\frac{\delta - 1}{2\rho} = \frac{1 - \delta}{2\rho} > L.$$

In other words, if the parameters of Algorithm 3.11 are such that $\frac{\rho}{1-\delta} < \frac{1}{2L}$, Lemma 3.9 and Corollary 3.10 hold. Hence, $\{F(x^k)\}$ is $\mathcal{K}$-monotone decreasing.

**Theorem 3.12.** *Assume that $L$ is Lipschitz constant of the Jacobian; that $L < \frac{1-\delta}{2\rho}$, and that there exists $\mathcal{F} \preceq_K F(x^k)$. Then, every accumulation point of $\{x^k\}$, if any, is $\mathcal{K}$-critical.*

*Proof.* Observe that

$$\begin{aligned} \mathcal{F} \preceq_{\mathcal{K}} F(x^{k+1}) &\preceq_{\mathcal{K}} F(x^k) + \alpha_k \delta f(x^k, v(x^k)) e \\ &\preceq_{\mathcal{K}} F(x^0) + \delta [\sum_{s=0}^{k} \alpha_s f(x^s, v(x^s))] e, \quad \text{for all } k. \end{aligned} \tag{3-9}$$

Then, series $\sum \alpha_k f(x^k, v(x^k))$ is summable and $\lim_{k\to\infty} \alpha_k f(x^k, v(x^k)) = \lim_{k\to\infty} \alpha_k \|v(x^k)\| = 0$. Let it be $\{x^{k_\ell}\}$, subsequence of $\{x^k\}$, convergent to $x^*$. By continuity of $v$, $\lim_{k\to\infty} v(x^{k_\ell}) = v(x^*)$. Our goal is to demonstrate that $f(x^*, v(x^*)) = 0$. By definition of $\nu_k$,

$$f(x^{k_\ell} + \rho v(x^{k_\ell}), v(x^{k_\ell})) + \frac{\nu_{k_\ell} \rho \|v(x^{k_\ell})\|^2}{2} \geq \delta f(x^{k_\ell}, v(x^{k_\ell})).$$

Since $f(.,v(x^{k_\ell}))$ is continuous, there exists $\tilde{\alpha}_{k_\ell} \in (\alpha_{k_\ell}, \omega^{-1}\alpha_{k_\ell}]$, such that

$$f(x^{k_\ell} + \tilde{\alpha}_{k_\ell} v(x^{k_\ell}), v(x^{k_\ell})) + \frac{\nu_{k_\ell}\tilde{\alpha}_{k_\ell}\|v(x^{k_\ell})\|^2}{2} = \delta f(x^{k_\ell}, v(x^{k_\ell})) \tag{3-10}$$

because, by (3-2),

$$f(x^{k_\ell} + \alpha_{k_\ell} v(x^{k_\ell}), v(x^{k_\ell})) + \frac{\nu_{k_\ell}\alpha_{k_\ell}\|v(x^{k_\ell})\|^2}{2} < \delta f(x^{k_\ell}, v(x^{k_\ell})).$$

Remember that $\{\alpha_{k_\ell}\} \subset [0,\rho)$. Then, we have the following two cases to consider.

- First, assume that $\lim_{l\to\infty} \alpha_{k_l} = 0$. Observe that

$$0 \le \nu_{k_\ell}\|v(x^{k_\ell})\|^2 \le 2|\delta f(x^{k_\ell}, v(x^{k_\ell})) - f(x^{k_\ell} + \rho_{k_\ell} v(x^{k_\ell}), v(x^{k_\ell}))|.$$

  Since $v$ and $f$ are continuous functions, we have that

$$\{|\delta f(x^{k_\ell}, v(x^{k_\ell}) - f(x^{k_\ell} + \rho_{k_\ell} v(x^{k_\ell}), v(x^{k_\ell}))|\}$$

  is convergent and, henceforth,

$$\{\nu_{k_\ell}\|v(x^{k_\ell})\|^2\}$$

  is bounded. Now, taking limit in (3-10), we get

$$\begin{aligned} f(x^*, v(x^*)) &= \lim_{l\to\infty}\left[f(x^{k_l} + \tilde{\alpha}_{k_l} v(x^{k_l}), v(x^{k_l})) + \frac{\nu_k\tilde{\alpha}_{k_l}\|v(x^{k_l})\|^2}{2}\right] \\ &= \delta f(x^*, v(x^*)). \end{aligned}$$

  Hence, $f(x^*, v(x^*)) = 0$ because $\delta \neq 1$.
- Now, assume that there exists a subsequence $\{\alpha_{k_{\ell_s}}\}$ of $\{\alpha_{k_\ell}\}$ such that $\lim_{s\to\infty} \alpha_{k_{\ell_s}} = 2\alpha > 0$. Then, for any given $\varepsilon > 0$, there exists $\kappa$ such that $s > \kappa$ implies

$$\varepsilon > \alpha_{k_{\ell_s}}\|v(x^{k_{\ell_s}})\| > \alpha\|v(x^{k_{\ell_s}})\|.$$

  Therefore, $\|v(x^*)\| = 0$, and consequently, $f(x^*, v(x^*)) = 0$.

□

### 3.3.1 Rate of convergence

In this subsection, $L$ is Lipschitz constant of $JF$.

We claim that zero is not an accumulation point of $\{\alpha_{k_s}\}$. Indeed, suppose that $\{\alpha_{k_s}\}$ is a subsequence of $\{\alpha_k\}$ with $\lim_{s\to\infty}\alpha_{k_s}=0$. Without loss of generality, we can state that

$$f(x^{k_s}+w^{-1}\alpha_{k_s}v(x^{k_s}),v(x^{k_s}))+\frac{\nu_{k_s}\omega^{-1}\alpha_{k_s}\|v(x^k)\|^2}{2}>\delta f(x^{k_s},v(x^{k_s})) \tag{3-11}$$

for all $s$. We have assumed that

$$2L<\frac{1-\delta}{\rho}.$$

Therefore,

$$2L+\frac{1-\delta}{\rho}<2\frac{1-\delta}{\rho}.$$

Hence,

$$\begin{aligned}\omega^{-1}\alpha_{k_s}\|v(x^{k_s})\|^2 2\left(\frac{1-\delta}{\rho}\right) &\geq \omega^{-1}\alpha_{k_s}\|v(x^{k_s})\|^2\left(2L+\frac{1-\delta}{\rho}\right)\\ &\geq \omega^{-1}\alpha_{k_s}\|v(x^{k_s})\|^2(L+\nu_{k_s}/2)\\ &= L\omega^{-1}\alpha_{k_s}\|v(x^{k_s})\|^2+\frac{\nu_{k_s}\omega^{-1}\alpha_{k_s}\|v(x^k)\|^2}{2}\\ &\geq (\delta-1)f(x^{k_s},v(x^{k_s})).\end{aligned}$$

Where the second inequality holds by Lemma 3.2 and we get the last inequality using (3-11) with the Lipschitz-continuity of $JF$. Therefore,

$$\alpha_{k_s}>\frac{-f(x^{k_s},v(x^{k_s}))\omega\rho}{2\|v(x^{k_s})\|^2}.$$

Noting that

$$f(x^{k_s},v(x^{k_s}))+\|v(x^{k_s})\|^2=0,$$

we get

$$\alpha_{k_s}>\frac{\omega\rho}{2},$$

which gives a contradiction. Without loss of generality, we can assume that

$$\alpha_k>\alpha>0,$$

for all $k$. We denote $f_n^*=\min_{0\leq k\leq n}|f(x^k,v(x^k))|$. Hence,

$$(n+1)\delta\alpha f_n^*\leq\delta\alpha\sum_{k=0}^{n}|f(x^k,v(x^k))|\leq\delta\sum_{k=0}^{n}\alpha_k|f(x^k,v(x^k))|=-\delta\sum_{k=0}^{n}\alpha_k f(x^k,v(x^k)).$$

Consider any $e \in \text{int}(K)$ such that $\max\{\langle e, w\rangle \colon w \in C\} = 1$. Substituting in (3-9), we get

$$\langle F(x^0) - F(x^*)\,,\, w\rangle \geq -\delta[\sum_{k=0}^{n} \alpha_k f(x^k, v(x^k))]\langle e, w\rangle \geq (n+1)\delta\alpha f_n^* \langle e, w\rangle$$

for all $w \in C$. Remember that $\|\omega\| = 1$. Therefore, using the Cauchy-Schwarz Inequality, we obtain the inequality

$$f_n^* \leq \frac{1}{n+1}\frac{1}{\alpha\delta}\|F(x^0) - F(x^*)\|,$$

which describes the convergence rate of $\{f_n^*\}$ to zero.

## 3.4 General case

In this section we will present an algorithm for the general case, this is, we will assume neither the convexity of $F$ nor Lipschitz continuity of $JF$.

**Algorithm 3.13.** *We need three exogenous constants: $\delta, \omega \in (0, 1)$, and $\rho > 0$*

**0. Initialization**: *Let it be $x^0 \in \mathbb{R}^n$. Compute $v(x^0)$, and initialize $k \leftarrow 0$.*

**1. Stopping criterium**: *If $v(x^k) = 0$, then STOP.*

**2. Line search**: *Compute*

$$\nu_k = \max\left\{0\,,\, 2\frac{\delta f(x^k, v(x^k)) - f(x^k + \rho v(x^k), v(x^k))}{\rho\|v(x^k)\|^2}\right\} \tag{3-12}$$

*and*

$$i_k = \min\left\{i \geq 1 \;\mid\; f(x^k + \rho\omega^i v(x^k), v(x^k)) + \frac{\nu_k \rho\omega^i \|v(x^k)\|^2}{2} < \delta f(x^k, v(x^k))\right\}. \tag{3-13}$$

**3. Iteration step**: *Define*

$$\alpha_k = \rho\omega^{i_k} \tag{3-14}$$

*and*

$$x^{k+1} = x^k + \alpha_k v(x^k). \tag{3-15}$$

*Compute $v(x^{k+1})$, set $k \leftarrow k+1$, and go to Step 1.*

We will continue to present the convergence results for Algorithm 3.13.

**Theorem 3.14.** *Let $\{x^k\}$ be the sequence generated by Algorithm 3.13. If $\{x^k\}$ is convergent, then its limit point is $\mathcal{K}$-critical.*

*Proof.* Let $x^* = \lim_{k\to\infty} x^k$. Then, by the continuity of $v$, $\lim_{k\to\infty} v(x^k) = v(x^*)$. Our goal is to demonstrate that $f(x^*, v(x^*)) = 0$.

By (3-12),

$$f(x^k + \rho v(x^k), v(x^k)) + \frac{\nu_k \rho \|v(x^k)\|^2}{2} \geq \delta f(x^k, v(x^k)).$$

Because $f(., v(x^k))$ is continuous, there exists $\tilde{\alpha}_k \in (\alpha_k, \omega^{-1}\alpha_k]$, such that

$$f(x^k + \tilde{\alpha}_k v(x^k), v(x^k)) + \frac{\nu_k \tilde{\alpha}_k \|v(x^k)\|^2}{2} = \delta f(x^k, v(x^k)) \tag{3-16}$$

because, by (3-13) and (3-14),

$$f(x^k + \alpha_k v(x^k), v(x^k)) + \frac{\nu_k \alpha_k \|v(x^k)\|^2}{2} < \delta f(x^k, v(x^k)).$$

So, we have two cases to consider.

- First, assume that $\lim_{k\to\infty} \alpha_k = 0$. By assumption $\{x^k\}$ is convergent. Therefore $\{\nu_k \|v(x^k)\|^2\}$ is bounded and $\lim_{k\to\infty} \alpha_k v(x^k) = 0$. Taking the limit in (3-16), we get

$$\begin{aligned} f(x^*, v(x^*)) &= \lim_{k\to\infty} \left[ f(x^k + \tilde{\alpha}_k v(x^k), v(x^k)) + \frac{\nu_k \tilde{\alpha}_k \|v(x^k)\|^2}{2} \right] \\ &= \delta f(x^*, v(x^*)). \end{aligned}$$

  Hence, $f(x^*, v(x^*)) = 0$ because $\delta \neq 1$.
- Now, assume that $\liminf \alpha_k = 2\alpha > 0$. Then, for all given $\varepsilon > 0$ there exists $\kappa$ such that $k > \kappa$ implies

$$\varepsilon > \alpha_k \|v(x^k)\| > \alpha \|v(x^k)\|, \quad \text{because} \quad \lim_{k\to\infty} \alpha_k \|v(x^k)\| = 0.$$

  Therefore, $\|v(x^*)\| = 0$, and consequently, $f(x^*, v(x^*)) = 0$.

□

To conclude this section, we will make an important observation. If it turns out that there exists $e \in int(\mathcal{K})$ such that $k = 0, 1, 2, \ldots$ and

$$F(x^{k+1}) \preceq_{\mathcal{K}} F(x^k) + \delta \alpha_k f(x^k, v(x^k))e, \tag{3-17}$$

as showed by Lemma 3.4 then we can prove that every accumulation point of the sequence $\{x^k\}$, generated by Algorithm 3.13, if any, is $\mathcal{K}$-critical. Note that

Algorithm 3.3 is a modification of Algorithm 3.13 for the convex case that guarantees the condition (3-17), as showed by Lemma 3.4. Just like this, Algorithm 3.11 is also a modification of procedure 3.13 that ensures condition (3-17) when the parameters satisfy $\frac{\rho}{1-\delta} < \frac{1}{2L}$ where $L$ is Lipschitz constant of the Jacobian, see Lemma 3.10.

## 3.5 Numerical experiments

This section presents the results of numerical experiments to evaluate the effectiveness of the algorithm and their ability to generate Pareto curves. All considered problems are multi-objective optimization-based. Thus, $\mathcal{K} = \mathbb{R}^m_+$ and $G$ is the canonical basis of $\mathbb{R}^m$.

The specifications of program, computer, stopping criteria and the maximum number of iterations are the same presented in Chapter 2 and therefore will be omitted here.

### 3.5.1 Finding Pareto points

We have tested 37 nonconvex problems with parameters $\rho = 2$, $\omega = 0.9$, and $\delta = 10^{-3}$, while Algorithm 3.3 was tested using 19 convex problems with parameters $\rho = \omega = 0.9$ and $\delta = 10^{-3}$.

In the tables below, the column "Problem" indicates the names. "Source" lists the source papers of the problems. Column $n$ and $m$ indicate the numbers of variables and objectives, respectively. All the problems were solved 200 times using starting points from a uniform random distribution inside a box specified in $x^0$. The last four columns list the corresponding results. The "%" column lists the percentages of runs that reached a critical point. "it" lists the average iterations per successful runs, and "evalg" lists how many times the Jacobian was computed.

Algorithm ended at critical points for 100% of runs for 26 problems of runs. Its performance was only unsatisfactory for Lov2 (35% success) and LTDZ (only 17% success). For the other 8 problems, it achieved critical points at least 69% of the times.

Only the FDS problem was a challenge for Algorithm 3.3. It ended 69,5% of the runs at a critical point. This problem is known to be difficult – see [20].

| **Problem** | **Source** | $n$ | $m$ | $x^0$ | % | it | evalg | time |
|---|---|---|---|---|---|---|---|---|
| AP3 | [1] | 2 | 2 | $[-10, 10]^n$ | 95 | 2641.2 | 7965.9 | 4.871 |
| DD1 | [10] | 5 | 2 | $[-20, 20]^n$ | 100 | 79.15 | 251.45 | 0.169 |
| DGO1 | [32] | 1 | 2 | $[-10, 13]^n$ | 100 | 5.07 | 21.45 | 0.014 |
| DTLZ2.2 | [11] | 3 | 3 | $[0, 1]^n$ | 86 | 15.91 | 69.34 | 0.039 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| FA1 | [32] | 3 | 3 | $[0.1, 1]^n$ | 100 | 12.76 | 39.90 | 0.027 |
| Far1 | [32] | 2 | 2 | $[-1, 1]^n$ | 100 | 90.22 | 293.77 | 0.179 |
| FF1 | [32] | 2 | 2 | $[-1, 1]^n$ | 100 | 33.52 | 104.94 | 0.069 |
| Hil1 | [31] | 2 | 2 | $[0, 1]^n$ | 100 | 22.70 | 97.52 | 0.050 |
| KW2 | [36] | 2 | 2 | $[-3, 3]^n$ | 91.5 | 216.93 | 673.54 | 0.397 |
| LE1 | [32] | 2 | 2 | $[-5, 10]^n$ | 100 | 80.11 | 272.28 | 0.151 |
| Lov2 | [38] | 2 | 2 | $[-0.75, 0.75]^n$ | 35 | 26.29 | 90.77 | 0.053 |
| Lov3 | [38] | 2 | 2 | $[-20, 20]^n$ | 100 | 7.33 | 36.01 | 0.018 |
| Lov4 | [38] | 2 | 2 | $[-100, 100]^n$ | 100 | 9.56 | 44.84 | 0.022 |
| Lov5 | [38] | 3 | 2 | $[-2, 2]^n$ | 100 | 47.41 | 146.13 | 0.097 |
| LTDZ | [37] | 3 | 3 | $[0, 1]^n$ | 17 | 2578.3 | 7750.2 | 4.769 |
| MGH7 | [43] | 3 | 3 | $[-2, 2]^n$ | 100 | 150.53 | 486.38 | 0.290 |
| MGH26 | [43] | 4 | 4 | $[-1, 1]^n$ | 100 | 14.04 | 55.59 | 0.031 |
| MLF1 | [32] | 1 | 2 | $[0, 20]^n$ | 100 | 4.96 | 20.61 | 0.015 |
| MLF2 | [32] | 2 | 2 | $[-100, 100]^n$ | 100 | 2.02 | 9.79 | 0.006 |
| MMR1 | [41] | 2 | 2 | $[0, 1]^n$ | 100 | 7.16 | 53.85 | 0.018 |
| MMR3 | [41] | 2 | 2 | $[-1, 1]^n$ | 69 | 14.93 | 45.84 | 0.045 |
| MMR5 | [41] | 50 | 2 | $[-50, 50]^n$ | 94 | 7923.9 | 23772.8 | 16.560 |
| MOP2 | [32] | 2 | 2 | $[-1, 1]^n$ | 100 | 35.40 | 113.73 | 0.071 |
| MOP3 | [32] | 2 | 2 | $[-\pi, \pi]^n$ | 100 | 16.72 | 73.00 | 0.035 |
| MOP5 | [32] | 2 | 3 | $[-1, 1]^n$ | 100 | 15.45 | 47.94 | 0.035 |
| QV1 | [32] | 10 | 2 | $[-5.12, 5.12]^n$ | 100 | 829.1 | 2497.4 | 1.611 |
| SK1 | [32] | 1 | 2 | $[-100, 100]^n$ | 100 | 0.82 | 5.79 | 0.004 |
| SK2 | [32] | 4 | 2 | $[-10, 10]^n$ | 100 | 41.51 | 135.41 | 0.080 |
| SLC2 | [50] | 100 | 2 | $[-10, 10]^n$ | 100 | 460.39 | 949.41 | 1.411 |
| SLCDT1 | [51] | 2 | 2 | $[-1.5, 1.5]^n$ | 100 | 2.77 | 20.32 | 0.009 |
| SSFYY2 | [32] | 1 | 2 | $[-10, 10]^n$ | 100 | 3.15 | 21.59 | 0.010 |
| TKLY1 | [32] | 4 | 2 | $[0.1, 1] \times [0, 1]^n$ | 78 | 647.4 | 1995.7 | 1.219 |
| Toi9 | [53] | 100 | 100 | $[-100, 100]^n$ | 100 | 32.17 | 122.75 | 0.678 |
| Toi10 | [53] | 10 | 9 | $[-2, 2]^n$ | 80 | 4335.0 | 13072.3 | 9.630 |
| VU1 | [32] | 2 | 2 | $[-3, 3]^n$ | 100 | 1361.61 | 4108.63 | 2.416 |

