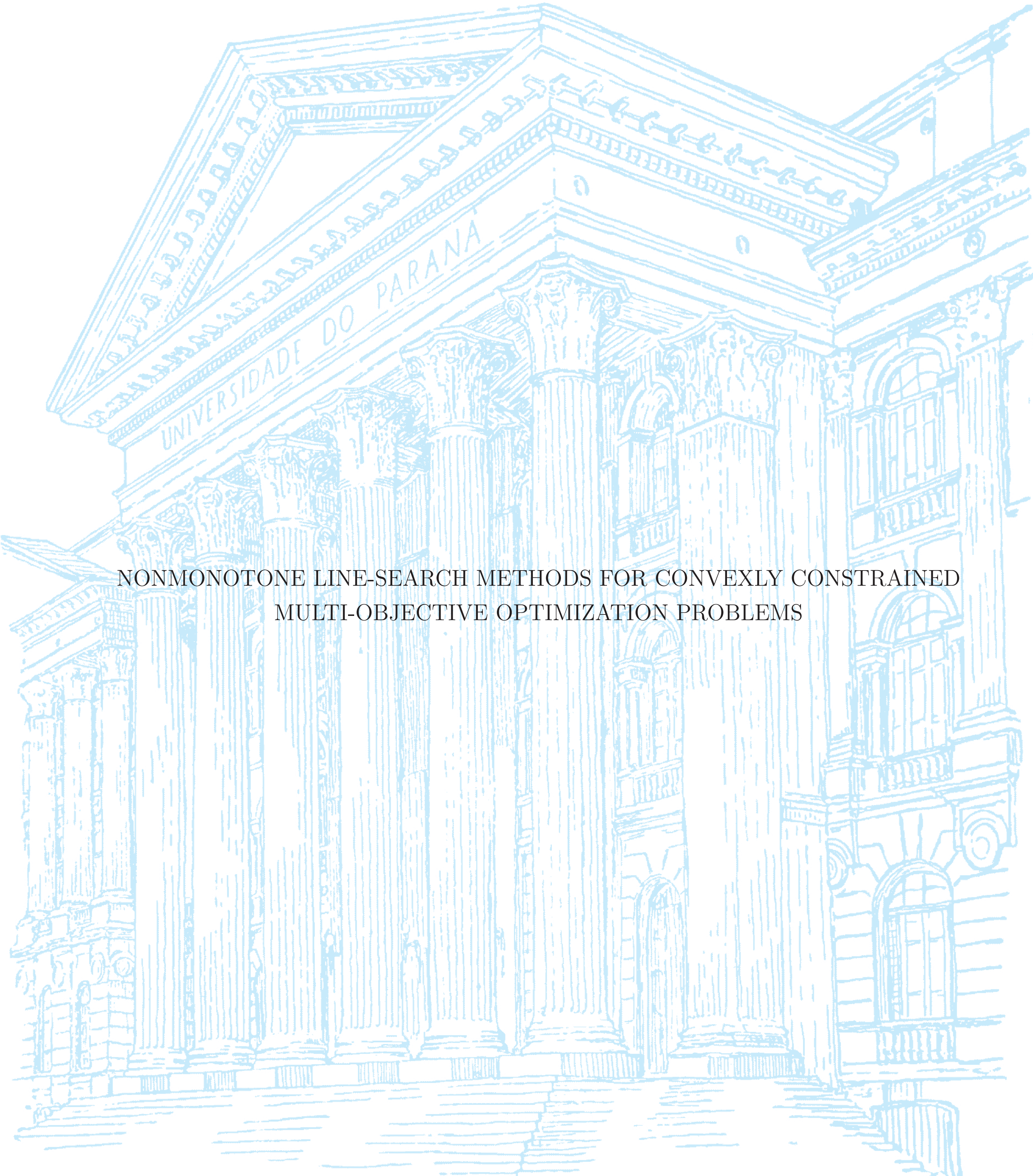
UNIVERSIDADE FEDERAL DO PARANÁ

MARIA EDUARDA PINHEIRO



NONMONOTONE LINE-SEARCH METHODS FOR CONVEXLY CONSTRAINED MULTI-OBJECTIVE OPTIMIZATION PROBLEMS

CURITIBA

2022

UNIVERSIDADE FEDERAL DO PARANÁ
MARIA EDUARDA PINHEIRO

NONMONOTONE LINE-SEARCH METHODS FOR CONVEXLY CONSTRAINED MULTI-OBJECTIVE OPTIMIZATION PROBLEMS

Dissertação apresentada ao curso de Pós-Graduação em Matemática, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial à obtenção do Título de Mestre em Matemática.

Orientador: Prof. Dr. Geovani Nunes Grapiglia.

CURITIBA
2022

Catalogação na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

P654n Pinheiro, Maria Eduarda

Non-monotone line-search methods for convexly constrained multi-objective optimization problems [recurso eletrônico] / Maria Eduarda Pinheiro - Curitiba, 2022.

Dissertação (mestrado) - Programa de Pós-Graduação em Matemática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientador: Prof. Dr. Geovani Nunes Grapiglia

1. Matemática. 2. Otimização matemática. 3. Lógica simbólica. I. Universidade Federal do Paraná. II. Grapiglia, Geovani Nunes. III. Título.