**Table 3.1:** *Performance of nonconvex problems.*

| **Problem** | **Source** | $n$ | $m$ | $x^0$ | % | it | evalg | time |
|---|---|---|---|---|---|---|---|---|
| AP1 | [1] | 2 | 3 | $[-10, 10]^n$ | 100 | 964.32 | 1939.56 | 1.8758 |
| AP2 | [1] | 1 | 2 | $[-100, 100]^n$ | 100 | 4.43 | 14.87 | 0.0131 |
| AP4 | [1] | 3 | 3 | $[-10, 10]^n$ | 100 | 853.39 | 1718.03 | 1.6716 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| BK1 | [32] | 2 | 2 | $[-5, 10]^n$ | 100 | 3.76 | 13.52 | 0.0111 |
| DGO2 | [32] | 1 | 2 | $[-9, 9]^n$ | 100 | 71.83 | 144.66 | 0.1502 |
| FDS | [20] | 50 | 3 | $[-2, 2]^n$ | 69.5 | 1799.16 | 3614.50 | 4.1121 |
| IKK1 | [32] | 2 | 3 | $[-50, 50]^n$ | 100 | 3.12 | 11.15 | 0.0095 |
| JOS1 | [35] | 10 | 2 | $[-10^4, 10^4]^n$ | 100 | 91.17 | 183.34 | 0.1700 |
| Lov1 | [38] | 2 | 2 | $[-100, 100]^n$ | 100 | 4.64 | 15.29 | 0.0139 |
| MGH33 | [43] | 10 | 10 | $[-1, 1]^n$ | 100 | 2.19 | 35.61 | 0.0122 |
| MHHM2 | [32] | 2 | 3 | $[0, 1]^n$ | 100 | 2.81 | 11.56 | 0.0093 |
| MOP1 | [32] | 1 | 2 | $[-5, 5]^n$ | 100 | 2.71 | 10.49 | 0.0091 |
| MOP7 | [32] | 2 | 3 | $[-400, 400]^n$ | 100 | 159.37 | 319.73 | 0.3117 |
| PNR | [48] | 2 | 2 | $[-2, 2]^n$ | 100 | 12.37 | 41.13 | 0.0301 |
| SLCDT2 | [51] | 10 | 3 | $[-1, 1]^n$ | 100 | 20.49 | 56.05 | 0.0433 |
| SP1 | [32] | 2 | 2 | $[-100, 100]^n$ | 100 | 41.94 | 97.79 | 0.0815 |
| Toi4 | [53] | 4 | 2 | $[-100, 100]^n$ | 100 | 4.99 | 15.98 | 0.0133 |
| Toi8 | [53] | 3 | 3 | $[-1, 1]^n$ | 100 | 3.92 | 30.93 | 0.0127 |
| ZLT1 | [32] | 10 | 5 | $[-1000, 1000]^n$ | 100 | 5.96 | 18.80 | 0.0161 |

**Table 3.2:** *Performance of Algorithm 3.3.*

### 3.5.2 Building Pareto fronts

We tested the ability of our methods to generate Pareto frontiers appropriately. We considered four non-convex examples from Table 3.1 and four convex problems from Table 3.2, which are all bicriteria problems. The results are shown in Figures 3.1 and 3.3. For each problem, there are three graphics. The first ones were obtained by discretizing the corresponding boxes by a fine grid and plotting all the image points. These provide good representations of the image of $F$ and a geometric notion of the Pareto frontiers. The second graphics were obtained by running Algorithm for each non-convex and Algorithm 3.3 for each convex problem, 300 times, using randomly generated starting points belonging to the corresponding boxes. The third graphics were obtained in a similar manner, but by running the algorithms only 20 times. In these graphics, the image of a starting point is represented by a start, while black and blue points represent images of the final iterate and intermediately computed iterates, respectively. Straight segments link images of consecutive iterates.

Figure 3.1 shows that for the chosen set of test problems, considering a reasonable number of starting points, our algorithms were able to satisfactorily estimate the Pareto frontiers. We emphasize that the non-monotonous behavior of the sequences generated by Algorithm A, where it begins from some of the initial points, is observable in the third graphics, especially for FAR1, Hil1 and KW2.

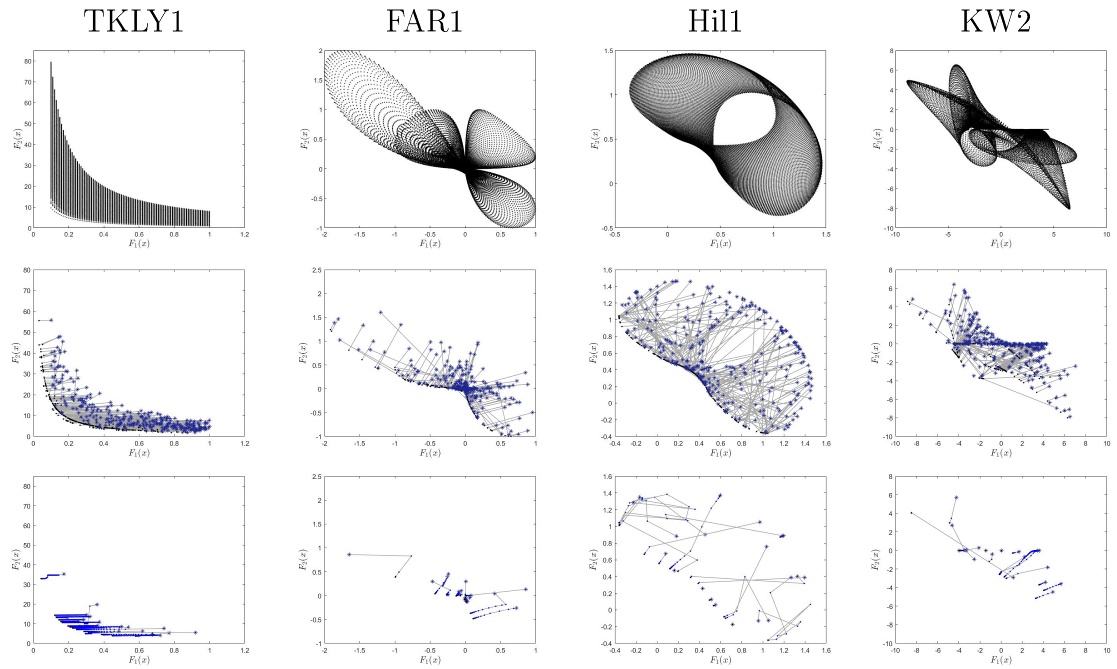



**Figure 3.1:** *Non-convex problems*

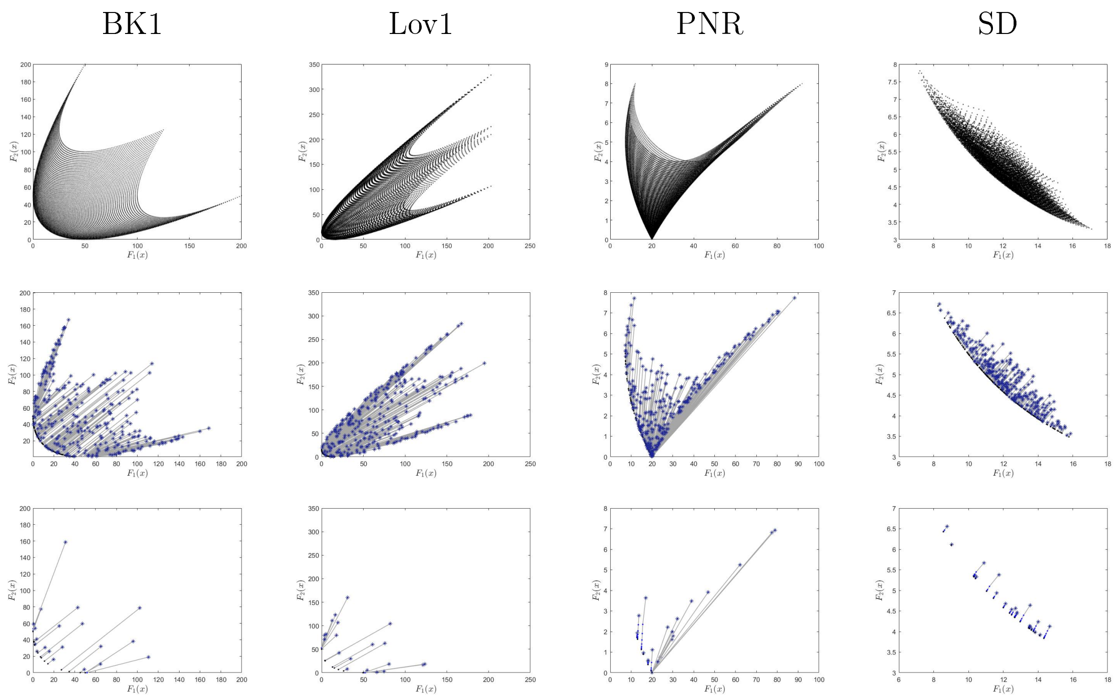



**Figure 3.3:** *Convex problems*

## 3.6 A new group of testing problems

Since our algorithm makes use of gradient values of the objective function only, unlike most of the algorithms from the literature, we will propose a set of test problems in which the computation of a gradient value requires less computational effort than that required for calculating an objective function value.

Given $b \in \mathbb{R}^n$ and $A$, symmetric positive definite $n \times n$ matrix, we define

$$q(x) = \frac{1}{2} x^T A x + b^T x.$$

Moreover, let $\zeta$, $\Upsilon$ and $\Gamma\colon \mathbb{R}^n \to \mathbb{R}$ defined as

$\zeta(x) = \sum_{i=1}^n x_i^2 + \arctan(x_i)$

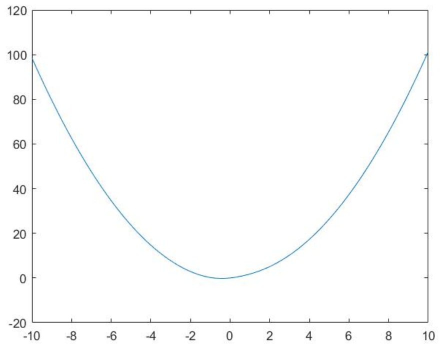



**Figure 3.4:** *Graphic of* $\zeta(x)$ *for* $n = 1$

$\Gamma(x) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^n \int_{-\infty}^{x_i} e^{-s^2/2} ds$

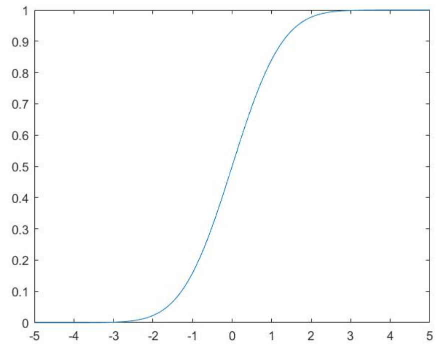



**Figure 3.5:** *Graphic of* $\Gamma(x)$ *for* $n = 1$

and

$$\Upsilon(x) = \sum_{i=1}^n \ln(x_i^2 + e)$$

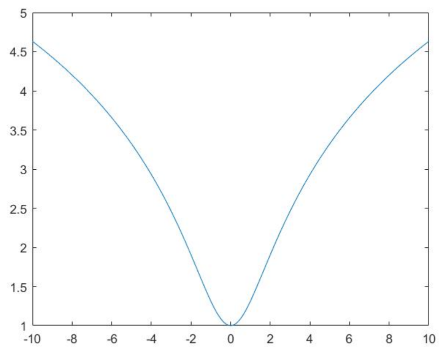



**Figure 3.6:** *Graphic of* $\Upsilon(x)$ *for* $n = 1$

A simple computation shows that

$$\nabla q(x) = Ax + b, \quad \nabla\zeta(x)_i = 2x_i + \frac{1}{1 + x_i^2}, \quad \nabla\Gamma(x)_i = \frac{e^{-x_i^2/2}}{\sqrt{2\pi}} \quad \text{and}$$

$$\nabla \Upsilon(x)_i = \frac{2x_i}{x_i^2 + e}.$$

Function $q$ is strongly convex with module $\|A\|$. $\zeta$ is separable and convex with only one minimizer. The other two functions are separable and non-convex. $\Gamma$ has not minimizer and $(0, n)$ is its image. $\Upsilon$ achieves its minimum at the origin.

In order to build a set of test problems, we combine functions $q$, $\zeta$, $\Upsilon$ and $\Gamma$ forming different objectives. Altogether there are 11 problems, which are presented in the following table. Columns # show the number that will identify each problem from now on.

| # | $F(x)$ | # | $F(x)$ | # | $F(x)$ |
|---|---|---|---|---|---|
| 1 | $(q(x)\,,\, \zeta(x))^T$ | 7 | $(q(x)\,,\, \zeta(x)\,,\, \Gamma(x))^T$ | 11 | $(q(x)\,,\, \zeta(x)\,,\, \Gamma(x)\,,\, \Upsilon(x))^T$ |
| 2 | $(q(x)\,,\, \Gamma(x))^T$ | 8 | $(q(x)\,,\, \zeta(x)\,,\, \Upsilon(x))^T$ | | |
| 3 | $(q(x)\,,\, \Upsilon(x))^T$ | 9 | $(q(x)\,,\, \Gamma(x)\,,\, \Upsilon(x))^T$ | | |
| 4 | $(\zeta(x)\,,\, \Gamma(x))^T$ | 10 | $(\zeta(x)\,,\, \Gamma(x)\,,\, \Upsilon(x))^T$ | | |
| 5 | $(\zeta(x)\,,\, \Upsilon(x))^T$ | | | | |
| 6 | $(\Gamma(x)\,,\, \Upsilon(x))^T$ | | | | |

**Table 3.3:** *New problems.*

The following figures illustrate the image, Pareto fronts and optimal solutions sets of the first six problems when $n = 2$ and

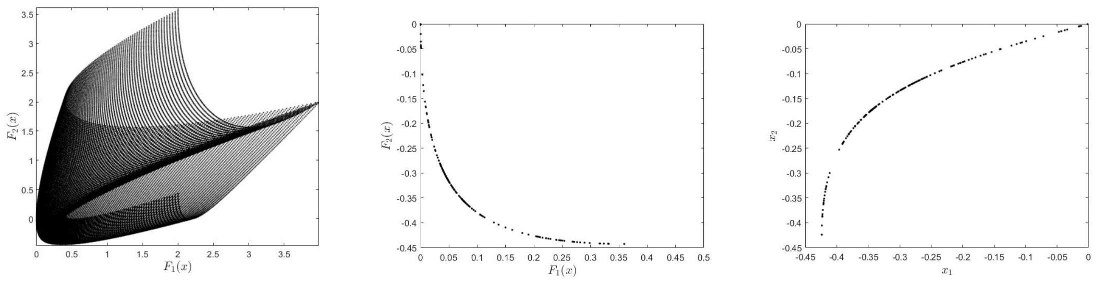
$$A = \begin{pmatrix} 1 & -1 \\ -1 & 5 \end{pmatrix}.$$

For Problems 1 and 2, $b$ is the zero vector of $\mathbb{R}^2$ and for Problem 3, $b = (0, 4)^T$. The starting point was taken in the box $[-1, 1]$. For each of the six problems we have three graphs. The first ones were obtained by discretizing the box $[-1, 1]$ by a fine grid and plotting the image points. The second graphics were constructed compiling the algorithm for each problem 300 times, thus obtaining a critical point and then plotting its image. The third graphics were obtained plotting the 300 critical points. The performance of the algorithm for this situation is shown in Table 3.4.

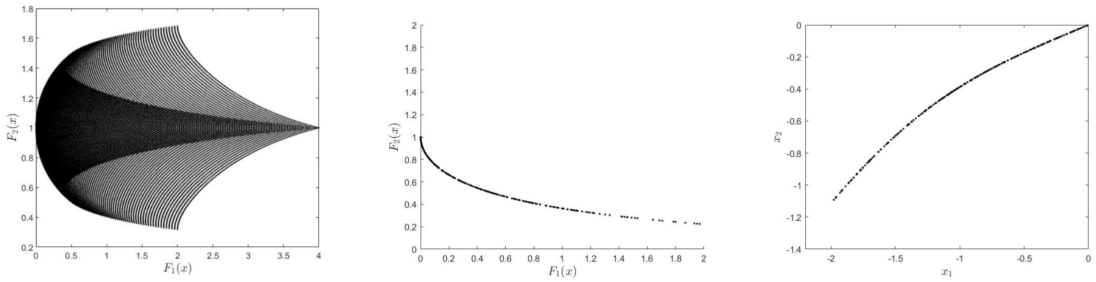
| *Problem* | $x^0$ | % | *it* |
|---|---|---|---|
| 1 | $[-1,1]^n$ | 100.00 | 7.13 |
| 2 | $[-1,1]^n$ | 100.00 | 19.56 |
| 3 | $[-1,1]^n$ | 100.00 | 6.26 |
| 4 | $[-1,1]^n$ | 100.00 | 19.50 |
| 5 | $[-1,1]^n$ | 100.00 | 6.05 |
| 6 | $[-1,1]^n$ | 100.00 | 7.15 |
| 7 | $[-1,1]^n$ | 100.00 | 13.20 |
| 8 | $[-1,1]^n$ | 100.00 | 5.10 |
| 9 | $[-1,1]^n$ | 100.00 | 7.40 |
| 10 | $[-1,1]^n$ | 100.00 | 6.91 |
| 11 | $[-1,1]^n$ | 100.00 | 7.52 |