CDD 519.6

Bibliotecário: Nilson Carlos Vieira Junior CRB9/1797

ATA Nº104

ATA DE SESSÃO PÚBLICA DE DEFESA DE MESTRADO PARA A OBTENÇÃO DO GRAU DE MESTRA EM MATEMÁTICA

No dia quatorze de fevereiro de dois mil e vinte e dois às 14:00 horas, na sala https://meet.jit.si/DefesaMariaEduarda, remoto, foram instaladas as atividades pertinentes ao rito de defesa de dissertação da mestranda **MARIA EDUARDA PINHEIRO**, intitulada**: Nonmonotone Line-Search Methods for Convexly Constrained Multiobjective Optimization Problems**, sob orientação do Prof. Dr. GEOVANI NUNES GRAPIGLIA. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação MATEMÁTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: GEOVANI NUNES GRAPIGLIA (UNIVERSIDADE FEDERAL DO PARANÁ), LEANDRO DA FONSECA PRUDENTE (UNIVERSIDADE FEDERAL DE GOIÁS), ELIZABETH WEGNER KARAS (UNIVERSIDADE FEDERAL DO PARANÁ). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de mestra está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, GEOVANI NUNES GRAPIGLIA, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 14 de Fevereiro de 2022.

Assinatura Eletrônica
17/02/2022 08:08:40.0
GEOVANI NUNES GRAPIGLIA
Presidente da Banca Examinadora

Assinatura Eletrônica
17/02/2022 08:28:49.0
LEANDRO DA FONSECA PRUDENTE
Avaliador Externo (UNIVERSIDADE FEDERAL DE GOIÁS)

Assinatura Eletrônica
17/02/2022 07:57:12.0
ELIZABETH WEGNER KARAS
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MATEMÁTICA -
40001016041P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **MARIA EDUARDA PINHEIRO** intitulada: **Nonmonotone Line-Search Methods for Convexly Constrained Multiobjective Optimization Problems**, sob orientação do Prof. Dr. GEOVANI NUNES GRAPIGLIA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 14 de Fevereiro de 2022.

Assinatura Eletrônica
17/02/2022 08:08:40.0
GEOVANI NUNES GRAPIGLIA
Presidente da Banca Examinadora

Assinatura Eletrônica
17/02/2022 08:28:49.0
LEANDRO DA FONSECA PRUDENTE
Avaliador Externo (UNIVERSIDADE FEDERAL DE GOIÁS)

Assinatura Eletrônica
17/02/2022 07:57:12.0
ELIZABETH WEGNER KARAS
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

À minha mãe Marina e ao meu amor João.

AGRADECIMENTOS

Agradeço Àquele que nos fornece ferramentas para a construção de um mundo mais justo e igualitário.

À minha mãe Marina pelo afeto, amor e encorajamento para que eu sempre seguisse os meus sonhos. Ao meu primo Luiz Rafael por todas as conversas, conselhos e apoio.

Ao meu marido João Vitor que esteve ao meu lado durante todo o mestrado, compartilhando comigo todos os momentos, em especial os de comemoração. Obrigada por todo o carinho e compreensão.

Ao meu orientador, Professor Geovani Nunes Grapiglia, pelos ensinamentos, oportunidades e incentivos.

Aos professores Leandro da Fonseca Prudente e Elizabeh Wegner Karas por aceitarem o convite para participar da banca examinadora e pelas sugestões. À professora Elizabeh agradeço também por me apresentar o PPGM e pelo apoio no início do mestrado.

Aos professores que contribuíram para a minha caminhada, em especial ao professor Ademir Alves Ribeiro pela disponibilidade para conversar sobre ideias e interpretações.

A todos os colegas e amigos que tornaram a jornada mais prazerosa, sobretudo à Mariana da Rosa, que compartilhou esses dois anos comigo mesmo que de forma remota. Obrigada por todas as vídeo chamadas e fotos discutindo resultados e demonstrações, além de todas as conversas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

*"I like to learn.
That's an art and a science."*

Katherine Johnson

RESUMO

Nesta dissertação propomos uma classe de métodos de busca linear não-monótona para problemas de otimização multiobjetivo com restrições convexas. São obtidos limitantes de pior caso para o número de iterações que esses métodos precisam para gerar pontos Pareto-críticos aproximados. A generalidade da abordagem proposta permite o desenvolvimento de novos métodos para otimização multiobjetivo.

Palavras-chave: *Otimização multiobjetiva, Busca linear não-monótona, Complexidade de pior caso.*

ABSTRACT

In this work we propose a class of nonmonotone line-search methods for convexly constrained multiobjective optimization problems. Worst-case complexity bounds are obtained for the number of iterations that these methods need to generate an approximate Pareto critical point. The generality of our approach allows the development of new methods for multiobjective optimization

Keywords: *Multiobjective optimization, Nonmonotone line search, Worst-case complexity.*

Contents

Introduction

Motivation

In this work, we consider convex multiobjective optimization problems. This type of problem is characterized by a set of objective functions that have to be minimized simultaneously. It appears in several important applications, such as seismic design of buildings [27], planetary exploration [37, 35], maintenance of civil infrastructures [9], planning of cancer treatment [**12**, 40], drug design [21, 34], etc. In most cases, there is not a single point which minimizes all the objective functions at once. This fact motivates the notion of Pareto optimally or efficiency. Roughly speaking, a point is said to be Pareto optimal or efficient when from this point it is impossible to obtain an improvement in any of the objective functions without worsening the value of some other objective.

There are many strategies to solve multiobjective optimization problems [29, 30]. Among them, one of the most popular is the weighted sum method [28]. In recent years several methods that do not scalarize the problem by weighted sums have been developed (see, for instance, the surveys [3, 10]). Many of these methods were obtained by extending well-known scalar optimization algorithms. Notable representatives of this group are the steepest descent methods proposed in [8], the projected gradient methods proposed in [22, 12, 11, 25], the Newton's method proposed in [20], the proximal-point methods proposed in [13, 18] and the trust-region methods proposed in [14, 23].

So far, the theoretical study of the methods mentioned above has focused mainly in the asymptotic convergence analysis. However, in the context of scalar optimization, the worst-case complexity analysis of algorithms for smooth and possibly nonconvex problems has become a very active research area in the past few years (see, for instance, [2, 42, 41, 31, 36, 38, 24]). The goal of the worst-case complexity analysis is to estimate the maximal number of iterations (or function/gradient evaluations) required by the algorithm to produce an approximate first-order critical point of the problem.

In the context of multiobjective optimization, Grapglia, Yuan and Yuan [42] proved tha the trust-region algorithm proposed in [34] takes at most $\mathcal{O}(\epsilon^{-2})$ iterations to generate an ϵ-approximate Pareto critical point. More recently Fliege, Vaz and Vicente [19] obtained a bound of the same order for the gradient descent method proposed in [8] to unconstrained nonconvex multiobjective problems. In both contributions [19, 42] it is assumed that the gradients of the objectives are Lipschtiz continuous and only the unconstrained case is considered.

In this dissertation, we proposed a general class of nonmonotone line-search methods for convexly constrained multiobjective optimization problems. The methods are nonmonotone in the sense that they do not impose a decrease in the merit function at successive iterations. The proposed class generalizes the class of nonmonotone methods propose by Sachs and Sachs [33] for unconstrained single-objective optimization. The nonmonotonicity is controlled by a sequence of non-negative numbers $\{\nu_k\}$. We study the worst case complexity of the methods in the proposed class when each one of the m objective functions has Hölder continuous gradient with smoothness parameter $\theta_i \in (0, 1]$. In particular, when $\{\nu_k\}$ is summable we show that the corresponding nonmonotone methods take at most $\mathcal{O}\left(\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}\right)$ iterations to generate an ϵ-approximate Pareto critical point, where $\theta_{\min} = \min_{i=1,\ldots,m}\{\theta_i\}$. Numerical experiments with difference choices for $\{\nu_k\}$ are also presented. They suggest that nonmonotone methods compare favorably with the monotone method.

Contents

This dissertation is organized as follows. In Chapter 1, the problem is defined and results about multiobjective optimization are reviewed. Chapter 2 contains the new class of non-monotone methods and its worst-case complexity analysis. In Chapter 3, specific non-monotone methods are proposed and analyzed. Finally, in Chapter 4, numerical results are reported.

Notations

In that follows, $\|.\|$ denotes the euclidian norm and $I = \{1, \ldots, m\}$. Given $\Omega \subset \mathbb{R}^n$ and $x \in \Omega$, we define $\Omega - x = \{y \in \mathbb{R}^n y + x \in \Omega\}$. Besides that, given $a \in \mathbb{R}^m$, $\text{mean}(a) = \sum_{i=1}^m a_i/m$, and $\varphi(a) = \max_{i\in I}\{a_i\}$. $J_F(x)$ denotes the Jacobian of F at a point x and $L(x_0) = \{x \in \Omega : \varphi(F(x)) \le \varphi(F(x_0))\}$. We also consider the sets

$$\mathbb{R}^m_+ = \{z \in \mathbb{R}^m : z_i \ge 0,\ \forall i \in I\}$$

and

$$\mathbb{R}^m_{++} = \{z \in \mathbb{R}^m : z_i > 0,\ \forall i \in I\}.$$

Moreover, the relations $\succ$ and $\succ_\omega$ is given respectively by

$$y \succ x \iff y - x \in \mathbb{R}^m_+, \text{ and } y \succ_\omega x \iff y - x \in \mathbb{R}^m_{++}.$$

Chapter 1

Problem Definition and Auxiliary Results

In this work we consider methods for multi-objective problems on the form:

$$\begin{array}{ll} \min & F(x) = (f_1(x), \ldots, f_m(x))^T \\ \text{s.t} & x \in \Omega \end{array} \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continously differentiable function and $\Omega \subset \mathbb{R}^n$ is a closed and convex set.

Definition 1.1 (Guerraggio and Luc [16], page 619)**.** *Given a point $x^* \in \Omega$.*

(a) x^* is an efficient solution of (1.1) *when there is no $y \in \Omega$ such that $F(x^*) \succ F(y)$ and $F(x^*) \neq F(y)$;*

(b) x^* is said to be a weakly efficient solution of (1.1) *when there is no $y \in \Omega :\; F(x^*) \succ_\omega F(y)$;*

(c) x^* is a local (or local weakly) efficient solution of (1.1) *when there exist a neighborhood $N(x^*)$ of x^* for which there is no $y \in N(x^*) \cap \Omega : F(x^*) \succ F(y)$ and $F(x^*) \neq F(y)$ $(F(x^*) \succ_\omega F(y))$.*

Lemma 1.2. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a continous differentiable function and $\Omega \subset \mathbb{R}^n$ be a closed and convex set. If x^* is a local weakly solution of* (1.1) *then*

$$-\mathbb{R}^m_{++} \cap \{J_F(x^*)(x - x^*) : x \in \Omega\} = \emptyset. \tag{1.2}$$

Proof. See Theorem 5.1 (item(ii)-(a)) in Guerraggio and Luc [16]. □

Definition 1.3. *A point $x^* \in \Omega$ is said to be a Pareto critical point for* (1.1) *if it satisfies condition* (1.2).