**Table 3.4:** *Performance of new problems.*

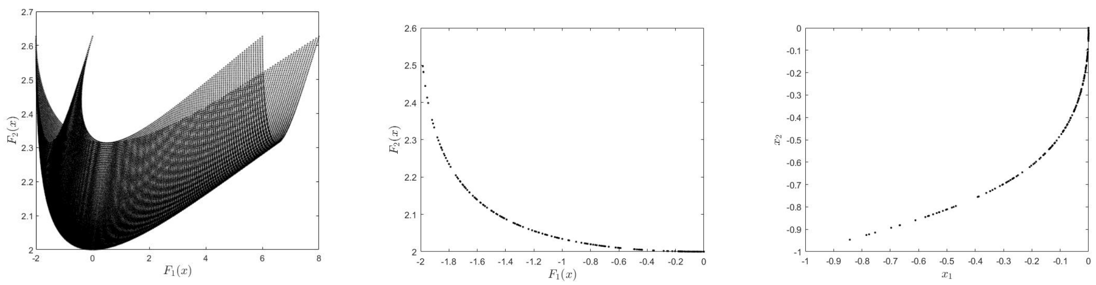
$$F_1(x) = (q(x)\,,\, \zeta(x))^T$$



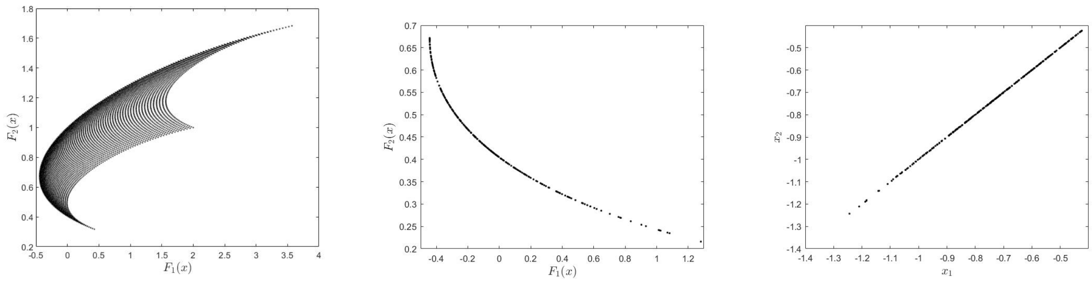

$$F_2(x) = (q(x)\,,\, \Gamma(x))^T$$



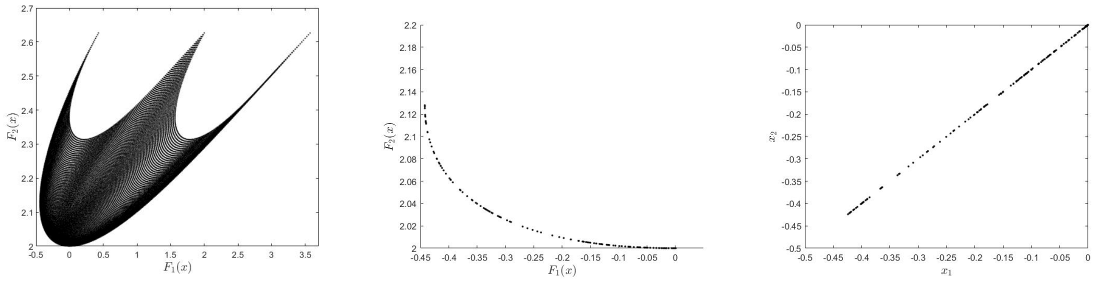

$$F_3(x) = (q(x)\,,\, \Upsilon(x))^T$$



$$F_4(x) = (\zeta(x)\,,\,\Gamma(x))^T$$




$$F_5(x) = (\zeta(x)\,,\,\Upsilon(x))^T$$




$$F_6(x) = (\Gamma(x)\,,\,\Upsilon(x))^T$$

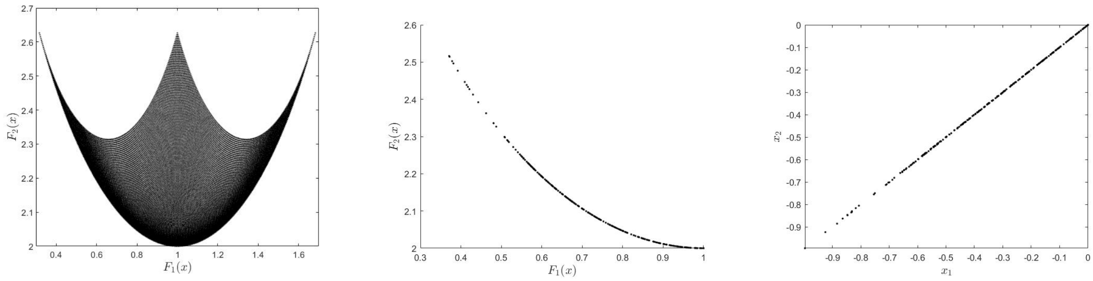
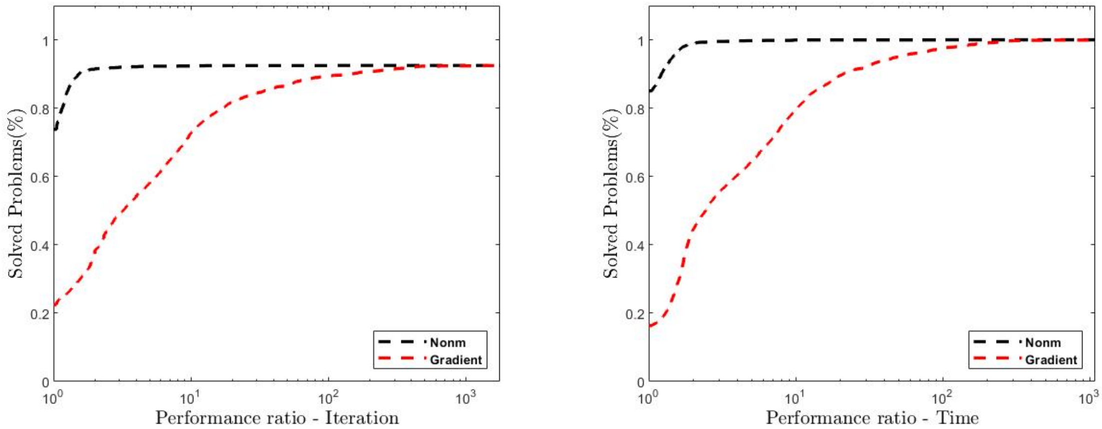


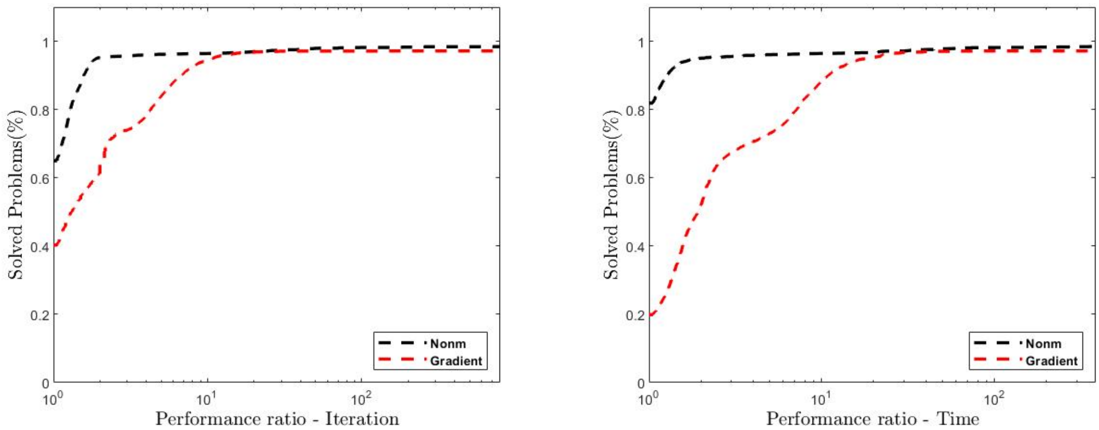

**Figure 3.9:** *Graphics of Image, Pareto Front and problem solving with two objective functions, two variables and the starting point taken in the range [-1,1].*

We compare the steepest descent algorithm with the two different line searches, the Armijo's one and ours, using as criterium the average number of iterations needed to find a critical point, and the average employed CPU time.

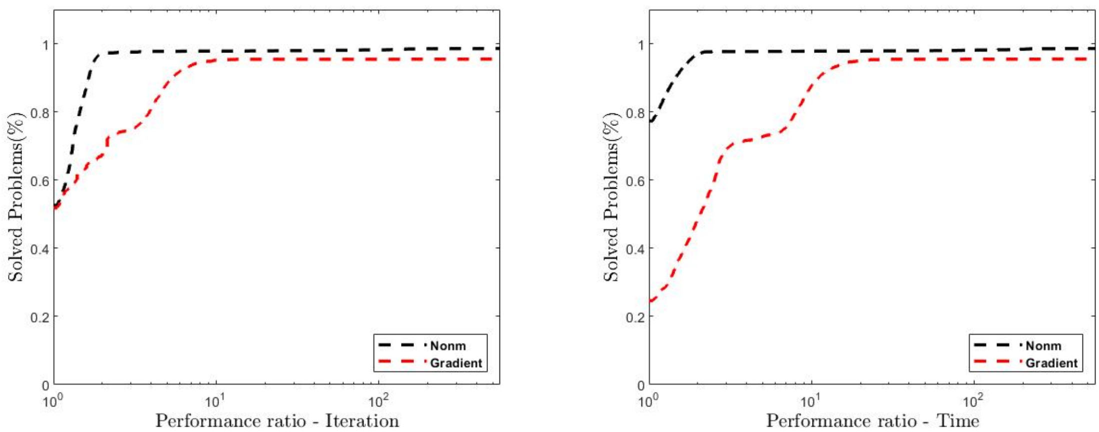
**n=2**


Solved Problems(%)
Performance ratio - Iteration
Nonm
Gradient
Solved Problems(%)
Performance ratio - Time
Nonm
Gradient

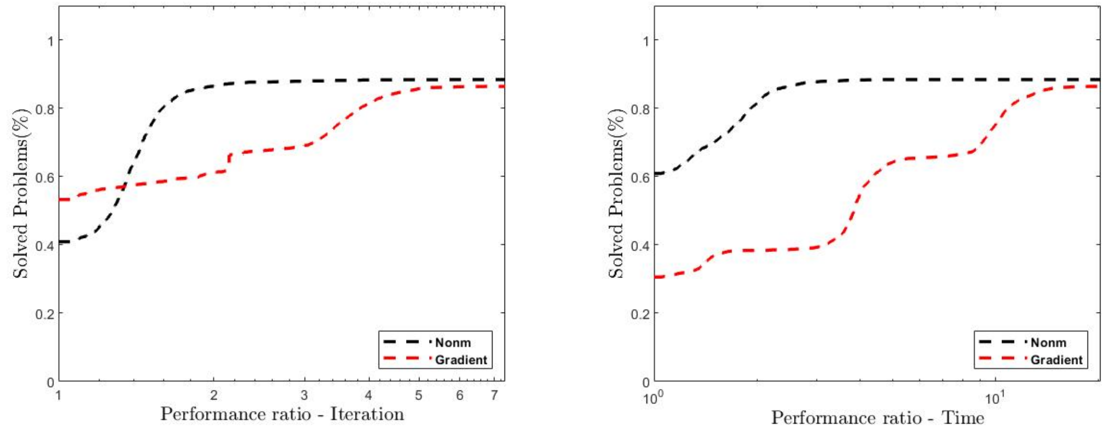
**n=10**


Solved Problems(%)
Performance ratio - Iteration
Nonm
Gradient
Solved Problems(%)
Performance ratio - Time
Nonm
Gradient

**n=20**


Solved Problems(%)
Performance ratio - Iteration
Nonm
Gradient
Solved Problems(%)
Performance ratio - Time
Nonm
Gradient

**n=50**




**n=100**

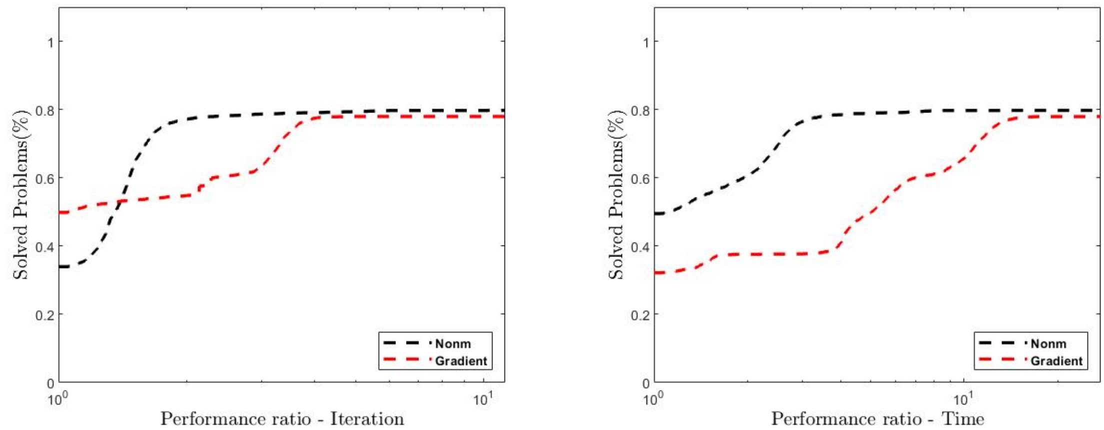



**Figure 3.12:** *Performance Profiles comparing iteration and time for the group of 11 new problems.*

With the same data obtained to generate the Performance Profile above, we decided, to check the quality of the Pareto front using the metrics: Purity metric, Spread Metric-Delta and Spread Metric-Gamma-see [8, 22] or Appendix A. So we can compare which Pareto curve has the least dominated points (Purity metric) and which ones have less holes (Spread Metric-Delta e Spread Metric-Gamma). The results obtained are shown in Figure 3.14.

We can observe that when the Gradiente algorithm is compared with the two searches (Armijo and the new search), there exists a highlight for the new search, both in terms of time and iteration. And when we look at the quality of the points that belong to the Pareto curve, we see that the new search does not lose in quality respect to the Armijo search. Therefore, the new search has significant contributions

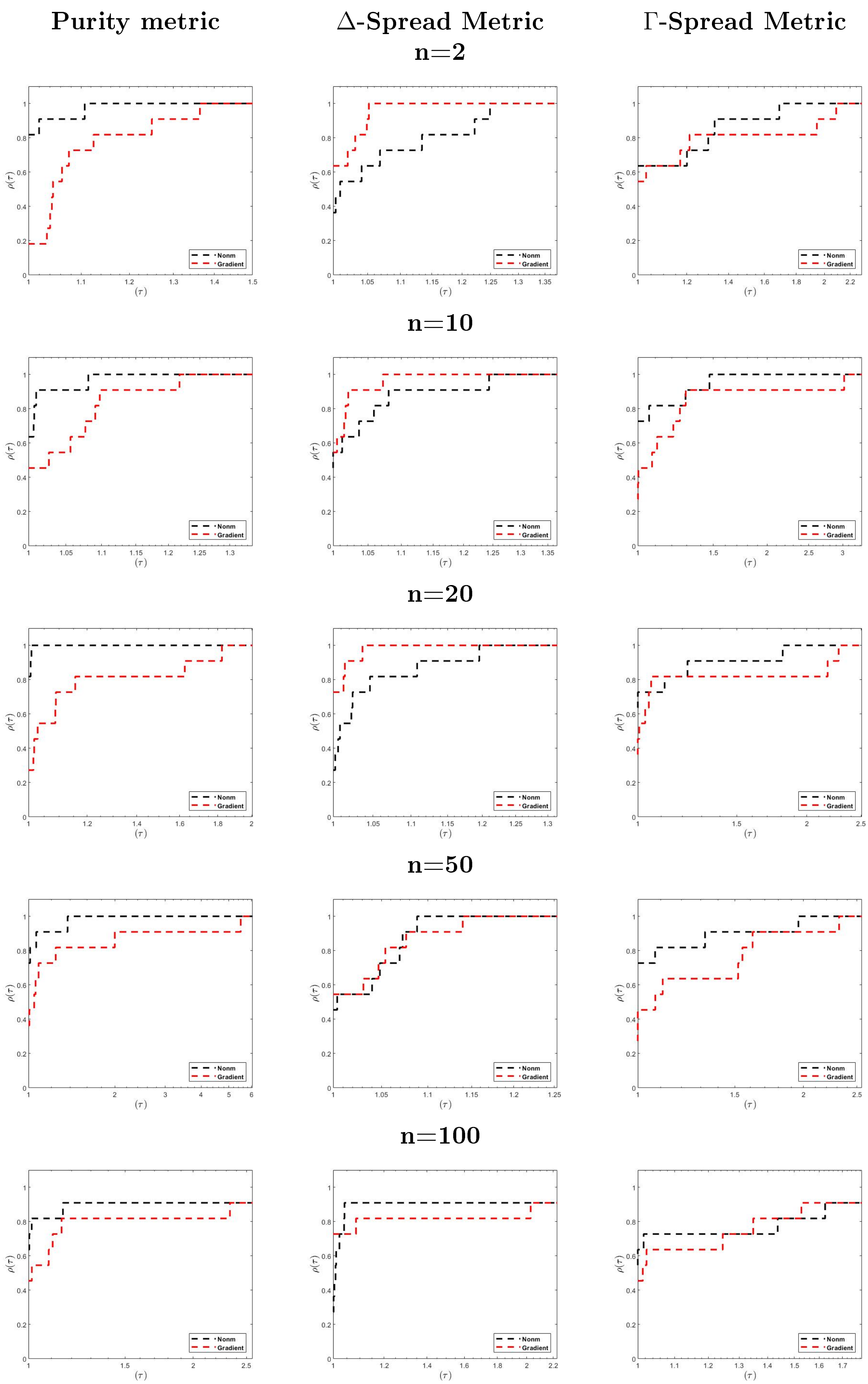



**Figure 3.14:** *Purity metric, Spread Metric-Delta, Spread Metric-Gamma for the group of 11 new problems.*

when is used in this new group of problems.

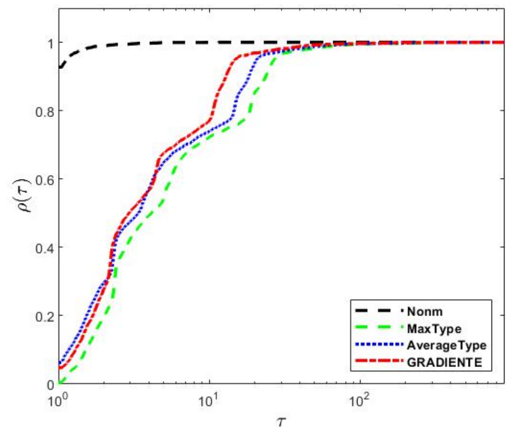
### 3.6.1 Non-monotone algorithms

Since in general our algorithm is non-monotone, we decided to compare it with others that have this same characteristic. We make the Performance Profile confronting iteration and time of the gradient algorithms with the linear search of Armijo, average-type and max-type, with the parameters suggested by the authors, see [42]. For this, we changed the number of variables and the interval where the starting point is taken, this information can be found in Table 3.5.

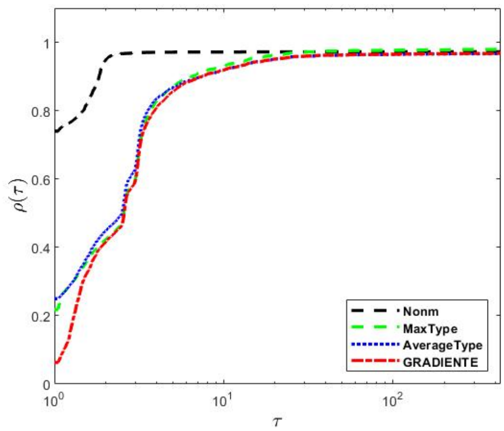
| Problem | $n$ | $x^0$ |
|---|---|---|
| 1 | 100 | $[-10^3, 10^3]$ |
| 2 | 5 | $[-5, 5]$ |
| 3 | 10 | $[-10^2, 10^2]$ |
| 4 | 50 | $[-5, 5]$ |
| 5 | 200 | $[-10^2, 10^2]$ |
| 6 | 500 | $[-5, 5]$ |
| 7 | 2 | $[-5, 5]$ |
| 8 | 10 | $[-10^2, 10^2]$ |
| 9 | 10 | $[-5, 5]$ |
| 10 | 100 | $[-5, 5]$ |
| 11 | 10 | $[-5, 5]$ |

**Table 3.5:** *New problems, variable numbers and starting point range.*

**Time**

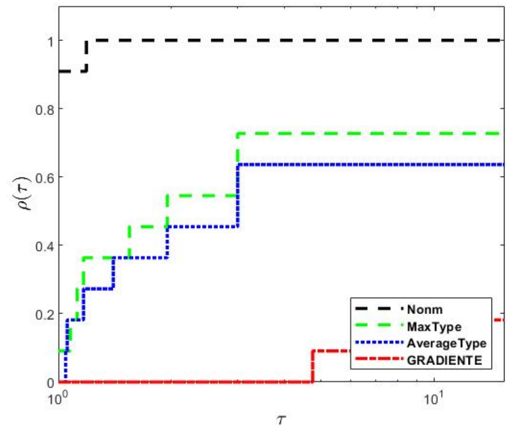


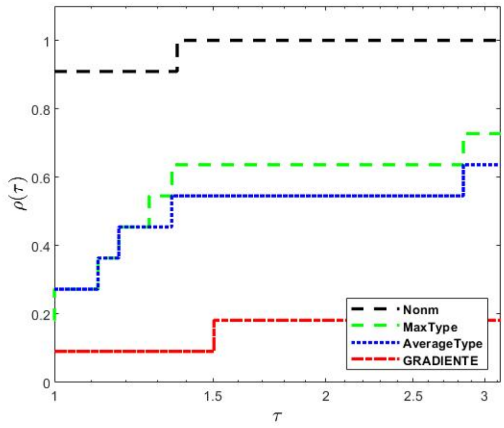

**Iteration**



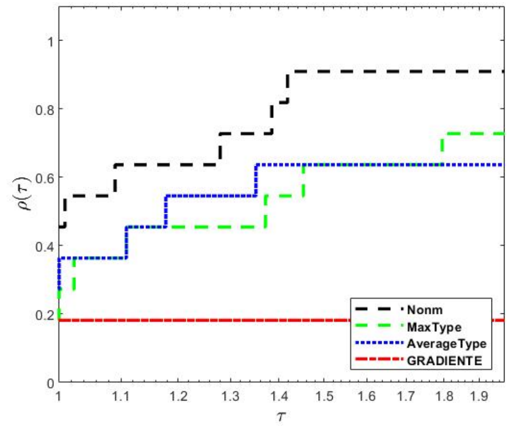

**Purity Metric**

**Γ−Spread Metri**




**Δ−Spread Metric**




We can see a very significant performance of the new linear search, both in terms of effectiveness and robustness of the Algorithm 3.13, when it is compared with itself using Armijo's search and with the Averagr-type and Max- type. In the graphs that compare Iteration and Time, we see that the new search is the one that stands out the most. And in the graphs that compare the quality of the Pareto Front (Purity Metric, $\Gamma-$Spread Metric and $\Delta-$Spread Metric) we observe that the algorithm with the new search generates a better Pareto curve than the other algorithms. So here we see a more significant highlight of the new search using this new group of problems.

### 3.6.2 Four new problems

We created a new set of test problems with four bi-objective optimization problems, that is, $m = 2$. Functions $\zeta$, $\Gamma$, $\Upsilon$, and $q$ were combined with the following one, $\theta : \mathbb{R}^n \to \mathbb{R}$, see [12, 41],

$$\theta(x) = \sum_{i=0}^{n} \left[ 2 - 0.8e^{-\left(\frac{x_i - 0.6}{0.4}\right)^2} - e^{-\left(\frac{x_i - 0.2}{0.004}\right)^2} \right].$$

As it was pointed before, $\Gamma$, $\Upsilon$, and $q$ have just one critical point which is the respective global minimizer. On the other hand, function $\theta$ has one local minimum at $0,6$ and one global minimum at $0,2$. Observe that the attraction valley of the local minimizer is much larger than the attraction valley of the global minimizer. That situation should challenge any procedure which intends to find global minimizers.

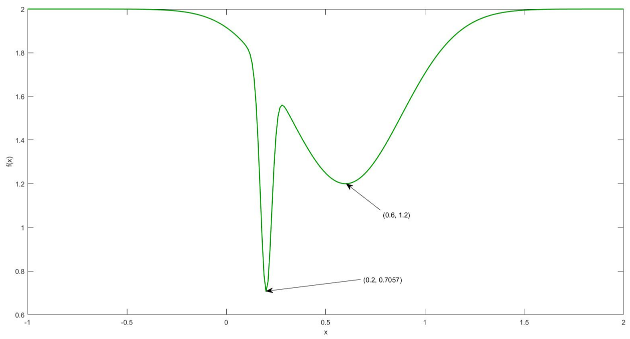



**Figure 3.16:** *Graphic of $\theta(x)$ for $n = 1$ :*

We created a new group of problems involving function $\theta(x)$ to check the efficiency of our algorithm:

| # | $F(x)$ |
|---|---|
| 12 | $(q(x)\,,\,\theta(x))^T$ |
| 13 | $(\zeta(x)\,,\,\theta(x))^T$ |
| 14 | $(\Gamma(x))\,,\,\theta(x))^T$ |
| 15 | $(\Upsilon(x)\,,\,\theta(x))^T$ |

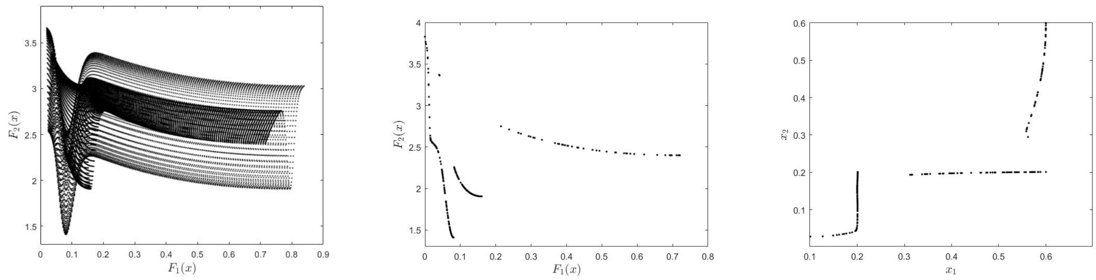
We will continue to describe our new set of problems. We will present the "Image" of the problems together with the "Pareto Front" and the Pareto Points of the problem from 12 to 15.

The graph of "Image" was constructed by discretizing the interval Pareto Points and plotting the image of the respective values.

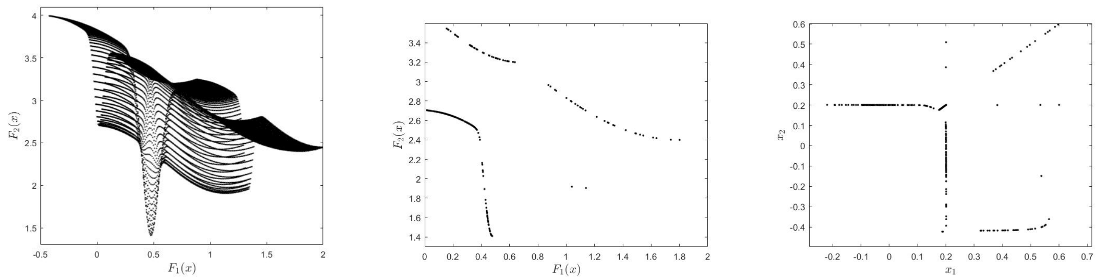
The Pareto Front was constructed plotting the image of solution of each problem obtained compiling our Algorithm 300 times, taking the starting point in the interval $[0, 1]$.

For the Pareto Points of the problem, we generated 300 starting points randomly in the interval $[0, 1]^n$ and the number of variables $n = 2$. The solutions were plotted and the results are as follows:

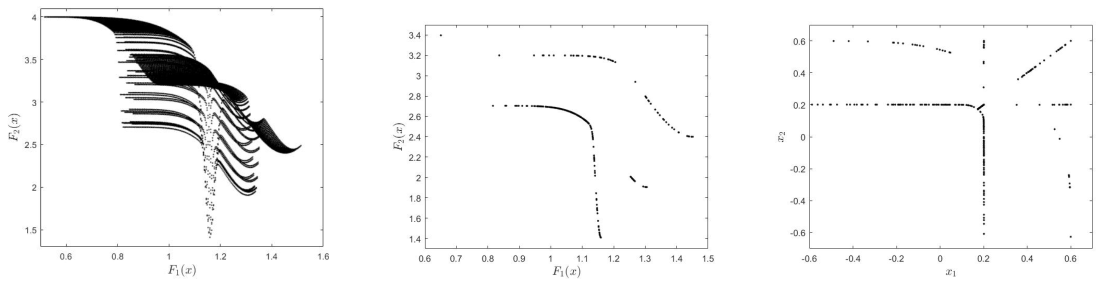
$$F_{12}(x) = (q(x)\,,\,\theta(x))^T$$



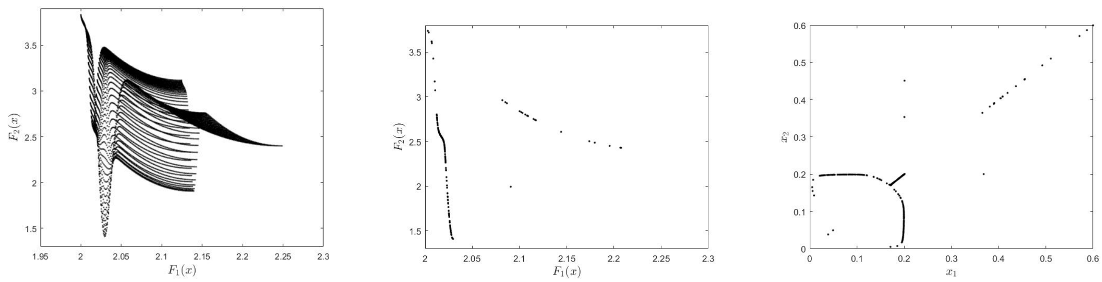

$$F_{13}(x) = (\zeta(x)\,,\,\theta(x))^T$$




$$F_{14}(x) = (\Gamma(x)\,,\,\theta(x))^T$$



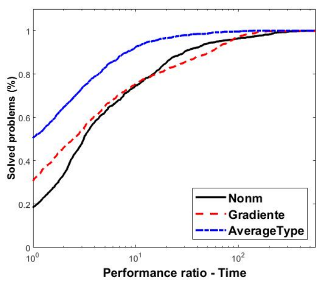

$$F_{15}(x) = (\Upsilon(x)\,,\,\theta(x))^T$$



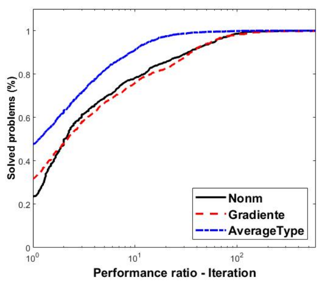

Now, it follows the numerical experiments involving the new set of problems. We will compare our algorithms with Gradiente and Average-Type [42], using different metrics, Iteration, Computational Time, the Purity Metric, $\Delta-$Spread Metric and $\Gamma-$Spread Metric, see [8, 22]. Figure 3.20 shows us the results of these experiments, which are presented in blocks by the number of variables, with six graphics in each. We did the experiments for $n = 2, 10, 20, 50$ and $100$, with the range of the starting point taken in $[0, 1]$. The first and second graphs of the block present the Performance Profile in relation to time and iteration, respectively. The third and fourth graphs compare our algorithm with the gradient and the Average-type, respectively, this is, check which of the algorithms has the Pareto front with less dominated points. The fifth and sixth ones compare the three algorithms in relation to the $\Delta-$Spread Metric and $\Gamma-$Spread Metric, respectively, this is, check which Pareto curve is more continuous, that is, which one has less holes.

**n=2**

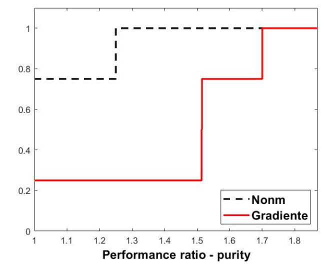
Time



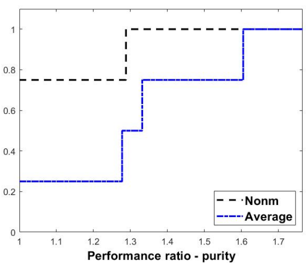

Iteration



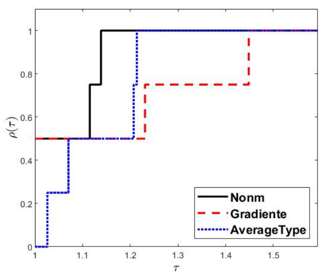

Purity Metric



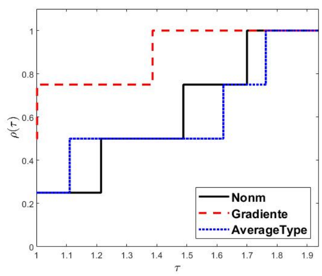

Purity Metric




$\Delta-$Spread Metric

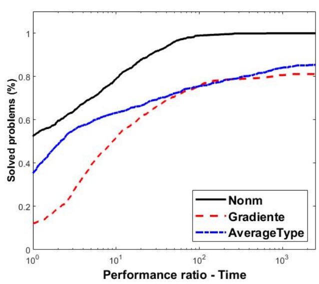



$\Gamma-$Spread Metric

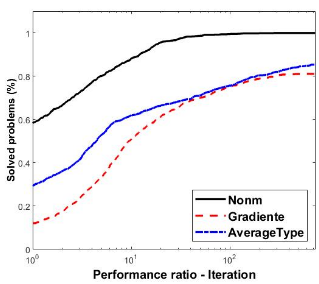



**n=10**

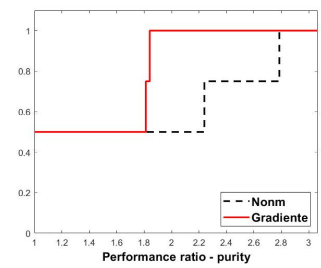
Time



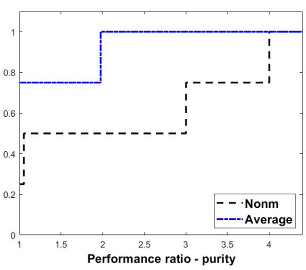

Iteration



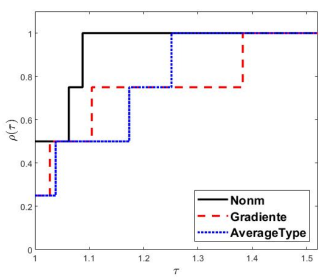

Purity Metric



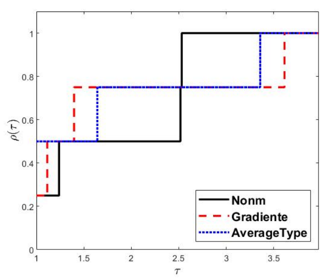

Purity Metric




$\Delta-$Spread Metric

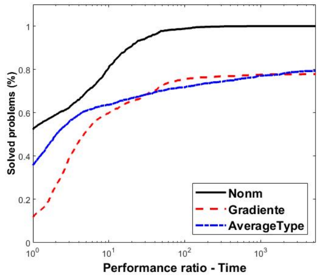



$\Gamma-$Spread Metric

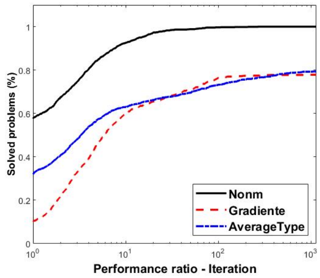



**n=20**

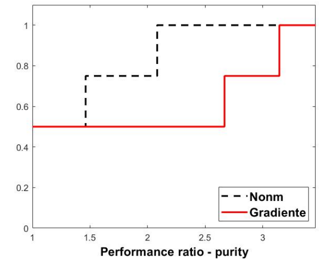
Time



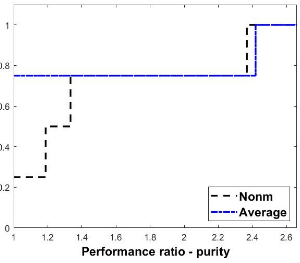

Iteration



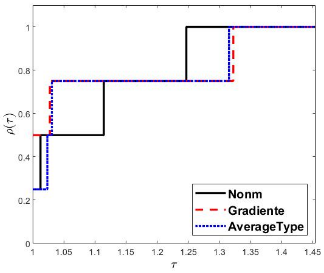

Purity Metric



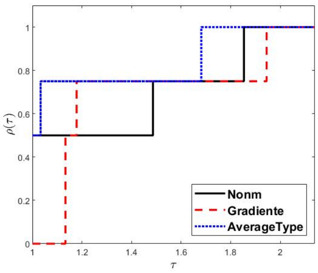

Purity Metric




Δ−Spread Metric

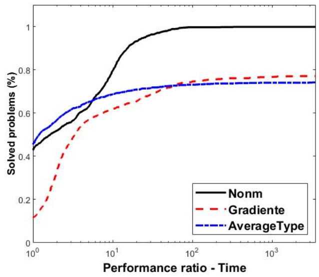



Γ−Spread Metric

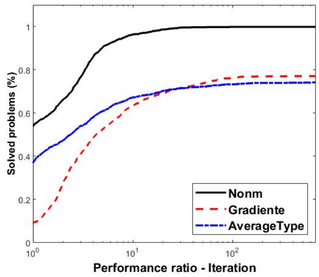



**n=50**

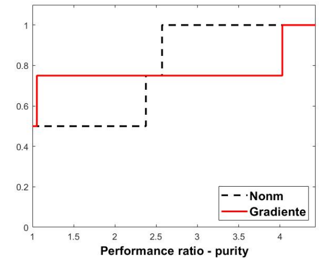
Time



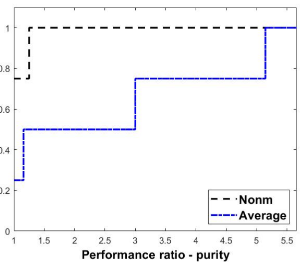

Iteration



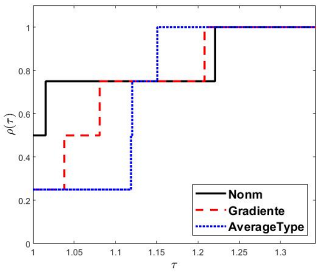

Purity Metric



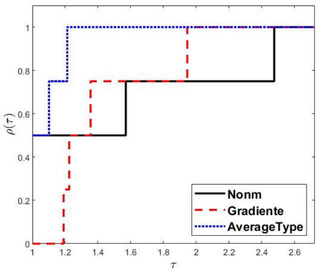

Purity Metric




Δ−Spread Metric




Γ−Spread Metric



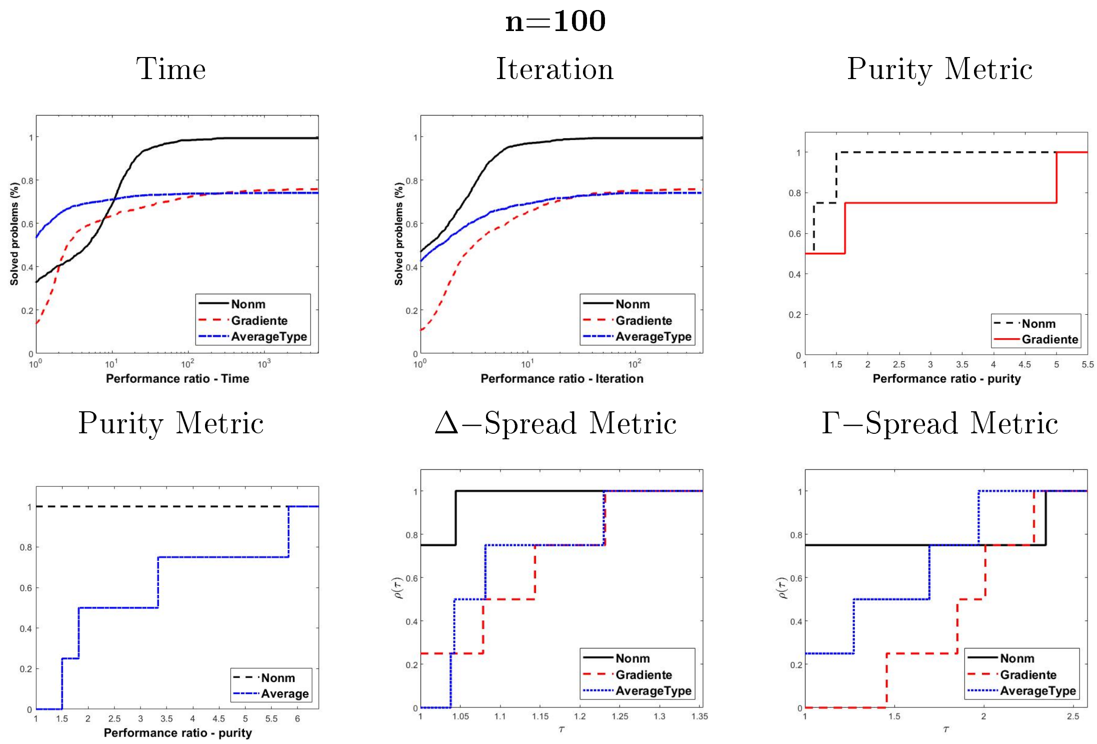



**Figure 3.20:** *Performance Profile in relation to time, iteration, Purity Metric, Δ−Spread Metric and Γ−Spread Metric with different values of variables.*

We can observe that for $n = 2$, the algorithm Average-Type has better performances both in time and iteration, and on Purity Metric our algorithm performs better, same in the $\Delta-$Spread Metric, but in the $\Gamma-$Spread Metric the gradient stands out.