Remark 1.4. *If* $m = 1$ *and* $\Omega = \mathbb{R}^n$, (1.1) *reduces to the unconstrained scalar optimization and the Pareto criticality condition* (1.2) *is equivalent to the usual critical condition* $\nabla f(x^*) = 0$.

Remark 1.5. *Let* $x \in \Omega$ *be a point which is not Pareto critical. Then there exists a direction* $d \in \Omega - x : J_F(x)d \in -\mathbb{R}^m_{++}$, *that is,* $J_F(x)d \prec_\omega 0$ *and* d *is a descent direction for* F *at* x.

For each $x \in \Omega$ consider the following subproblem

$$\min_{d \in \Omega - x} h_x(d) = \max_{i \in I}\{\nabla f_i(x)^T d\} + \frac{\|d\|^2}{2} \tag{1.3}$$

The next lemma establishes important properties of subproblem (1.3).

Lemma 1.6. *The following statments holds:*

(a) The subproblem (1.3) *has only one solution.*

(b) If x *is a Pareto critical point of* F *and* $d(x) = argmin_{d\in\Omega-x} h_x(d)$ *then* $d(x) = 0$ *and, consequently,* $h_x(d(x)) = 0$.

(c) If x *is not a Pareto critical point of* F*, therefore* $d(x) \neq 0$ *and* $h_x(d(x)) < 0$. *In particular,* $d(x)$ *is a descent direction for* F *at* x.

(d) The mapping $x \to d(x)$ *is continuous.*

Proof. See Proposition 3 in [5] and Lemma 1 [7]. □

Remark 1.7. *From statments (b) and (d), it follows that* $\|d(x)\|$ *is a suitable Pareto criticality measure for* x.

In what follows we will consider the following assumption:

A1 For $i = 1, \ldots, m$, $f_i \in C^{1,\theta_i}_{Li}(\Omega)$, that is, f_i is continously differentiable and its gradient ∇f_i is Hölder continuous with constant $L_i > 0$ and smoothness level $\theta_i \in (0, 1]$:

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq L_i \|x - y\|^{\theta_i},\ \forall x, y \in \Omega.$$

Lemma 1.8. *Suppose that* ***A1*** *holds and let* $i \in I$. *Then, for all* $x, y \in \Omega$

$$f_i(y) \leq f_i(x) + \nabla f_i(x)^T (y - x) + \frac{L_i}{\theta_i + 1} \|y - x\|^{1+\theta_i}.$$

Proof. See Lemma 1 in [39]. □

Lemma 1.9. *Suppose that **A1** holds and let*

$$x^+ = \bar{x} + \alpha d(\bar{x}) \tag{1.4}$$

for some $\bar{x} \in \Omega$ and $\alpha > 0$ where

$$d(\bar{x}) := \operatorname*{argmin}_{d \in \Omega - \bar{x}} h_{\bar{x}}(d) \tag{1.5}$$

with $h_{\bar{x}}(.)$ defined in (1.3)*. Given $\epsilon > 0$, define*

$$L(\epsilon) = \max_{i \in I} \left\{ \left(\frac{4L_i}{1+\theta_i} \right)^{\frac{1}{\theta_i}} \left(\frac{1}{\epsilon} \right)^{\frac{1-\theta_i}{\theta_i}} \right\}. \tag{1.6}$$

If

$$\|d(\bar{x})\| \geq \epsilon \text{ and } \alpha \leq \frac{1}{L(\epsilon)}, \tag{1.7}$$

then

$$\varphi(F(x^+)) \leq \varphi(F(\bar{x})) - \frac{\alpha}{4}\|d(\bar{x})\|^2. \tag{1.8}$$

Proof. By the first inequality in (1.7) and equality (1.4) we have, for all $i \in I$,

$$\epsilon \leq \|d(\bar{x})\| = \frac{1}{\alpha}\|x^+ - \bar{x}\| \implies \alpha\epsilon \leq \|x^+ - \bar{x}\| \implies \alpha^{1-\theta_i}\epsilon^{1-\theta_i} \leq \|x^+ - \bar{x}\|^{1-\theta_i}. \tag{1.9}$$

First we will show that

$$\frac{L_i}{1+\theta_i}\|x^+ - \bar{x}\|^{1+\theta_i} \leq \frac{\|x^+ - \bar{x}\|^2}{4\alpha}, \ \forall\, i \in I. \tag{1.10}$$

Indeed, by (1.9) we have

$$\begin{aligned} \frac{1}{4\alpha}\|x^+ - \bar{x}\|^2 &= \frac{1}{4\alpha}\|x^+ - \bar{x}\|^{1+\theta_i}\|x^+ - \bar{x}\|^{1-\theta_i} \\ &\geq \frac{1}{4\alpha}\|x^+ - \bar{x}\|^{1+\theta_i}\alpha^{1-\theta_i}\epsilon^{1-\theta_i} \\ &= \left(\frac{1}{4}\right)\left(\frac{1}{\alpha^{\theta_i}}\right)\epsilon^{1-\theta_i}\|x^+ - \bar{x}\|^{1+\theta_i}, \ \forall\, i \in I. \end{aligned} \tag{1.11}$$

On the other hand, by the second inequality in (1.7) and the definition (1.6) we get

$$
\begin{aligned}
\left(\frac{1}{4}\right)\left(\frac{1}{\alpha^{\theta_i}}\right)\epsilon^{1-\theta_i} &\geq \left(\frac{1}{4}\right) L(\epsilon)^{\theta_i}\epsilon^{1-\theta_i} \\
&\geq \left(\frac{1}{4}\right)\left[\frac{4L_i}{1+\theta_i}\left(\frac{1}{\epsilon}\right)^{1-\theta_i}\right]\epsilon^{1-\theta_i} \\
&= \frac{L_i}{1+\theta_i}, \ \forall\, i \in I. \qquad (1.12)
\end{aligned}
$$