For $n = 10$, our algorithm has better robustness and efficiency in relation to time and iteration, but it loses when we use it to Purity Metric, it performs well in $\Delta-$Spread Metric, but does not stand out in $\Gamma-$Spread Metric.

For $n = 20$, our algorithm performs well with both time and iteration, stands out in Purity Metric when compared to Gradient but loses to Average-Type. In $\Delta-$Spread Metric and $\Gamma-$Spread Metric the Average-type, performs better than the other algorithms.

With $n = 50$, when we look at the Performance Profile in relation to time, we observe that the Average-Type is more efficient, but our algorithm is more robust and in relation to iteration, ours has better performance. When we look at Purity Metric charts, we lose in efficiency and gain in robustness when we compare it with the gradient, and we have better performance if we compare it with the Average-Type. For the $\Delta-$Spread, Metric our algorithm is more efficient but Average-Type is more robust, while in $\Gamma-$Spread Metric we see better performance for Average-type.

If $n = 100$, we see that Average-type is more efficient but our algorithm is more robust when comparing time. In relation to iteration, ours has better performance as well as when we use Purity Metric, we also have significant performance for the $\Delta-$Spread Metric and $\Gamma-$Spread Metric.

## 3.7 Chapter conclusion

As we look at everything that has been presented in this chapter, both in theory and in numerical experiments, we emphasize that our algorithm has a satisfactory performance when we compare it with others using problems that have the characteristic cited by Yunda Dong, [14, 15] "problems that are easier to evaluate the gradient than functional values". However, when evaluating the performance of the non-monotonicity characteristic of line search, we see that our algorithm does not have a great advantage over others, such as the Average-Type. So this further underscores the importance of this work, that is, we present a new line search that does not make use of functional values and which has a significant performance in examples whose evaluation of gradients demands less computational effort than evaluating their respective functional values.

CHAPTER 4

# How to compute inexact $\mathcal{K}$-steepest descent directions

In this chapter we will present a practical way to calculate the inexact $\mathcal{K}$-steepest descent direction, this is, we will calculate some direction that is descending but not necessarily the steepest. Moreover, we will display numerical experiments that show the efficiency, the robustness and the quality of such procedure. This chapter was based on the work of Fliege and Svaiter [21], Drummond and Iusem [27], Drummond and Svaiter [28], and Fukuda and Drummond in [24].

## 4.1 $\sigma$-Approximate $\mathcal{K}$-steepest descent direction

We begin this section with two propositions known from the literature that can be found in [28], and which will be used in later results.

**Proposition 4.1.** *Let $d$ a $\mathcal{K}$-descent direction for $F$ at $x$ and $\beta \in (0,1)$. Then there exists $\hat{t}$ such that*

$$F(x+tv) \prec_{\mathcal{K}} F(x) + \beta t JF(x)d$$

*for all $t \in (0,\hat{t})$.*

*Proof.* See Proposition 2.1 in [28]. □

**Proposition 4.2.** *Let $\beta \in (0,1)$ and $d$ a $\mathcal{K}$-descent direction for $F$ at $x$. There exist $\hat{t}$, $\delta$ and $\delta' > 0$ such that $t < \hat{t}$, $\|x'-x\| < \delta$ and $\|d'-d\| < \delta'$ imply that $d'$ is $\mathcal{K}$-descent direction for $F$ at $x'$ and*

$$F(x'+tv') \prec_{K} F(x') + \beta t JF(x')v'.$$

*Proof.* See Proposition 3.6 in [28]. □

Next we will define a $\sigma$-Approximate $\mathcal{K}$-steepest descent direction, as well as the problem to calculate the direction and its dual, together with results that will be the basis for the conclusion of this chapter. Also note that we can define $f(x,d)$ about convex hull of $G$.

**Remark 4.3.** *Let $\tilde{G} = conv(G)$. Then,*

$$f(x,d) = \max\{\langle JF(x)d\,,\,w\rangle \mid w \in \tilde{G}\}.$$

*Proof.* It is clear that $f(x,d) \le \max\{\langle JF(x)d\,,\,w\rangle \mid w \in \tilde{G}\}$ because $G \subset \tilde{G}$. We have to proof that $f(x,d) \ge \max\{\langle JF(x)d\,,\,w\rangle \mid w \in \tilde{G}\}$. Take $\bar{w} \in \tilde{G}$ such that

$$\max\{\langle JF(x)d\,,\,w\rangle \mid w \in \tilde{G}\} = \langle JF(x)d\,,\,\bar{w}\rangle.$$

For each $w \in G$, there exists $\lambda(w) \ge 0$ such that $\bar{w} = \sum_{w\in G}\lambda(w)w$ and $\sum_{w\in G}\lambda(w) = 1$. Therefore,

$$\langle JF(x)d\,,\,\bar{w}\rangle = \sum_{w\in G}\lambda(w)\langle JF(x)d\,,\,w\rangle \le f(x,d).$$

□

An equivalent formulation for Problem (1-1) is

$$\min\left\{\max\left\{\frac{\|d\|^2}{2} + \langle JF(x)d\,,\,w\rangle \mid w \in \tilde{G}\right\} \mid d \in \mathbb{R}^n\right\}. \tag{4-1}$$

Indeed, first, observe that $v(x)$ is optimal solution of (4-1) because, taking in account Remark 4.3 and the definition of $v(x)$, for all $d \in \mathbb{R}^n$ it holds that

$$\begin{aligned}\max\left\{\frac{\|d\|^2}{2} + \langle JF(x)d\,,\,w\rangle \mid w \in \tilde{G}\right\} &= \frac{\|d\|^2}{2} + f(x,d)\\ &\ge \frac{\|v(x)\|^2}{2} + f(x,v(x))\\ &= \max\left\{\frac{\|v(x)\|^2}{2} + \langle JF(x)v(x), w\rangle \mid w \in G\right\}\\ &= \max\left\{\frac{\|v(x)\|^2}{2} + \langle JF(x)v(x), w\rangle \mid w \in \tilde{G}\right\}.\end{aligned}$$

Second, if d is optimal solution of (4-1), then $\hat{d} = v(x)$. Taking in account

Remark 4.3 again, we get, for all $d \in \mathbb{R}^n$, that

$$\begin{aligned}
\frac{\|\hat{d}\|^2}{2} + f(x, \hat{d}) &= \max\left\{\frac{\|\hat{d}\|^2}{2} + \langle JF(x)\hat{d}\,,\, w\rangle \mid w \in G\right\} \\
&= \max\left\{\frac{\|\hat{d}\|^2}{2} + \langle JF(x)\hat{d}\,,\, w\rangle \mid w \in \tilde{G}\right\} \\
&\leq \max\left\{\frac{\|d\|^2}{2} + \langle JF(x)d\,,\, w\rangle \mid w \in \tilde{G}\right\} \\
&= \frac{\|d\|^2}{2} + f(x, d).
\end{aligned}$$

Then, $\hat{d} = v(x)$ because Problem (1-1) has only one optimal solution.

The dual problem of (4-1) is

$$\max\left\{\min\left\{\frac{\|d\|^2}{2} + \langle JF(x)d\,,\, w\rangle \mid d \in \mathbb{R}^n\right\} \mid w \in \tilde{G}\right\}. \tag{4-2}$$

Problems (4-1) and (4-2) are convex and Problem (4-1) has optimal solution. Then, it does not exist a duality gap. Moreover, since

$$\min\left\{\frac{\|d\|^2}{2} + \langle JF(x)d\,,\, w\rangle \mid d \in \mathbb{R}^n\right\} = -\frac{\|JF(x)^\top w\|^2}{2}$$

for $x \in \mathbb{R}^n$ and $w \in \tilde{G}$ given, (4-2) is equivalent to

$$\max\left\{-\frac{\|JF(x)^\top w\|^2}{2} \mid w \in \tilde{G}\right\}. \tag{4-3}$$

**Definition 4.4.** *Given $\sigma \in [0, 1)$, we say that $d$ is a $\sigma$-approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$ if*

$$\frac{\|d\|^2}{2} + f(x, d) \leq (1 - \sigma)\theta(x). \tag{4-4}$$

Observe that $v(x)$ is the 0-approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$. Another immediate consequence of the definition is that every $\sigma$-approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$ is also $\mathcal{K}$-descent direction for $F$ at $x$.

## 4.2 How to compute a $\sigma$-$\mathcal{K}$-descent direction

In this section will be present a practical way of how to compute a $\sigma$-Approximate $\mathcal{K}$-steepest descent direction. Thus, it becomes possible to perform numerical experimentation.

From now on, we assume that

$$G = \{w_1, \ldots, w_\ell\},$$

i.e., $\mathcal{K}^*$ and (consequently) $\mathcal{K}$ are finitely generated cones.

We need to introduce some additional notations. In the corresponding linear space, $e$ identifies the vector which has all its components equal to 1, and $e_k$ identifies the vector with all components equal to zero except the $k$-th, which is equal to 1. For $y \in \mathbb{R}^\ell \times \mathbb{R}$ given, $\tilde{y}$ is the projection of $y$ onto $\mathbb{R}^\ell$. Matrix

$$W = [w_1 \ldots\ldots w_\ell]$$

is $m \times \ell$ and $B = W^\top JF(x)JF(x)^\top W$ is $\ell \times \ell$, then

$$A = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix} \quad \textit{and} \quad M = \begin{pmatrix} -B & -e \end{pmatrix}$$

are $(\ell+1) \times (\ell+1)$ and $\ell \times (\ell+1)$, respectively. The simplex in $\mathbb{R}^\ell$ will be denoted by $\mathcal{S}$, i.e.,

$$\mathcal{S} = \left\{ y \in \mathbb{R}^\ell \mid \sum_{i=1}^{\ell} y_i = 1, \quad y_1 \geq 0, \ldots, y_\ell \geq 0 \right\}$$

and, finally, for a given $\sigma \in [0, 1)$,

$$\mathcal{D} = \left\{ y \in \mathbb{R}^\ell \times \mathbb{R} \quad \Bigg| \quad \begin{array}{rcl} (1 - \sigma/2) y^\top A y + \langle e_{\ell+1} \,,\, y \rangle & \leq & 0 \\ My & \leq & 0 \\ \tilde{y} & \in & \mathcal{S} \end{array} \right\}.$$

As the most important consequence of our assumption, there exists $\tilde{y} \in \mathcal{S}$ such that $v(x) = -JF(x)^\top W \tilde{y}$. Then, for any given $\sigma \in [0, 1)$, $\hat{y} = (\tilde{y} \,,\, f(x, v(x)))^\top \in \mathcal{D}$ because

$$M\hat{y} = -B\tilde{y} - f(x, v(x))e = W^\top JF(x)v(x) - f(x, v(x))e \leq 0$$

and

$$(1-\sigma/2)\hat{y}^\top A\hat{y} + \langle e_{\ell+1} \,,\, \hat{y} \rangle = \frac{\|v(x)\|^2}{2} + f(x, v(x)) + \frac{1-\sigma}{2}\|v(x)\|^2 = -\sigma\|v(x)\|^2/2 \leq 0.$$

Obviously, $\mathcal{D}$ is closed and convex. We claim that $\mathcal{D}$ is compact. Indeed, compactness of $\mathcal{S}$ implies that $B(\mathcal{S}) = \{Bz \mid z \in \mathcal{S}\}$ is compact. Observe that $y^\top A y =$

$\|JF(x)^\top W\tilde{y}\|^2 \geq 0$. Then, for any $y \in \mathcal{D}$ it holds that $-a \leq y_{l+1} \leq 0$, where

$$a = \max\left\{\max\{y_i \mid y \in B(\mathcal{S})\} \mid i = 1, \ldots, \ell\right\}.$$

Hence,

$$\mathcal{D} \subset \mathcal{S} \times [-a, 0].$$

The next lemma shows how to calculate the $\mathcal{K}$-steepest descent direction for $F$ at $x$ when $\mathcal{K}$ is finitely generated.

**Lemma 4.5.** *Consider problem*

$$(AP) \qquad \min \frac{1}{2} y^\top A y + \langle e_{\ell+1}, y\rangle \quad s.t. \quad My \leq 0, \quad \tilde{y} \in \mathcal{S}.$$

- *(a) Take $y^* \in \mathbb{R}^{\ell+1}$ such that $\tilde{y^*} \in \mathcal{S}$, $v(x) = -JF(x)^\top W\tilde{y^*}$, and $y_{\ell+1} = f(x, v(x))$. Then, $y^*$ is optimal solution of (AP).*
- *(b) If $y^*$ is optimal solution of (AP), then $v(x) = -JF(x)^\top W\tilde{y^*}$, and $y_{\ell+1} = f(x, v(x))$.*

*Proof.* For item (a), observe that

$$\begin{aligned} My^* &= -f(x, v(x))e + W^\top JF(x)[-JF(x)^\top W\tilde{y^*}] && \text{(4-5)} \\ &= -f(x, v(x))e + \begin{pmatrix} \langle w_1, JF(x)v(x)\rangle \\ \vdots \\ \langle w_\ell, JF(x)v(x)\rangle \end{pmatrix} && \text{(4-6)} \\ &\leq 0. && \text{(4-7)} \end{aligned}$$

Then, $y^*$ is solution of (AP) and, since

$$\frac{1}{2}(y^*)^\top A y^* + \langle e_{\ell+1}, y^*\rangle = \frac{\|v(x)\|^2}{2} + f(x, v(x)),$$

$y^*$ is optimal solution of (AP). For item (b), denote $d^* = -JF(x)^\top W\tilde{y^*}$. Observe that $y^*_{\ell+1} = f(x, d^*)$. Then,

$$\frac{1}{2}(y^*)^\top A y^* + y^*_{\ell+1} = \frac{1}{2}\|d^*\|^2 + f(x, d^*) \leq \frac{1}{2}\|v(x)\|^2 + f(x, v(x)).$$

Since $v(x)$ is the only one $\mathcal{K}$-steepest descent direction for $F$ at $x$, we get $d^* = v(x)$. □

**Lemma 4.6.** *Take $\sigma \in [0, 1)$ and $y \in \mathcal{D}$ such that $y_{\ell+1} = f(x, d)$, where $d = -JF(x)^\top W\tilde{y}$. Then, $d$ is a $\sigma$- approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$.*

*Proof.* First, observe that $W\tilde{y} \in \tilde{G}$ and therefore,

$$-\frac{\|JF(x)^\top W\tilde{y}\|^2}{2} = -\frac{\|d\|^2}{2} \leq \theta(x)$$

because (4-3) is the dual problem of (1-1). Then,

$$-\frac{(1-\sigma)}{2}\|d\|^2 \leq (1-\sigma)\theta(x).$$

Second, it holds that

$$(1-\sigma/2)y^\top Ay + \langle e_{\ell+1}\,,\, y\rangle = \frac{\|d\|^2}{2} + f(x,d) + \frac{1-\sigma}{2}\|d\|^2 \leq 0.$$

Consequently,

$$\frac{\|d\|^2}{2} + f(x,d) \leq -\frac{1-\sigma}{2}\|d\|^2 \leq (1-\sigma)\theta(x),$$

i.e., $d$ is $\sigma$–approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$. □

Now, we present our main contribution: a practical way to compute $\sigma$–approximate $\mathcal{K}$-steepest descent directions for $F$ at $x$.

Take $y^* \in S$ such that $(1-\sigma/2)\|JF(x)^\top W\lambda\|^2 + f(x, -JF(x)^\top W\lambda) \leq 0$. Then $d = -JF(x)^\top W\lambda$ is $\sigma-$approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$.

## 4.3 Computational experiments

In this section we will present the well-known Armijo line search along $\sigma$-approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$, the Algorithm and its Convergence Results. Moreover, present the numerical experiments that show the efficiency and robustness of an $\sigma$–approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$.

### 4.3.1 An algorithm using an Armijo-type line-search along $\sigma$-approximate $K$–steepest descent direction

In this section we will recall an algorithm proposed by Graña Drummond and Svaiter, in [28], and the corresponding convergence analyses.

Let $i_k$ be defined as

$$i_k = \min\{i \in \mathbb{N} \mid F(x + 2^{-i}d) \preceq_{\mathcal{K}} F(x) + 2^{-i}\beta JF(x)d\}. \tag{4-8}$$

Function $i_k$ defines a line-search of Armijo-type. Proposition 4.1 assures that if $d$ is a $\mathcal{K}$-descent direction for $F$ at $x$ then there exists $i_k$. So the following algorithm is well-defined.

**Algorithm 4.7.** *Consider two exogenous constants: $\sigma$ and $\beta \in (0,1)$.*

0. **Initialization**: *Choose $x^0 \in \mathbb{R}^n$. Compute $d(x^0)$ and initialize $k \leftarrow 0$.*
1. **Stopping criterium**: *If $d(x^k) = 0$, then STOP.*
2. **Direction**: *Compute $d^k$, a $\sigma$-$\mathcal{K}$-descent direction for $F$ at $x^k$.*
3. **Line search**: *Compute $i_k$ as in (4-8) and define the steplength $t_k = 2^{-i_k}$.*
4. **Iteration step**: *Set*
$$x^{k+1} = x^k + t_k d^k$$
*and $k \leftarrow k+1$. GOTO Step 1.*

In what follows, $\{x^k\}$, $\{d^k\}$ and $\{t_k\}$ are the sequences generated by Algorithm 4.7. If it stops at some iteration $k$, then $x^k$ is $\mathcal{K}$-critical for $F$ and it was successful. Let us assume then, that the algorithm does not stop, i.e. $\{x^k\}$, $\{d^k\}$ and $\{t_k\}$ are infinite sequences. Regarding convergence we have the following results.

**Lemma 4.8.** *If $\bar{x}$ is an accumulation point of $\{x^k\}$ then $F(\bar{x}) \preceq_{\mathcal{K}} F(x^k)$ for all $k$ and $\lim_{k\to\infty} F(x^k) = F(\bar{x})$. In particular, $F$ is constant in the set of accumulation points of $\{x^k\}$.*

*Proof.* See Proposition 2.2 in [28]. □

**Lemma 4.9.** *If there exists $\mathcal{F} \prec_{\mathcal{K}} F(x^k)$ for all $k$ then*

$$\sum t_k |\theta(x^k)| < \infty \qquad \text{and} \qquad \sum t_k \|d^k\|^2 < \infty.$$

*Proof.* By assumption, $\{x^k\}$ is infinite sequence, then

$$\frac{\|d^k\|^2}{2} + f(x^k, d^k) \leq (1-\sigma)\theta(x^k) < 0$$

and

$$F(x^{k+1}) \preceq_{\mathcal{K}} F(x^k) + \beta t_k JF(x^k) d^k \preceq_{\mathcal{K}} F(x^k)$$

for any $k$. Take some $k \geq 0$. Using the properties of $\varphi$ and the definition of $f$, we get

$$\begin{aligned}
\varphi(\mathcal{F}) &\leq \varphi(F(x^{k+1})) \leq \varphi(F(x^k)) + \beta t_k \varphi(JF(x^k)d^k) \leq \varphi(F(x^k)) + \beta t_k f(x^k, d^k) \\
&= \varphi(F(x^k)) + \beta t_k \left[ f(x^k, d^k) + \tfrac{\|d^k\|^2}{2} - \tfrac{\|d^k\|^2}{2} \right] \\
&\leq \varphi(F(x^k)) + \beta t_k \left[ (1-\sigma)\theta(x^k) - \tfrac{\|d^k\|^2}{2} \right]
\end{aligned}$$

because $d^k$ is $\sigma$-$\mathcal{K}$-approximate steepest descent direction for $F$ at $x^k$. Then,

$$\begin{aligned}\varphi(\mathcal{F}) &\leq \varphi(F(x^0)) + \beta \sum_{s=0}^{k} t_s \left[(1-\sigma)\theta(x^s) - \frac{\|d^s\|^2}{2}\right] \\ &= \varphi(F(x^0)) - \beta \sum_{s=0}^{k} t_s \left[(1-\sigma)|\theta(x^s)| + \frac{\|d^s\|^2}{2}\right].\end{aligned}$$

In other words,

$$\begin{aligned}\sum_{s=0}^{k} t_s \left[(1-\sigma)|\theta(x^s)| + \frac{\|d^s\|^2}{2}\right] &= (1-\sigma)\sum_{s=0}^{k} t_s|\theta(x^s)| + \frac{1}{2}\sum_{s=0}^{k} t_s\|d^s\|^2 \\ &\leq \frac{\varphi(F(x^0)) - \varphi(\mathcal{F})}{\beta},\end{aligned}$$

for all $k$, and the conclusion follows. □

**Theorem 4.10.** *If there exists $\mathcal{F} \prec_{\mathcal{K}} F(x^k)$ for all, $k$ then any accumulation point of $\{x^k\}$ is $\mathcal{K}$-critical.*

*Proof.* See Theorem 4.2 in [28]. □

The results below there are for case in that $F$ is $\mathcal{K}$-convex.