Then, combining (1.11) and (1.12) we see that (1.10) is true. Now, it follows from **A1**, Lemma 1.8, (1.10) and (1.4) that

$$
\begin{aligned}
f_i(x^+) &\leq f_i(\bar{x}) + \nabla f_i(\bar{x})^T(x^+ - \bar{x}) + \frac{L_i}{1+\theta_i}\|x^+ - \bar{x}\|^{1+\theta_i} \\
&\leq f_i(\bar{x}) + \nabla f_i(\bar{x})^T(x^+ - \bar{x}) + \frac{1}{4\alpha}\|x^+ - \bar{x}\|^2 \\
&= f_i(\bar{x}) + \nabla f_i(\bar{x})^T(x^+ - \bar{x}) + \frac{\alpha^2}{4\alpha}\|d(\bar{x})\|^2 \\
&= f_i(\bar{x}) + \nabla f_i(\bar{x})^T(x^+ - \bar{x}) + \frac{\alpha}{4}\|d(\bar{x})\|^2. \qquad (1.13)
\end{aligned}
$$

In view of (1.5) and (1.3) we have

$$
\begin{aligned}
h_{\bar{x}}(d(\bar{x})) \leq h_{\bar{x}}(0) &\implies \max_{i\in I}\{\nabla f_i(\bar{x})^T d(\bar{x})\} + \frac{\|d(\bar{x})\|^2}{2} \leq 0 \\
&\implies \nabla f_i(\bar{x})^T d(\bar{x}) \leq -\frac{\|d(\bar{x})\|^2}{2}, \ \forall\, i \in I. \qquad (1.14)
\end{aligned}
$$

Combining (1.13) and (1.14), it follows that

$$
\begin{aligned}
& f_i(x^+) \leq f_i(\bar{x}) - \frac{\alpha}{2}\|d(\bar{x})\|^2 + \frac{\alpha}{4}\|d(\bar{x})\|^2 \\
\implies & f_i(x^+) \leq f_i(\bar{x}) - \frac{\alpha}{4}\|d(\bar{x})\|^2. \qquad (1.15)
\end{aligned}
$$

Since (1.15) holds for all $i \in \{1, \dots, m\}$ we have

$$
\varphi(F(x^+)) \leq \varphi(F(\bar{x})) - \frac{\alpha}{4}\|d(\bar{x})\|^2.
$$

□

Chapter 2

A Class of Nonmonotone Line-Search Methods

Let us consider the following general algorithm

Algorithm 1: General Nonmonotone Line-Search Method

Given : $x_0 \in \Omega$, $\epsilon, \beta \in (0,1)$, $\alpha_0 \in (0,1]$, set $k := 0$.

Step 1 Compute $d(x_k) = \operatorname{argmin}_{d\in\Omega - x_k} h_{x_k}(d)$ where

$$h_{x_k}(d) = \max_{i\in I}\{\nabla f_i(x_k)^T d\} + \frac{\|d\|^2}{2}.$$

If $\|d(x_k)\| \leq \epsilon$, then STOP.

Step 2.1 Find the smallest integer $\ell \geq 0$ such that $\alpha_k\beta^\ell \leq \alpha_0$.

Step 2.2 Choose $\nu_k \geq 0$. If

$$\varphi(F(x_k + \alpha_k\beta^\ell d(x_k))) \leq \varphi(F(x_k)) - \frac{\alpha_k\beta^\ell}{4}\|d(x_k)\|^2 + \nu_k \tag{2.1}$$

set $\ell_k = \ell$ and go to step 3. Otherwise, set $\ell := \ell + 1$ and go back to Step 2.2.

Step 3 Set $x_{k+1} = x_k + \alpha_k\beta^{\ell_k} d(x_k)$, $\alpha_{k+1} = \alpha_k\beta^{\ell_k - 1}$, $k := k+1$ and go back to Step 1.

Remark 2.1. *The applicability of Algorithm 1 strongly depends on our ability to solve subproblems of the form*

$$\min_{d\in\Omega - x}\max_{i\in I}\{\nabla f_i(x)^T d\} + \frac{\|d\|^2}{2}. \tag{2.2}$$

In many pratical applications, Ω is a polyedron, $\Omega = \{x \in \mathbb{R}^n : a_j^T x \leq b_j, j = 1, \ldots, p\}$.

In this case, the solution $d(x) = y^* - x$ of (2.2) can be obtained by solving

$$\begin{array}{lll} \min_{y\in\mathbb{R}^n,\nu\in\mathbb{R}} & \nu + \frac{\|y-x\|^2}{2} & \\ \text{s.t} & \nabla f_i(x)^T(y-x) \leq \nu,\ i = 1,\dots,m & \quad (2.3)\\ & a_j^T y \leq b_j, j = 1,\dots,p. & \end{array}$$

Remark 2.2. *When $\nu_k > 0$ we may have $\varphi(F(x_{k+1})) \geq \varphi(F(x_k))$ resulting in a non-monotone sequence $\{\varphi(F(x_k))\}_{k\geq 0}$.*

Remark 2.3. *When $\Omega = \mathbb{R}^n$ and $m = 1$, Algorithm 1 reduces to a version of Algorithm 1 in [15], which is a general line-search method for single-objective optimization based on a nonmonotone Armijo line-search.*

In view of Lemma 1.9, Step 2 of Algorithm 1 is well defined. The next lemma gives a lower bound for $\{\alpha_k\}$.