**Theorem 4.11.** *Suppose that $F$ is $\mathcal{K}$-convex and that $d^k$ is $\sigma-$approximate $\mathcal{K}-$steepest descent direction for $F$ at $x$. If $F(\hat{x}) \preccurlyeq_{\mathcal{K}} F(x^k)$ then*

$$\|\hat{x} - x^{k+1}\|^2 \leq \|\hat{x} - x^k\|^2 + \|x^{k+1} - x^k\|^2.$$

*Proof.* See Lemma 6.1 in [28]. □

**Theorem 4.12.** *Suppose that $F$ is $\mathcal{K}$-convex and there exists $\mathcal{F} \prec_{\mathcal{K}} F(x^k)$ for all $k$. Then, $\{x^k\}$ converges to a $\mathcal{K}$-critical point $x^*$.*

*Proof.* See Theorem 6.3 in [28]. □

### 4.3.2 Computational experiments

We will now present the numerical experiments that show the efficiency and robustness of a $\sigma$–approximate $\mathcal{K}$-steepest descent direction for $F$ at $x$. The implementation was divided into three groups of problems. The first formed by simple convex examples, the second formed by more elaborate convex examples and a third group formed by nonconvex examples. This way we could see the behavior of the algorithm in several examples with different structures.

The specifications of program, computer, stopping criteria and the maximum number of iterations are the same presented in Chapter 2 and therefore will be omitted here.

To calculate the $\sigma$–approximate $\mathcal{K}$-steepest descent direction, described in a practical way in the previous section, we created a subroutine using the Conditional Gradient Method-CGM. So the subroutine stops when we find a $y$ satisfying (4-4) or when it satisfies inequality $S(x^k) = 0$. We will continue presenting this algorithm which can be found in [3].

We created the diagram 4.1 to illustrate how Algorithm 4.7 was implemented with the search $\sigma$–approximate. Note that rectangles are part of the routine of Algorithm 4.7 and the circles refer to the subroutine referring to algorithm 4.13.

**Algorithm 4.13.** *Start with* $x^0 \in S \subset \mathbb{R}^n$. *Generate the sequence* $\{x^k\}$, $\forall k = 1, 2, \ldots$ *via the following steps:*

1. *Compute*
$$p^k = argmin\{\langle p - x^k, \bigtriangledown f(x^k)\rangle : p \in S\}. \tag{4-9}$$
2. *Stopping Criteria: Let*
$$S(x) := \min_{p \in S} \langle p - x, \bigtriangledown f(x)\rangle. \tag{4-10}$$
*If* $S(x^k) = \langle p^k - x^k, \bigtriangledown f(x^k)\rangle = 0$, *STOP. Else, goto Step 3.*
3. *Line search: Compute*
$$\lambda^{k-1} = argmin\ f(x^k + \lambda(p^k - x^k)). \tag{4-11}$$
*Update* $x^{k+1} = x^k + \lambda^k(p^k - x^k)$.
4. *Set* $k \leftarrow k + 1$. *Goto Step 1.*

All problems were solved 300 times. Tables 4.1, 4.2 and 4.3 were mounted in blocks of four lines. The four lines are dedicated to the corresponding results. In the first line, "%" are the percentages of runs that reached a critical point. In the second line, "it", for the successful runs, displays the average of iteration's numbers. In the third line, "Itint" is the average of the internal interactions, this is, the subroutine that calculates the $\sigma$–approximate. Finally, in the fourth line of the "Time" block, we have the average time taken for the algorithm to find a critical point. The first columns of the tables are dedicated to the identification of the problems, the paper where we found it, "$n$" and "$m$" give the number of variables and objectives respectively and "$x^0$" represents the starting points from a uniform

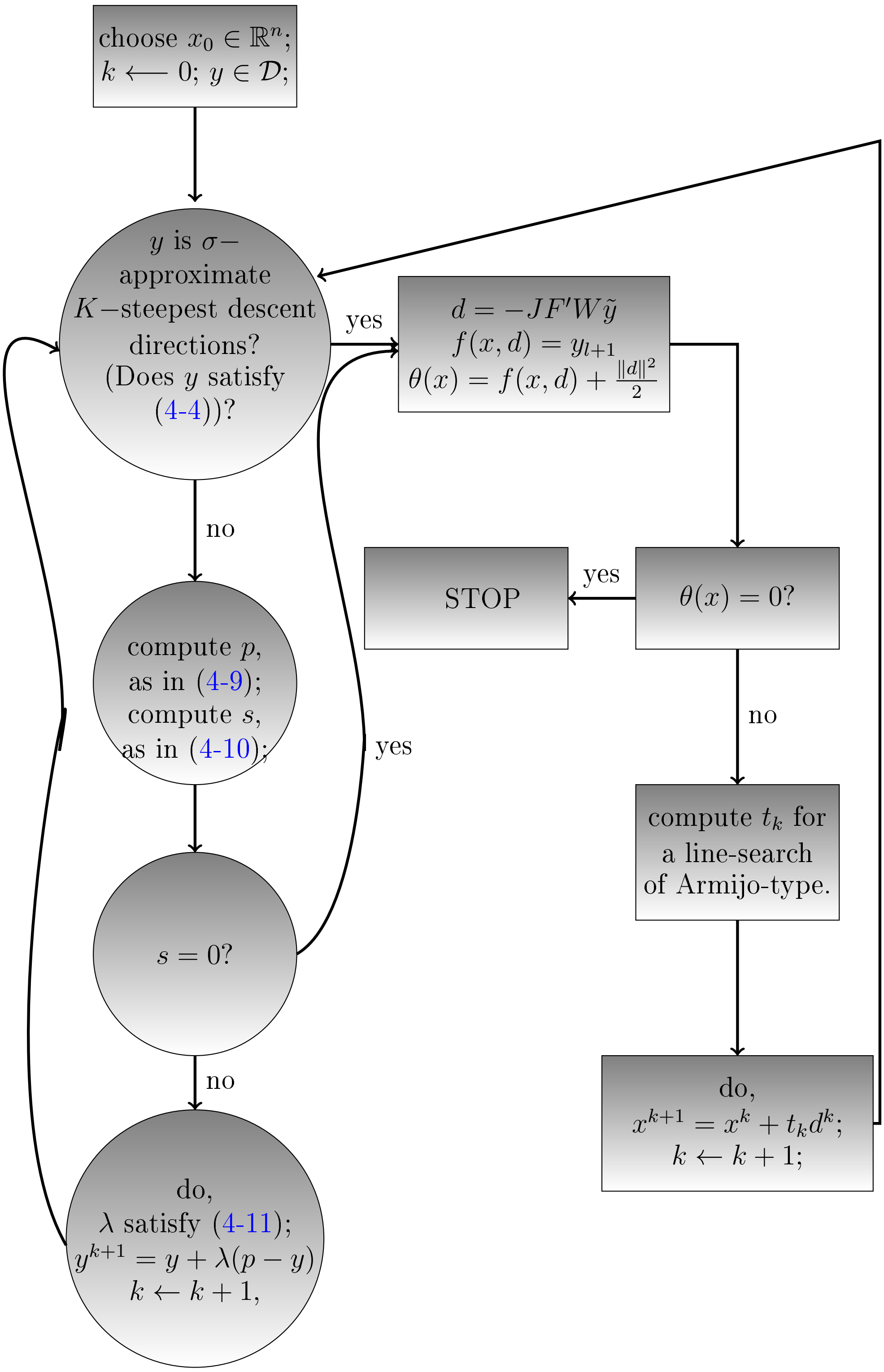



**Figure 4.1:** *Fluxogram of Algorithm 4.7.*

random distribution inside a box specified. The last ten columns are dedicated to the corresponding results with the variation of $\sigma$ between 0.0 and 0.9.

We did the performance profile over time to compare the exact search ($\sigma = 0$) with the search $\sigma$–approximate varying the $\sigma$ the same way as in the tables. For more details on performance profile, see [13].

We need to answer a question: The quality of Pareto points when using a $\sigma$–approximate direction is as good as when we use the exact direction? To answer this question let us present the purity metric and spread metrics graphs, which show the pareto curve quality. See [8, 22].

Furthermore, the direction $\sigma$–approximate gives us a $\theta(x)$ also approximate, thus, it is necessary to verify the quality of the critical points obtained by the algorithm. So, based on articles [8, 22] or Appendix A, we present the performance profile for the Purity metric, which compares the non-dominated points belonging to the Pareto front generated by each $\sigma$, compared two by two. In this way, the algorithm that presents less dominated points has a better performance. Graphs were generated comparing the exact direction "$\sigma = 0$" with the best performance in terms of time, that being $\sigma = 0.8$.

In order to evaluate the quality of pareto points, we will also present the "$\Gamma$–Spread metrics" and the "$\Delta$–Spread metrics". The "$\Gamma$–Spread metrics" compare the size of "holes" of pareto front, thus, the best pareto front is the one that presents minors "holes". The "$\Delta$–Spread metrics" measure how well the points are distributed over a pareto front.

### 4.3.3 First group of problems

Twelve simple convex examples form the first group of problems. So, it was relatively easy for the algorithm to find critical points. The numerical results appear in Table 4.1.

Figure 4.2 presents a performance profile, using time as criterion, comparing the results, obtained by the algorithm for different values of $\sigma$: from $\sigma = 0$, which corresponds to the use of exact $\mathcal{K}$-steepest descent direction, to $\sigma = 0.9$, which corresponds to values close to the maximum allowed for $\sigma$. This graphic suggests that, regarding computational time and considering only these ten values, the best option for $\sigma$ is 0.8.

In Figures 4.3, 4.4 and 4.5, we show a comparison of quality of the Pareto fronts computed by the algorithm when $\sigma = 0$ and when $\sigma = 0,8$. We did these comparisons by three different criteria: the purity metric in Figure 4.3, the $\Gamma$-spread metric in Figure 4.4 and the $\Delta$-spread metric in Figure 4.5.

| | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AP1, [1] | % | 99.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 369.85 | 315.11 | 335.58 | 259.95 | 246.10 | 250.85 | 205.24 | 192.75 | 165.22 | 160.83 |
| $m = 3$ | Itint | 1.00 | 0.79 | 0.75 | 0.77 | 0.76 | 0.71 | 0.72 | 0.73 | 0.72 | 0.74 |
| $x^0 \in [-10, 10]^n$ | Time | 2.0487 | 0.4882 | 0.3179 | 0.2001 | 0.1866 | 0.2058 | 0.1750 | 0.1626 | 0.1366 | 0.1263 |
| IKK1, [32] | % | 100.00 | 100.00 | 99.00 | 99.00 | 99.00 | 100.00 | 100.00 | 99.00 | 100.00 | 100.00 |
| $n = 2$ | It | 134.89 | 66.11 | 35.61 | 69.89 | 37.73 | 45.70 | 63.51 | 71.03 | 97.43 | 108.67 |
| $m = 3$ | Itint | 1.54 | 1.25 | 1.12 | 0.98 | 1.02 | 0.96 | 0.88 | 0.76 | 0.75 | 0.68 |
| $x^0 \in [-50, 50]^n$ | Time | 0.0476 | 0.0379 | 0.0274 | 0.0279 | 0.0254 | 0.0246 | 0.0251 | 0.0237 | 0.0248 | 0.0225 |
| JOS1,[35] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 100$ | It | 510.47 | 510.27 | 510.11 | 510.30 | 510.01 | 510.35 | 510.40 | 510.00 | 510.00 | 509.61 |
| $m = 2$ | Itint | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $x^0 \in [-100, 100]^n$ | Time | 0.0674 | 0.0693 | 0.0712 | 0.0689 | 0.0709 | 0.0657 | 0.0711 | 0.0729 | 0.0676 | 0.0676 |
| Lov1, [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 3.19 | 3.21 | 3.16 | 3.10 | 3.13 | 2.98 | 2.92 | 2.93 | 2.85 | 2.85 |
| $m = 2$ | Itint | 1.25 | 1.13 | 1.09 | 1.01 | 0.93 | 0.85 | 0.82 | 0.79 | 0.78 | 0.77 |
| $x^0 \in [-10, 10]^n$ | Time | 0.0450 | 0.0416 | 0.0398 | 0.0365 | 0.0336 | 0.0297 | 0.0288 | 0.0257 | 0.0239 | 0.0278 |
| MHHM2, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 | 1.00 |
| $m = 3$ | Itint | 0.52 | 0.50 | 0.23 | 0.14 | 0.09 | 0.05 | 0.06 | 0.04 | 0.03 | 0.02 |
| $x^0 \in [0, 1]^n$ | Time | 0.0083 | 0.0089 | 0.0041 | 0.0027 | 0.0018 | 0.0012 | 0.0012 | 0.0010 | 0.0007 | 0.0007 |
| MOP7, [32] | % | 93.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 149.94 | 142.75 | 139.58 | 136.73 | 128.15 | 124.06 | 124.07 | 108.74 | 113.98 | 109.69 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $m = 3$ | Itint | 0.30 | 0.29 | 0.28 | 0.25 | 0.28 | 0.27 | 0.26 | 0.28 | 0.26 | 0.26 |
| $x^0 \in [-400, 400]^n$ | Time | 0.3336 | 0.3331 | 0.3105 | 0.2708 | 0.2749 | 0.2592 | 0.2450 | 0.2376 | 0.2153 | 0.2152 |
| PNR, [48] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 5.52 | 5.46 | 6.14 | 6.45 | 6.48 | 6.66 | 5.65 | 5.49 | 5.26 | 5.94 |
| $m = 2$ | Itint | 0.70 | 0.70 | 0.72 | 0.73 | 0.73 | 0.74 | 0.70 | 0.68 | 0.67 | 0.69 |
| $x^0 \in [-2, 2]^n$ | Time | 0.0528 | 0.0504 | 0.0534 | 0.0564 | 0.0549 | 0.0577 | 0.0490 | 0.0464 | 0.0442 | 0.0501 |
| SD, [52] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 4$ | It | 18.69 | 18.45 | 18.32 | 18.84 | 18.39 | 18.53 | 18.63 | 18.03 | 18.78 | 18.50 |
| $m = 2$ | Itint | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 | 0.98 | 0.97 | 0.97 |
| $x^0 \in [1,3] \times [\sqrt{2}, 3] \times [\sqrt{2}, 3] \times [1,3]^n$ | Time | 0.2077 | 0.1841 | 0.1862 | 0.1911 | 0.1855 | 0.1894 | 0.1826 | 0.1819 | 0.1792 | 0.1803 |
| SLC2, [50] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 100$ | It | 30.59 | 28.94 | 28.60 | 30.85 | 29.64 | 30.21 | 33.55 | 32.38 | 32.20 | 32.18 |
| $m = 2$ | Itint | 1.02 | 0.94 | 0.88 | 0.82 | 0.83 | 0.82 | 0.78 | 0.77 | 0.77 | 0.76 |
| $x^0 \in [-10, 10]^n$ | Time | 0.3021 | 0.2278 | 0.2118 | 0.1872 | 0.1988 | 0.2242 | 0.2262 | 0.1897 | 0.2104 | 0.2149 |
| SLCDT2, [51] | % | 95.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 10$ | It | 8.83 | 8.77 | 8.69 | 8.70 | 9.10 | 8.96 | 9.04 | 9.06 | 8.84 | 9.06 |
| $m = 3$ | Itint | 1.16 | 1.14 | 1.12 | 1.10 | 1.08 | 1.06 | 1.05 | 1.04 | 1.02 | 1.01 |
| $x^0 \in [-1, 1]^n$ | Time | 0.1110 | 0.2167 | 0.1623 | 0.2016 | 0.1881 | 0.2094 | 0.1839 | 0.1525 | 0.1779 | 0.1668 |
| SP1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 11.96 | 12.50 | 11.83 | 12.49 | 11.93 | 12.41 | 11.90 | 12.28 | 12.37 | 12.44 |
| $m = 2$ | Itint | 1.51 | 1.16 | 1.03 | 0.92 | 0.88 | 0.82 | 0.78 | 0.72 | 0.69 | 0.63 |
| $x^0 \in [-100, 100]^n$ | Time | 0.1657 | 0.1304 | 0.1072 | 0.0984 | 0.0920 | 0.0844 | 0.0804 | 0.0740 | 0.0632 | 0.0656 |
| Toi4, [53] | % | 94.00 | 95.00 | 96.00 | 96.00 | 98.00 | 96.00 | 97.00 | 97.00 | 98.00 | 98.00 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 4$ | It | 805.55 | 580.95 | 515.90 | 429.19 | 320.38 | 487.33 | 307.06 | 438.74 | 264.49 | 348.37 |
| $m = 2$ | Itint | 0.16 | 0.09 | 0.09 | 0.08 | 0.06 | 0.04 | 0.04 | 0.03 | 0.02 | 0.02 |
| $x^0 \in [-100, 100]^n$ | Time | 0.0699 | 0.0480 | 0.0463 | 0.0358 | 0.0288 | 0.0394 | 0.0242 | 0.0332 | 0.0203 | 0.0270 |

**Table 4.1:** *Convex Problem; Group I*

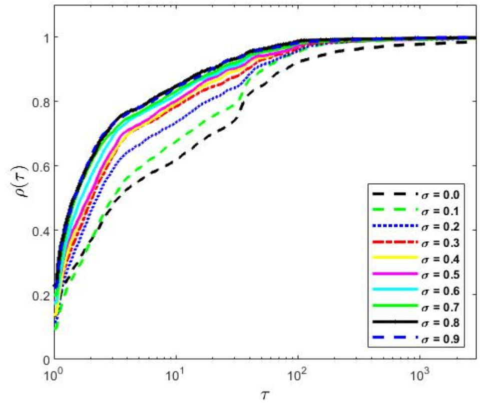



**Figure 4.2:** *Performance profile-Time*

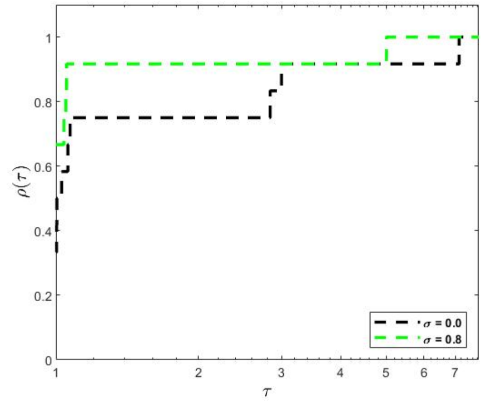



**Figure 4.3:** *Purity metric-G1*

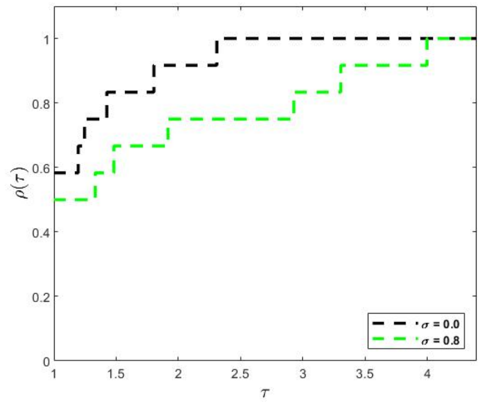



**Figure 4.4:** $\Gamma$*–Spread metric*

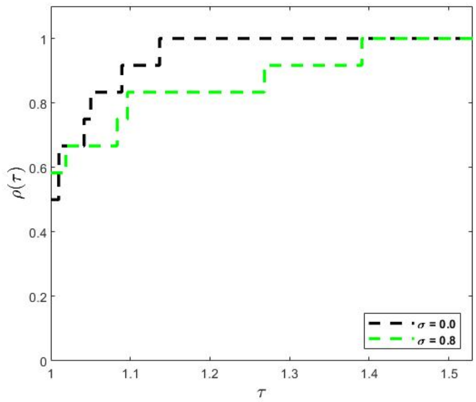



**Figure 4.5:** $\Delta$*–Spread metric*

Looking at Table 4.1 we see that in general directions $\sigma$–approximate have a more satisfying performance than exact direction ($\sigma = 0$), moreover, we also conclude that as “$\sigma$” increases, the results improve, showing that the values 0.8 and 0.9 are the best results. In this group of examples, almost all the problems were solved, this is, in almost 100% of the 300 times, each problem was compiled until the algorithm stopped at a critical point. Iterations “It” of the algorithm decrease as the $\sigma$ increases, and the same happens to internal iterations “Itint” of the routine, which give the direction $\sigma$–approximate, and the time “Time”, that represents the time spent for the algorithm to find a critical point."Time", that represents the time spent for the algorithm to find a critical point.

We can also see in the graph of performance profile (Figure 4.2) comparing time, that both robustness and effectiveness of the search $\sigma$–approximate are better, also having a better prominence for the $\sigma$ with higher values.

Figure 4.3 shows that the 0.8-approximate direction is better than the exact $\mathcal{K}$-steepest descent direction in the sense that it generated fewer dominated points.

On the contrary, Figure 4.4 shows that the exact direction is better than the approximate direction because the computed Pareto set has, in this case, fewer "holes".

Figure 4.5 shows us that, according to the $\Delta-$Sprenad metric criterion, the generation of the Pareto set, when using the exact direction, is more robust than when using the 0.8-approximate direction. On the other hand, the generation of the Pareto set, when using the 0.8-approximate direction, is more efficient than when using the exact direction.

### 4.3.4 Second group of problems

The second group of problems is formed by six convex problems more elaborate. Table 4.2 presents information of this group of problems.

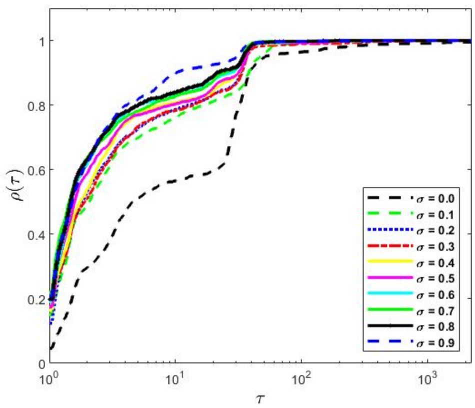



**Figure 4.6:** *Time*

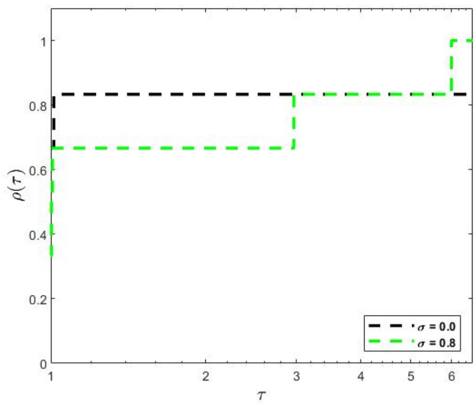



**Figure 4.7:** *Purity metric*

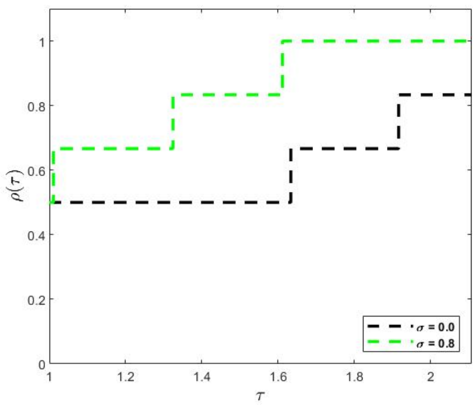



**Figure 4.8:** $\Gamma$–*Spread metric*

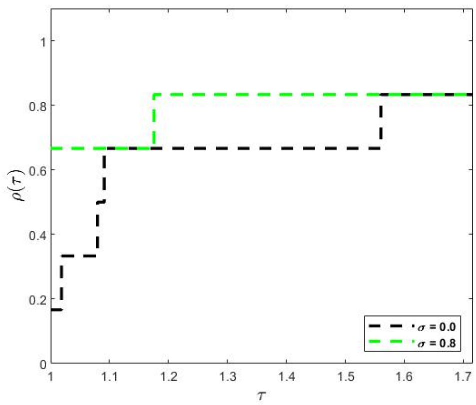



**Figure 4.9:** $\Delta$–*Spread metric*

| | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AP4, [1] | % | 99.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 3$ | It | 397.32 | 354.68 | 316.00 | 386.89 | 322.27 | 253.82 | 222.02 | 217.02 | 209.02 | 145.53 |
| $m = 3$ | Itint | 1.13 | 0.87 | 0.88 | 0.88 | 0.87 | 0.81 | 0.80 | 0.74 | 0.73 | 0.77 |
| $x^0 \in [-10, 10]^n$ | Time | 3.9611 | 1.1299 | 0.7104 | 1.4628 | 0.8469 | 0.7314 | 0.4491 | 0.3507 | 0.4005 | 0.5194 |
| BK1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $m = 2$ | Itint | 0.50 | 0.47 | 0.44 | 0.41 | 0.37 | 0.35 | 0.31 | 0.31 | 0.28 | 0.20 |
| $x^0 \in [-5, 10]^n$ | Time | 0.0084 | 0.0089 | 0.0069 | 0.0066 | 0.0059 | 0.0061 | 0.0052 | 0.0050 | 0.0046 | 0.0033 |
| FDS, [20] | % | 98.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 5$ | It | 88.10 | 88.69 | 86.18 | 87.27 | 82.61 | 79.83 | 75.16 | 88.66 | 80.70 | 78.92 |
| $m = 3$ | Itint | 1.03 | 0.99 | 0.97 | 0.95 | 0.94 | 0.92 | 0.91 | 0.89 | 0.88 | 0.85 |
| $x^0 \in [-2, 2]^n$ | Time | 1.0638 | 0.9437 | 0.8986 | 0.8942 | 0.8147 | 0.7897 | 0.7376 | 0.8696 | 0.7750 | 0.7452 |
| MGH33, [43] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 10$ | It | 2.92 | 2.80 | 2.87 | 2.89 | 2.85 | 2.90 | 2.88 | 2.91 | 2.83 | 2.96 |
| $m = 10$ | Itint | 0.82 | 0.81 | 0.79 | 0.79 | 0.80 | 0.79 | 0.76 | 0.74 | 0.76 | 0.71 |
| $x^0 \in [-1, 1]^n$ | Time | 0.0244 | 0.0224 | 0.0228 | 0.0226 | 0.0226 | 0.0237 | 0.0215 | 0.0220 | 0.0210 | 0.0209 |
| Toi8, [53] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 3.01 | 2.78 | 2.95 | 3.77 | 3.98 | 4.50 | 4.21 | 4.26 | 4.41 | 3.16 |
| $m = 2$ | Itint | 0.69 | 0.66 | 0.67 | 0.73 | 0.74 | 0.77 | 0.74 | 0.74 | 0.73 | 0.61 |
| $x^0 \in [-10, 10]^n$ | Time | 0.0379 | 0.0329 | 0.0311 | 0.0440 | 0.0460 | 0.0512 | 0.0478 | 0.0473 | 0.0468 | 0.0248 |
| ZLT1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 10$ | It | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $m = 5$ | Itint | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $x^0 \in [-1000, 1000]^n$ | Time | 0.0093 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0005 |

**Table 4.2:** *Convex Problem; Group II*

Looking at Table 4.2, we see the same characteristics of group 1 of problems, this is, almost all examples are resolved 100% of the 300 times they have been compiled, and as $\sigma$ increases, iterations "It", internal iterations "itint", and time "Time" are decreasing, respectively.

Figure 4.6 shows that the biggest $\sigma$, "0.8" and "0.9", are the ones with the best performances. Figure 4.7 shows that the exact search ($\sigma = 0$) has fewer dominated points than the search $\sigma$–approximate. Figures 4.8 and 4.9 show that both in $\Gamma$–Spread metric and $\Delta$–Spread metric, the $\sigma$–approximate has better performance.

### 4.3.5 Third group of problems

The third group of problems is formed by twenty two non-convex problems. Tables 4.3 presents information of this group of problems.

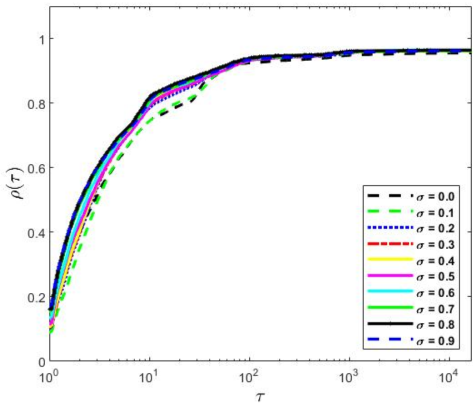



**Figure 4.10:** *Performace Profile-Time*

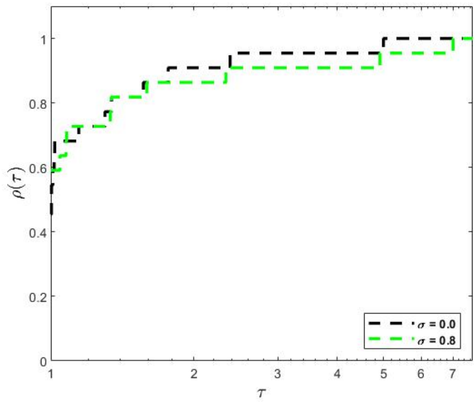



**Figure 4.11:** *Purity metric-NC*

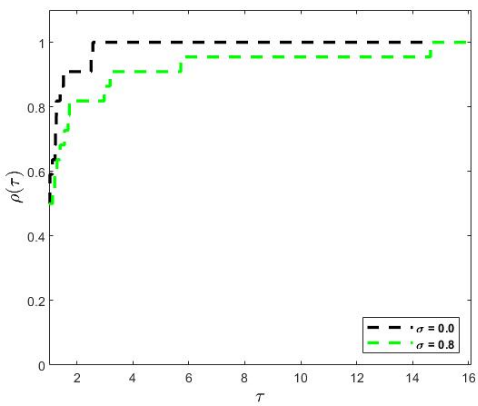



**Figure 4.12:** $\Gamma$-*Spread metric*

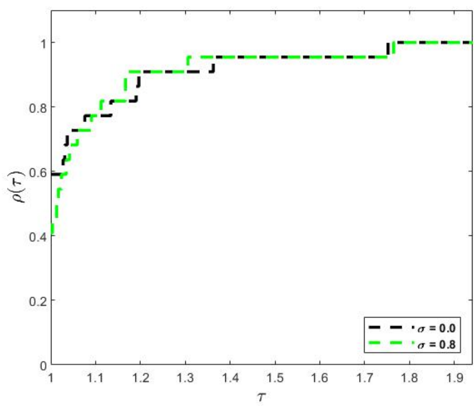



**Figure 4.13:** $\Delta$-*Spread metric*

| | | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| AP3,[1] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 75.25 | 75.78 | 73.41 | 71.38 | 67.37 | 67.41 | 61.93 | 56.48 | 48.52 | 41.16 |
| $m = 2$ | Itint | 0.43 | 0.32 | 0.29 | 0.28 | 0.30 | 0.27 | 0.26 | 0.24 | 0.29 | 0.31 |
| $x^0 \in [-1, 1]^n$ | Time | 0.2008 | 0.1373 | 0.1005 | 0.1086 | 0.1130 | 0.0926 | 0.0837 | 0.0629 | 0.0602 | 0.0608 |
| DD1, [10] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 5$ | It | 62.09 | 61.81 | 62.66 | 56.38 | 59.44 | 60.88 | 55.54 | 56.02 | 58.87 | 59.30 |
| $m = 2$ | Itint | 1.02 | 1.01 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 |
| $x^0 \in [-20, 20]^n$ | Time | 0.7591 | 0.6917 | 0.7226 | 0.6283 | 0.6556 | 0.6787 | 0.6332 | 0.6536 | 0.6710 | 0.6834 |
| Far1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 64.89 | 58.82 | 57.59 | 64.47 | 60.91 | 65.58 | 56.61 | 56.99 | 56.60 | 55.97 |
| $m = 2$ | Itint | 0.90 | 0.91 | 0.89 | 0.83 | 0.84 | 0.87 | 0.82 | 0.83 | 0.76 | 0.78 |
| $x^0 \in [-1, 1]^n$ | Time | 0.6178 | 0.5635 | 0.5702 | 0.6525 | 0.5417 | 0.6068 | 0.4663 | 0.4752 | 0.4777 | 0.5733 |
| FF1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 32.67 | 34.67 | 31.92 | 33.99 | 33.45 | 32.85 | 34.96 | 31.81 | 34.51 | 33.47 |
| $m = 2$ | Itint | 1.00 | 0.99 | 0.95 | 0.96 | 0.96 | 0.94 | 0.94 | 0.91 | 0.93 | 0.93 |
| $x^0 \in [-1, 1]^n$ | Time | 0.3118 | 0.3796 | 0.3456 | 0.3808 | 0.3677 | 0.3505 | 0.3819 | 0.3367 | 0.3713 | 0.3585 |
| Hil1, [31] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 15.36 | 14.96 | 15.49 | 14.61 | 14.96 | 17.01 | 14.16 | 16.27 | 17.33 | 17.58 |
| $m = 2$ | Itint | 1.21 | 1.13 | 1.08 | 1.07 | 1.05 | 1.01 | 1.00 | 0.98 | 0.98 | 0.96 |
| $x^0 \in [0, 1]^n$ | Time | 0.1686 | 0.1617 | 0.1697 | 0.1486 | 0.1509 | 0.1948 | 0.1661 | 0.1791 | 0.2016 | 0.2026 |
| KW2, [36] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 107.03 | 94.60 | 86.63 | 119.99 | 139.14 | 111.19 | 105.81 | 91.24 | 103.89 | 85.68 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $m = 2$ | Itint | 0.92 | 0.91 | 0.87 | 0.80 | 0.79 | 0.80 | 0.79 | 0.79 | 0.77 | 0.79 |
| $x^0 \in [-3, 3]^n$ | Time | 0.5374 | 0.4775 | 0.4025 | 0.6046 | 0.7177 | 0.4563 | 0.4439 | 0.4392 | 0.4227 | 0.4285 |
| LE1, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 92.01 | 84.34 | 80.70 | 82.16 | 75.66 | 78.58 | 83.42 | 82.77 | 78.62 | 79.47 |
| $m = 2$ | Itint | 0.32 | 0.25 | 0.22 | 0.23 | 0.23 | 0.22 | 0.22 | 0.20 | 0.21 | 0.21 |
| $x^0 \in [-5, 10]^n$ | Time | 0.2781 | 0.1949 | 0.1513 | 0.1582 | 0.1562 | 0.1645 | 0.2139 | 0.1409 | 0.1454 | 0.1696 |
| Lov3, [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 5.37 | 81.23 | 78.11 | 82.14 | 77.58 | 76.44 | 82.01 | 83.11 | 82.29 | 80.71 |
| $m = 2$ | Itint | 0.99 | 0.25 | 0.24 | 0.25 | 0.23 | 0.24 | 0.20 | 0.21 | 0.20 | 0.19 |
| $x^0 \in [-20, 20]^n$ | Time | 0.0598 | 0.1573 | 0.1429 | 0.1565 | 0.1569 | 0.1501 | 0.1421 | 0.1747 | 0.1413 | 0.1362 |
| Lov4, [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 1.81 | 1.98 | 3.92 | 4.93 | 5.95 | 6.39 | 6.35 | 6.81 | 6.79 | 6.93 |
| $m = 2$ | Itint | 0.61 | 0.59 | 0.70 | 0.76 | 0.83 | 0.86 | 0.85 | 0.88 | 0.87 | 0.88 |
| $x^0 \in [-20, 20]^n$ | Time | 0.0163 | 0.0184 | 0.0365 | 0.0464 | 0.0588 | 0.0636 | 0.0643 | 0.0697 | 0.0646 | 0.0686 |
| Lov5, [38] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 3$ | It | 88.90 | 90.89 | 87.32 | 88.54 | 88.58 | 88.66 | 86.80 | 86.26 | 84.97 | 81.08 |
| $m = 2$ | Itint | 1.00 | 0.93 | 0.85 | 0.77 | 0.70 | 0.64 | 0.61 | 0.57 | 0.53 | 0.46 |
| $x^0 \in [-2, 2]^n$ | Time | 1.1990 | 1.1334 | 0.9108 | 0.8210 | 0.7580 | 0.6679 | 0.6119 | 0.5818 | 0.4642 | 0.4476 |
| LTDZ, [37] | % | 11.00 | 16.00 | 18.00 | 19.00 | 16.00 | 17.00 | 22.00 | 17.00 | 20.00 | 16.00 |
| $n = 3$ | It | 500.47 | 606.79 | 510.00 | 692.51 | 617.49 | 643.40 | 742.74 | 511.90 | 590.66 | 727.13 |
| $m = 3$ | Itint | 1.02 | 1.02 | 1.02 | 1.01 | 1.00 | 1.00 | 0.98 | 0.98 | 0.99 | 0.99 |
| $x^0 \in [0, 1]^n$ | Time | 5.6996 | 8.3519 | 6.1970 | 7.8335 | 5.4859 | 7.3747 | 7.0770 | 4.5815 | 5.0813 | 6.8475 |
| MGH16, [43] | % | 92.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n = 4$ | It | 67.13 | 72.17 | 77.45 | 76.71 | 76.16 | 89.17 | 89.91 | 82.53 | 93.26 | 93.15 |
| $m = 5$ | Itint | 1.10 | 1.05 | 1.05 | 1.06 | 1.02 | 0.98 | 0.94 | 0.94 | 0.93 | 0.92 |
| $x^0 \in [-25, 25] \times [-5, 5]^2 \times [-1, 1]^n$ | Time | 0.9752 | 0.8866 | 1.0009 | 1.3290 | 1.1231 | 1.2700 | 1.2989 | 1.1522 | 1.3020 | 1.2609 |
| MGH33, [43] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 10$ | It | 2.96 | 2.97 | 2.95 | 2.98 | 2.96 | 2.77 | 2.82 | 2.84 | 2.91 | 2.90 |
| $m = 10$ | Itint | 0.81 | 0.79 | 0.80 | 0.77 | 0.76 | 0.80 | 0.78 | 0.76 | 0.73 | 0.72 |
| $x^0 \in [-1, 1]^n$ | Time | 0.0246 | 0.0281 | 0.0275 | 0.0292 | 0.0268 | 0.0278 | 0.0258 | 0.0255 | 0.0254 | 0.0253 |
| MHHM2, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 1.00 | 0.99 | 0.98 | 0.99 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 0.99 |
| $m = 3$ | Itint | 0.50 | 0.49 | 0.23 | 0.15 | 0.11 | 0.05 | 0.06 | 0.04 | 0.02 | 0.03 |
| $x^0 \in [0, 1]^n$ | Time | 0.0082 | 0.0093 | 0.0041 | 0.0028 | 0.0024 | 0.0013 | 0.0013 | 0.0011 | 0.0007 | 0.0008 |
| MMR1, [41] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 7.72 | 7.67 | 20.13 | 6.20 | 10.28 | 11.23 | 10.99 | 11.47 | 6.55 | 11.62 |
| $m = 2$ | Itint | 1.10 | 1.04 | 1.01 | 1.00 | 0.99 | 0.96 | 0.97 | 0.95 | 0.94 | 0.91 |
| $x^0 \in [0.1, 1]^n$ | Time | 0.1015 | 0.1000 | 0.1052 | 0.0869 | 0.0920 | 0.0934 | 0.0881 | 0.0802 | 0.0859 | 0.0796 |
| MOP2, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 18.24 | 26.92 | 31.25 | 19.38 | 24.41 | 19.60 | 17.40 | 19.76 | 28.11 | 27.18 |
| $m = 2$ | Itint | 0.63 | 0.58 | 0.55 | 0.54 | 0.57 | 0.59 | 0.60 | 0.53 | 0.51 | 0.52 |
| $x^0 \in [-4, 4]^n$ | Time | 0.0459 | 0.0419 | 0.0442 | 0.0289 | 0.0341 | 0.0375 | 0.0352 | 0.0340 | 0.0370 | 0.0394 |
| MOP3, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 7.27 | 7.55 | 9.02 | 8.72 | 7.97 | 7.54 | 8.00 | 7.49 | 7.29 | 7.36 |
| $m = 2$ | Itint | 1.08 | 1.06 | 1.01 | 1.00 | 0.96 | 0.95 | 0.93 | 0.91 | 0.91 | 0.88 |
| $x^0 \in [-\pi, \pi]^n$ | Time | 0.0832 | 0.1200 | 0.1313 | 0.1222 | 0.1096 | 0.0995 | 0.1028 | 0.0967 | 0.0918 | 0.0896 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MOP5, [32] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 2.32 | 5.04 | 3.84 | 3.71 | 3.86 | 4.43 | 6.08 | 1.92 | 5.25 | 3.10 |
| $m = 3$ | Itint | 1.72 | 1.67 | 1.53 | 1.54 | 1.34 | 0.99 | 0.95 | 0.97 | 0.95 | 0.96 |
| $x^0 \in [-30, 30]^n$ | Time | 0.0158 | 0.0191 | 0.0160 | 0.0169 | 0.0141 | 0.0099 | 0.0108 | 0.0098 | 0.0104 | 0.0124 |
| QV1, [1] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 10$ | It | 700.33 | 723.20 | 709.67 | 692.11 | 693.37 | 674.43 | 670.82 | 675.58 | 659.79 | 631.78 |
| $m = 2$ | Itint | 0.99 | 0.95 | 0.92 | 0.88 | 0.86 | 0.83 | 0.80 | 0.78 | 0.74 | 0.74 |
| $x^0 \in [-5.12, 5.12]^n$ | Time | 8.3231 | 11.6993 | 10.4359 | 8.3772 | 9.5114 | 8.9112 | 8.4489 | 8.2915 | 7.7951 | 7.2378 |
| SK2, [1] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 4$ | It | 35.94 | 33.68 | 34.47 | 35.14 | 33.64 | 33.44 | 33.73 | 31.64 | 33.37 | 33.58 |
| $m = 2$ | Itint | 0.63 | 0.58 | 0.57 | 0.55 | 0.53 | 0.60 | 0.61 | 0.58 | 0.55 | 0.59 |
| $x^0 \in [-10, 10]^n$ | Time | 0.3448 | 0.3524 | 0.3806 | 0.3788 | 0.3483 | 0.3755 | 0.3668 | 0.3260 | 0.3531 | 0.3644 |
| SLCDT1, [51] | % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 2$ | It | 19.69 | 20.03 | 19.69 | 19.98 | 19.61 | 19.91 | 19.69 | 19.76 | 19.60 | 19.97 |
| $m = 2$ | Itint | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.04 |
| $x^0 \in [-5, 5]^n$ | Time | 0.0098 | 0.0096 | 0.0092 | 0.0093 | 0.0096 | 0.0095 | 0.0091 | 0.0090 | 0.0092 | 0.0097 |
| TKLY1, [32] | % | 99.00 | 99.00 | 99.00 | 100.00 | 99.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| $n = 4$ | It | 879.49 | 572.92 | 374.07 | 361.48 | 376.06 | 413.44 | 452.57 | 454.70 | 265.69 | 301.40 |
| $m = 2$ | Itint | 1.00 | 0.91 | 0.84 | 0.82 | 0.81 | 0.80 | 0.77 | 0.78 | 0.76 | 0.77 |
| $x^0 \in [0.1, 1] \times [0, 1]^3$ | Time | 10.2427 | 9.2630 | 5.6913 | 5.6817 | 5.8290 | 6.5136 | 7.2856 | 5.9682 | 2.8845 | 3.5755 |