Lemma 2.4. *Suppose **A1** holds, then, for all $k \in \mathbb{N}$*

$$\alpha_k \geq \min\left\{\frac{1}{L(\epsilon)}, \alpha_0,\right\} = \bar{\alpha}(\epsilon). \tag{2.4}$$

Proof. For $k = 0$ the result is true. Now, assume by contradiction that

$$\alpha_{k+1} < \min\left\{\frac{1}{L(\epsilon)}, \alpha_0\right\} = \bar{\alpha}(\epsilon)$$

for some $k > 0$.

By the definition of α_{k+1} it follows that

$$\alpha_k\beta^{\ell_k-1} < \min\left\{\frac{1}{L(\epsilon)}, \alpha_0\right\} = \bar{\alpha}(\epsilon).$$

In this case, Lemma 1.9 implies that

$$\varphi(F(x_k + \alpha_k\beta^{\ell_k-1}d)) \leq \varphi(F(x_k)) - \frac{\alpha_k\beta^{\ell_k-1}}{4}\|d(x_k)\|^2 + \nu_k$$

which contradicts the minimality of ℓ_k. Therefore, we must have

$$\alpha_{k+1} \geq \min\left\{\frac{1}{L(\epsilon)}, \alpha_0\right\} = \bar{\alpha}(\epsilon),\ \forall k \geq 0.$$

□

Lemma 2.5. *Suppose that **A1** holds and let N_k be the total number of function evaluations up to the $k-th$ iteration of Algorithm 1. Then*

$$N_k \leq 3 + 2k + \frac{1}{\log(\beta)}(\log(\bar{\alpha}(\epsilon)) - \log(\alpha_0)).$$

Proof. Since $\alpha_{k+1} = \beta^{\ell_k - 1}\alpha_k,\ \log(\alpha_{k+1}) = (\ell_k - 1)\log(\beta) + \log(\alpha_k)$.

This means that $\ell_k + 1 = 2 + \dfrac{\log(\alpha_{k+1}) - \log(\alpha_k)}{\log(\beta)}$.

So, since $\beta \in (0,1),\ log(\beta) < 0$ and

$$\begin{aligned} N_k &\leq 1 + \sum_{\ell=0}^{k} \ell_i + 1 = 1 + \sum_{\ell=0}^{k}\left(2 + \frac{\log(\alpha_{\ell+1}) - \log(\alpha_\ell)}{\log(\beta)}\right) \\ &= 1 + 2(k+1) + \sum_{\ell=0}^{k} \frac{\log(\alpha_{\ell+1}) - \log(\alpha_\ell)}{\log(\beta)} \\ &= 3 + 2k + \frac{\log(\alpha_{k+1}) - \log(\alpha_0)}{\log(\beta)} \leq 2(k+1) + \frac{\log(\bar{\alpha}(\epsilon)) - \log(\alpha_0)}{\log(\beta)}. \end{aligned}$$

□

From now on, let us consider the following assumptions:

A2 There exist $\varphi_{\ell ow} : \varphi(F(x)) \geq \varphi_{\ell ow},\ \forall x \in \Omega$.

A3 $\lim_{T\to\infty} \dfrac{1}{T}\displaystyle\sum_{k=0}^{T-1} \nu_k = 0$.

Under assumption **A3**, given $\epsilon > 0$ there exists $T_0(\epsilon)$ such that

$$T \geq T_0(\epsilon) \implies \frac{1}{T}\sum_{k=0}^{T-1} \nu_k \leq \frac{\bar{\alpha}(\epsilon)\beta}{8}\epsilon^2 \tag{2.5}$$

where $\bar{\alpha}(\epsilon)$ is defined in (2.4). The next theorem establishes that Algorithm 1 performs at most $\mathcal{O}(\max\{T_0(\epsilon), \bar{\alpha}(\epsilon)^{-1}\epsilon^{-2}\})$ iterations to generate x_k such that $\|d(x_k)\| \leq \epsilon$.

Theorem 2.6. *Suppose that **A1-A3** hold and let $\{x_k\}_{k=0}^{T-1}$ be generated by Algorithm 1 with*

$$\|d(x_k)\| > \epsilon \ \ for\ k = 0, \ldots, T-1. \tag{2.6}$$

Then,

$$T \leq \max\{T_0(\epsilon), 8(\varphi(F(x_0)) - \varphi_{\ell ow})\beta^{-1}\bar{\alpha}(\epsilon)^{-1}\epsilon^{-2}\}. \tag{2.7}$$

Proof. By (2.1) and Lemma 2.4 we have

$$\begin{aligned} \nu_k + \varphi(F(x_k)) - \varphi(F(x_{k+1})) &\geq \frac{\alpha_k \beta^{\ell_k}}{4}\|d(x_k)\|^2 = \frac{\alpha_{k+1}\beta}{4}\|d(x_k)\|^2 \\ &\geq \frac{\bar{\alpha}(\epsilon)\beta}{4}\|d(x_k)\|^2 \geq \frac{\bar{\alpha}(\epsilon)\beta}{4}\epsilon^2, \end{aligned}$$

for $k = 0, \ldots, T-1$. Summing up these inequalities and using **A2** it follows that

$$\begin{aligned} T\frac{\bar{\alpha}(\epsilon)\beta}{4}\epsilon^2 &= \sum_{k=0}^{T-1}\frac{\bar{\alpha}(\epsilon)\beta}{4}\epsilon^2 \\ &\leq \sum_{k=0}^{T-1}(\nu_k + \varphi(F(x_k)) - \varphi(F(x_{k+1}))) \\ &= \sum_{k=0}^{T-1}\nu_k + \sum_{k=0}^{T-1}\varphi(F(x_k)) - \varphi(F(x_{k+1})) \\ &= \sum_{k=0}^{T-1}\nu_k + \varphi(F(x_0)) - \varphi(F(x_T)) \\ &\leq \sum_{k=0}^{T-1}\nu_k + \varphi(F(x_0)) - \varphi_{\ell ow}. \end{aligned}$$