**Table 4.3:** *Non-Convex Problem: Group III*

Table 4.3 shows us that with the exception of problem $LTDZ$, practically all the examples were 100% solved and the standard observed in the other two groups of problems remains, that is, as $\sigma$ increases, the performance of the algorithm improves, so much in iteration, internal iteration as in time. Problem $LTDZ$ does not perform well, the percentage of problems solved ranges from 11 to 20 percent, not having a standard in the other criteria. Figure 4.10, referring to Performance Profile and comparing the time we observe that the pattern of the two previous groups is maintained, this is, the $\sigma$'s largest value has the best performance in both efficiency and robustness. In Figure 4.11 the result is inconclusive, that is, we can not decide which algorithm has the least dominated points. The same is observed in the $\Delta$–Spread metric, having a slight tendency to say that in the $\Gamma$–spread metric the exact direction tends to perform better, that is, has "fewer holes".

### 4.3.6 Chapter conclusion

In summary, we observe that the performance of a $\sigma$–approximate $\mathcal{K}$-steepest descent direction is better than the exact direction for both convex and non-convex examples. And when we look at the quality of the critical points, it does not get worse when comparing a $\sigma$–approximate $\mathcal{K}$-steepest descent direction to an exact direction. Therefore we conclude that the way to calculate a $\sigma$–approximate $\mathcal{K}$-steepest descent direction presented in this chapter gives satisfactory results when compared to the exact direction.

The procedure presented in this chapter is not directly related to the linear search used, then it becomes very suggestive to use the way we calculated the $\sigma$–approximate $\mathcal{K}$-steepest descent direction with the new linear search introduced in Chapters 2 and 3. We intend to present these results in future works.

# Final remarks

This work was divided into basically two parts. In the first, a new way of computing the step length $\alpha_k$ was presented. This search is relevant because it uses only gradient information i.e., it does not work with functional values. Our numerical experiments suggest that it has better performance than other algorithms when evaluating the gradient is easier that evaluating the function itself. In Chapter 2, the new line-search procedure was used replacing the Wolf conditions in conjugate gradients algorithm and, in Chapter 3, the same procedure replaced Armijo-type search in the steepest descent algorithm. Numerical experiments were performed to test the efficiency and performance of our approach.

In the second part of the thesis - Chapter 4, we discussed a practical way of computing $\sigma$ – approximate $\mathcal{K}$ – steepest descent directions. We compared, also, the performance of descent algorithm when $\sigma$ assumes different values. Our numerical experiments suggest that $\sigma = 0.8$ has best performance without lost of quality of the generated Pareto front.

We can foresee three different continuations for our work.

First, we intend to study the behavior of the line-search proposed by us when applied over $\sigma$-approximate steepest descent direction, instead of the exact one, and when $\sigma$-approximate steepest descent direction replaces the steepest descent direction in the computation of conjugate gradient directions.

Secondly, it is worth studying the behavior of other classes of algorithms, for instance, Newton and projected gradient, when the line-search is performed according to our proposal.

Finally, it seems to be possible to apply our framework to vectorial variational inequality problems, since the line-search use only information of first order.

# Bibliography


[1] Ansary, M. A.; Panda, G. **A modified quasi-newton method for vector optimization problem**. *Optimization*, 64(11):2289–2306, 2015.

[2] Armijo, L. **Minimization of functions having Lipschitz continuous first partial derivatives**. *Pacific J. Math.*, 16:1–3, 1966.

[3] Beck, A.; Teboulle, M. **A conditional gradient method with linear rate of convergence for solving convex linear systems**. *Math. Methods Oper. Res.*, 59(2):235–247, 2004.

[4] Bello Cruz, J. Y.; Lucambio Pérez, L. R.; Melo, J. G. **Convergence of the projected gradient method for quasiconvex multiobjective optimization**. *Nonlinear Anal.*, 74(16):5268–5273, 2011.

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

[6] Bertsekas, D. P. **Convex analysis and optimization**. Athena Scientific, Belmont, MA, 2003. With Angelia Nedić and Asuman E. Ozdaglar.

[7] Burachik, R.; Graña Drummond, L. M.; Iusem, A. N.; Svaiter, B. F. **Full convergence of the steepest descent method with inexact line searches**. *Optimization*, 32(2):137–146, 1995.

[8] Custódio, A. L.; Madeira, J. F. A.; Vaz, A. I. F.; Vicente, L. N. **Direct multisearch for multiobjective optimization**. *SIAM J. Optim.*, 21(3):1109–1140, 2011.

[9] Dai, Y. H.; Yuan, Y. **A nonlinear conjugate gradient method with a strong global convergence property**. *SIAM J. Optim.*, 10(1):177–182, 1999.

[10] Das, I.; Dennis, J. E. **Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems**. *SIAM J. Optim.*, 8(3):631–657, 1998.

[11] Deb, K.; Thiele, L.; Laumanns, M.; Zitzler, E. **Scalable Multi-Objective Optimization Test Problems**. In: *Congress on Evolutionary Computation (CEC 2002)*, p. 825–830. IEEE Press, 2002.

[12] Deb, K. **Multi-objective genetic algorithms: Problem difficulties and construction of test problems**. *Evolutionary computation*, 7(3):205–230, 1999.

[13] Dolan, E. D.; Moré, J. J. **Benchmarking optimization software with performance profiles**. *Mathematical programming*, 91(2):201–213, 2002.

[14] Dong, Y. **New step lengths in conjugate gradient methods**. *Comput. Math. Appl.*, 60(3):563–571, 2010.

[15] Dong, Y. **A practical PR+ conjugate gradient method only using gradient**. *Appl. Math. Comput.*, 219(4):2041–2052, 2012.

[16] Ermol'ev, Y. M. **On the method of generalized stochastic gradients and quasi-fejér sequences**. *Cybernetics*, 5(2):208–220, 1969.

[17] Fazzio, N. S.; Schuverdt, M. L. **Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems**. *Optimization Letters*, 13(6):1365–1379, 2019.

[18] Fletcher, R.; Reeves, C. M. **Function minimization by conjugate gradients**. *Comput. J.*, 7:149–154, 1964.

[19] Fletcher, R. **Practical Methods of Optimization: Volume 1 Unconstrained Optimization**. J. WILEY & SONS, 1980.

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

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

[22] Fliege, J.; Vaz, A. I. F. **A method for constrained multiobjective optimization based on SQP techniques**. *SIAM J. Optim.*, 26(4):2091–2119, 2016.

[23] Fukuda, E.; Graña Drummond, L. M. **A survey on multiobjective descent methods**. *Pesquisa Operacional*, 34:585–620, 09 2014.

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

[25] Gilbert, J. C.; Nocedal, J. **Global convergence properties of conjugate gradient methods for optimization**. *SIAM J. Optim.*, 2(1):21–42, 1992.

[26] Gonçalves, M. L. N.; Prudente, L. F. **On the extension of the Hager-Zhang conjugate gradient method for vector optimization**. *Comput. Optim. Appl.*, 76(3):889–916, 2020.

[27] Graña Drummond, L. M.; Iusem, A. N. **A projected gradient method for vector optimization problems**. *Comput. Optim. Appl.*, 28(1):5–29, 2004.

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

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

[30] Hestenes, M. R.; Stiefel, E. **Methods of conjugate gradients for solving linear systems**. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.

[31] Hillermeier, C. **Generalized homotopy approach to multiobjective optimization**. *J. Optim. Theory Appl.*, 110(3):557–583, 2001.

[32] Huband, S.; Hingston, P.; Barone, L.; While, L. **A review of multiobjective test problems and a scalable test problem toolkit**. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, 2006.

[33] Iusem, A. N.; Svaiter, B. F.; Teboulle, M. **Entropy-like proximal methods in convex programming**. *Math. Oper. Res.*, 19(4):790–814, 1994.

[34] Jahn, J. **Vector Optimization**. Springer-Verlag, Berlin, 2011.

[35] Jin, Y.; Olhofer, M.; Sendhoff, B. **Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how?** In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCOâ€™2001)*, p. 1042–1049, 2001.

[36] Kim, I. Y.; De Weck, O. L. **Adaptive weighted-sum method for bi-objective optimization: Pareto front generation**. *Structural and multidisciplinary optimization*, 29(2):149–158, 2005.

[37] Laumanns, M.; Thiele, L.; Deb, K.; Zitzler, E. **Combining convergence and diversity in evolutionary multiobjective optimization**. *Evolutionary computation*, 10(3):263–282, 2002.

[38] Lovison, A. **Singular continuation: generating piecewise linear approximations to Pareto sets via global analysis**. *SIAM J. Optim.*, 21(2):463–490, 2011.

[39] Luc, D. T. **Theory of vector optimization**, volume 319 de **Lecture Notes in Economics and Mathematical Systems**. Springer-Verlag, Berlin, 1989.

[40] Lucambio Pérez, L. R.; Prudente, L. F. **Nonlinear conjugate gradient methods for vector optimization**. *SIAM J. Optim.*, 28(3):2690–2720, 2018.

[41] Miglierina, E.; Molho, E.; Recchioni, M. C. **Box-constrained multi-objective optimization: a gradient-like method without a priori scalarization**. *European Journal of Operational Research*, 188(3):662–682, 2008.

[42] Mita, K.; Fukuda, E. H.; Yamashita, N. **Nonmonotone line searches for unconstrained multiobjective optimization problems**. *J. Global Optim.*, 75(1):63–90, 2019.

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

[44] Nesterov, Y. **Introductory lectures on convex programming volume i: Basic course**. *Lecture notes*, 3(4):5, 1998.

[45] Nocedal, J.; Wright, S. J. **Numerical optimization**. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[46] Polak, E.; Ribière, G. **Note sur la convergence de méthodes de directions conjuguées**. *Rev. Française Informat. Recherche Opérationnelle*, 3(16):35–43, 1969.

[47] Polyak, B. **The conjugate gradient method in extremal problems**. *USSR Computational Mathematics and Mathematical Physics*, 9(4):94–112, 1969.

[48] Preuss, M.; Naujoks, B.; Rudolph, G. **Pareto set and emoa behavior for simple multimodal multiobjective functions**. In: *Parallel Problem Solving from Nature-PPSN IX*, p. 513–522. Springer, 2006.

[49] Qu, S.; Ji, Y.; Jiang, J.; Zhang, Q. **Nonmonotone gradient methods for vector optimization with a portfolio optimization application**. *European J. Oper. Res.*, 263(2):356–366, 2017.

[50] Schütze, O.; Lara, A.; Coello, C. C. **The directed search method for unconstrained multi-objective optimization problems**. *Proceedings of the EVOLVE–A Bridge Between Probability, Set Oriented Numerics, and Evolutionary Computation*, p. 1–4, 2011.

[51] Schütze, O.; Laumanns, M.; Coello, C. A. C.; Dellnitz, M.; Talbi, E.-G. **Convergence of stochastic search algorithms to finite size pareto set approximations**. *Journal of Global Optimization*, 41(4):559–577, 2008.

[52] Stadler, W.; Dauer., J. **Multicriteria optimization in engineering: A tutorial and survey.** *Progress in Astronautics and Aeronautics*, 150:209–209, 1993.

[53] Toint., P. L. **Test problems for partially separable optimization and results for the routine pspmin.** The University of Namur, Department of Mathematics, Belgium, Tech. Rep, 1983.

[54] Zhang, H.; Hager, W. W. **A nonmonotone line search technique and its application to unconstrained optimization**. *SIAM J. Optim.*, 14(4):1043–1056, 2004.

APPENDIX A

# Metric

The information in this appendix was taken from [8, 22].

## A.1 Purity metric

"Let $F_{p,s}$ be the approximation to the Pareto front computed by solver s for problem $p$. Let $F_p$ be the approximation to the Pareto front obtained by the union of all individual Pareto approximation, $\cup_{s \in S} F_{p,s}$, where all dominated points are removed. Since the true Pareto front is not known for all problems in our problems database, we consider $F_p$ in place of the true Pareto front. We define the purity metric as the number of points in $F_p$ divided by the number of points solver $s$ is able to compute that are not dominated by any other point computed, i.e.,

$$t_{p,s} = \frac{|F_p|}{|F_{p,s} \cap F_p|}$$

The purity metric measures the inverse of how many nondominated points a solver is able to compute from the set of all nondominated points computed. In our version of the metric, small values are better, as necessitated when using performance profiles. In case $|F_{p,s} \cap F_p| = 0$ we set $t_{p,s} := \infty$, meaning that solver s was unable to provide even a single nondominated point for problem $p$."

## A.2 Spread metrics

"While the purity metric measures how well a solver is able to compute nondominated points, the purity metric is unable to provide any information about how points are spread over the Pareto front. In order to understand whether a given solver is able to provide an approximation to the Pareto front whose points are "well distributed," we consider two additional metrics for our performance profiles. Let the approximated Pareto front computed by solver s for problem p be formed of $N$ points $x_1, \ldots, x_N$, and let these points be sorted by objective function $j$, i.e.,

$f_j(xi) \leq f_j(x_{i+1})$ $(i = 1, \ldots, N)$. Furthermore, let $x_0$ and $x_{N+1}$ be the extreme values for objective $j$; i.e., $x_0$ is the best known approximation to a global minimum of $f_j$, and $x_{N+1}$ is the best known approximation to a global maximum of $f_j$, computed over all Pareto front approximations obtained. Define $\delta_{i,j} = |f_j(x_{i+1}) - f_j(x_i)|$, and let $\bar{\delta}_j$ $(j = 1, \ldots, m)$ be the average of the distances $\delta_{i,j}$. The $\Gamma > 0$ and $\Delta > 0$ metrics are then defined as

$$\Gamma_{p,s} = \max_{j \in 1,\ldots,m} \max_{i \in 0,\ldots,N} \delta_{i,j},$$

and

$$\Delta_{p,s} = \max_{j \in 1,\ldots,m} \left( \frac{\delta_{0,j+\delta_{N,j}+\sum_{i=1}^{N} |\delta_{i,j} - \bar{\delta}_j|}}{\delta_{0,j+\delta_{N,j}+(N-1)\bar{\delta}_j}} \right).$$

Including $x_0$ and $x_{N+1}$ in the above is important, as $f(x_1)$ and $f(x_N)$ may be close to each other but far away from the true Pareto front extremes. This inclusion ensures that the metric $\Gamma$ is always well defined, while $\Delta$ is not defined in the case $N = 1$, $x_0 = x_1 = x_{N+1}$. While the $\Gamma$ metric measures the largest gap in the Pareto front, the $\Delta$ metric measures the scaled deviation from the average gap in the Pareto front."