So,

$$\epsilon^2 \leq \left(\frac{1}{T}\sum_{k=0}^{T-1}\nu_k\right)\frac{4}{\beta\bar{\alpha}(\epsilon)} + \frac{4(\varphi(F(x_0)) - \varphi_{\ell ow})}{T\beta\bar{\alpha}(\epsilon)}. \tag{2.8}$$

Assume that (2.7) is not true, i.e.,

$$T > \max\{T_0(\epsilon), 8(\varphi(F(x_0)) - \varphi_{\ell ow})\beta^{-1}\bar{\alpha}(\epsilon)^{-1}\epsilon^{-2}\}.$$

Then, by (2.5) in (2.8) we would have the contradiction

$$\begin{aligned} \epsilon^2 &< \frac{\bar{\alpha}(\epsilon)\beta}{8}\epsilon^2\frac{4}{\beta\bar{\alpha}(\epsilon)} + \frac{\beta\bar{\alpha}(\epsilon)\epsilon^2}{8(\varphi(F(x_0)) - \varphi_{\ell ow})}\frac{4(\varphi(F(x_0)) - \varphi_{\ell ow})}{\beta\bar{\alpha}(\epsilon)} \\ &= \frac{\epsilon^2}{2} + \frac{\epsilon^2}{2} = \epsilon^2. \end{aligned}$$

Therefore, (2.7) must be true. □

Remark 2.7. *By* (2.4) *and* (1.6)*, we have*

$$\begin{aligned} \bar{\alpha}(\epsilon)^{-1}\epsilon^{-2} &= \frac{1}{\min\{\frac{1}{L(\epsilon)}, \alpha_0\}}\epsilon^{-2} \\ &= \max\{L(\epsilon), \tfrac{1}{\alpha_0}\}\epsilon^{-2} \\ &= \max\left\{\max_{i\in I}\left\{\left(\frac{4L_i}{1+\theta_i}\right)^{\frac{1}{\theta_i}}\left(\frac{1}{\epsilon}\right)^{\frac{1-\theta_i}{\theta_i}}\right\}, \frac{1}{\alpha_0}\right\}\epsilon^{-2}. \end{aligned}$$

Thus, if we consider $\epsilon \in (0,1)$ and denote $\theta_{\min} = \min_{i\in I}\{\theta_i\}$, it follows that

$$\begin{aligned}
\bar{\alpha}(\epsilon)^{-1}\epsilon^{-2} &= \max\left\{\max_{i\in I}\left\{\left(\frac{4L_i}{1+\theta_i}\right)^{\frac{1}{\theta_i}}\left(\frac{1}{\epsilon}\right)^{\frac{1-\theta_i}{\theta_i}}\right\}, \frac{1}{\alpha_0}\right\}\epsilon^{-2} \\
&\leq \max\left\{\max_{i\in I}\left\{\left(\frac{4L_i}{1+\theta_i}\right)^{\frac{1}{\theta_i}}\right\}\epsilon^{-\left(\frac{1-\theta_{\min}}{\theta_{\min}}+2\right)}\right\} \\
&= \max\left\{\max_{i\in I}\left\{\left(\frac{4L_i}{1+\theta_i}\right)^{\frac{1}{\theta_i}}, \frac{1}{\alpha_0}\right\}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}\right. \\
&= \kappa_\theta \epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}
\end{aligned} \tag{2.9}$$

where

$$\kappa_\theta = \max\left\{\max_{i\in I}\left\{\left(\frac{4L_i}{1+\theta_i}\right)^{\frac{1}{\theta_i}}\right\}, \frac{1}{\alpha_0}\right\}. \tag{2.10}$$

Consequently,

$$\frac{\beta\bar{\alpha}(\epsilon)\epsilon^2}{8} \geq \frac{\beta}{8\kappa_\theta}\epsilon^{\left(\frac{1}{\theta_{\min}}+1\right)} \tag{2.11}$$

In view of Theorem 2.6 and Remark 2.7, Algorithm 1 performs at most

$$\mathcal{O}\left(\max\left\{T_0(\epsilon), \epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}\right\}\right)$$

iterations.

The next two corollaries describe situations in which a $T_0(\epsilon)$ can be obtained.

Corollary 2.8. *Suppose that* ***A1-A2*** *hold and that $\sum_{k=1}^{\infty}\nu_k < \infty$. Let $\{x_k\}_{k=0}^{T-1}$ be generated by Algorithm 1 with*

$$\|d(x_k)\| > \epsilon, \quad \textit{for } k = 1, \ldots, T-1,$$

for some $\epsilon \in (0,1)$. Then,

$$T \leq \max\left\{8\left(\sum_{k=1}^{\infty}\nu_k\right)\beta^{-1}\kappa_\theta, 8(\varphi(F(x_0)) - \varphi_{\ell ow})\beta^{-1}\right\}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)} \tag{2.12}$$

where κ_θ is defined in (2.10).

Proof. Note that $0 \leq \frac{1}{T}\sum_{k=1}^{T-1}\nu_k \leq \sum_{k=1}^{\infty}\nu_k$ for all $T > 1$. Since $\sum_{k=1}^{\infty}\nu_k < \infty$, it follows that **A3** holds. Moreover, using (2.9), since $T > 8\left(\sum_{k=1}^{\infty}\nu_k\right)\beta^{-1}\kappa_\theta\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}$, we have

$$T \geq 8\left(\sum_{k=1}^{\infty}\nu_k\right)\beta^{-1}\bar{\alpha}(\epsilon)^{-1}\epsilon^{-2} \text{ which means } \frac{1}{T}\left(\sum_{k=1}^{\infty}\nu_k\right) \leq \frac{\beta\bar{\alpha}(\epsilon)}{8}\epsilon^2.$$

Therefore, we can consider $T_0(\epsilon) = 8\left(\sum_{k=1}^{\infty}\nu_k\right)\beta^{-1}\kappa_\theta\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}$. Then by Theorem 2.6 and Remark 2.7, (2.12) holds. □

Corollary 2.9. *Suppose that **A1-A2** hold and that* $\lim_{k\to\infty}\nu_k = 0$*. Let* $c > 0$ *such that* $\nu_k \leq C$ *for all* $k \geq 0$*. Moreover, given* $\delta > 0$*, let* $k_0(\delta)$ *be a positive integer such that*

$$k \geq k_0(\delta) \implies \nu_k \leq \delta.$$

If $\{x_k\}_{k=0}^{T-1}$ *is generated by Algorithm 1 with*

$$\|d(x_k)\| > \epsilon \text{ for } k = 1, \ldots, T-1$$

Then,

$$T \leq \max\left\{\frac{16k_0(\frac{\delta}{2})C\kappa_\theta}{\beta}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}, 1 + k_0\left(\frac{\delta}{2}\right), 8(\varphi(F(x_0)) - \varphi_{low})\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}\right\} \quad (2.13)$$

for $\delta = \dfrac{\beta}{8\kappa_\theta}\epsilon^{\left(\frac{1}{\theta_{\min}}+1\right)}$.

Proof. In the proof of Corollary 2 in [15] it was shown that if

$$\Delta \geq \max\left\{\frac{2k_0(\frac{\delta}{2})C}{\delta}, 1 + k_0\left(\frac{\delta}{2}\right)\right\},$$

then $\dfrac{1}{\Delta}\sum_{k=1}^{\Delta-1}\nu_k \leq \delta$. Therefore $\{\nu_k\}$ satisfies **A3** and by (2.11) we can consider

$$T_0(\epsilon) = \max\left\{\frac{2k_0(\frac{\delta}{2})C}{\delta}, 1 + k_0\left(\frac{\delta}{2}\right)\right\},$$

with $\delta = \dfrac{\beta}{8\kappa_\theta}\epsilon^{\left(\frac{1}{\theta_{\min}}+1\right)}$. Then, by Theorem 2.6 and (2.9), (2.13) holds.

□

Chapter 3

Specific Nonmonotone Methods

The generality of our results allows several choices for the sequence $\{\nu_k\}$ in Algorithm 1. Each choice gives a different method for multi-objective optimization problems. In the context of single-objective unconstrained optimization, one of the most efficient nonmonotone line-search techniques is the one proposed by Zhang and Hager [43]. The next algorithm is a particular version of Algorithm 1 inspired by Zhang and Hager nonmonotone line-search.

Algorithm 2: Nonmonotone Line-Search Method inspired by Zhang and Hager

---

Given : $x_0 \in \Omega$, $\epsilon, \beta \in (0,1)$, $\alpha_0 \in (0,1]$ and a sequence $\{\eta_k\} \subset (0, \eta_{\max}]$ with $\eta_{\max} \in (0,1)$, set $Q_0 = 1$, $C_0 = \varphi(F(x_0))$ and $k := 0$.

Step 1 Compute $d(x_k) = \text{argmin}_{d \in \Omega - x_k} h_{x_k}(d)$ where

$$h_{x_k}(d) = \max_{i \in I}\{\nabla f_i(x_k)^T d\} + \frac{\|d\|^2}{2}.$$

If $\|d(x_k)\| \leq \epsilon$, then STOP.

Step 2.1 Find the smallest integer $\ell \geq 0$ such that $\alpha_k \beta^\ell \leq \alpha_0$.

Step 2.2 Set $\nu_k = C_k - \varphi(F(x_k))$.

If

$$\varphi(F(x_k + \alpha_k \beta^\ell d(x_k))) \leq \varphi(F(x_k)) - \frac{\alpha_k \beta^\ell}{4}\|d(x_k)\|^2 + \nu_k \tag{3.1}$$

go to step 3. Otherwise, set $\ell := \ell + 1$ and go back to Step 2.2.

Step 3 Set $x_{k+1} = x_k + \alpha_k \beta^{\ell_k} d(x_k),\ \alpha_{k+1} = \alpha_k \beta^{\ell_k - 1},\ Q_{k+1} = \eta_k Q_k + 1,$

$$C_{k+1} = \frac{\eta_k Q_k C_k + \varphi(F(x_{k+1}))}{Q_{k+1}}$$

$k := k + 1$ and go back to Step 1.

Theorem 3.1. *Suppose that* ***A1*** *and* ***A2*** *hold and let* $\{x_k\}_{k=0}^{T-1}$ *be generated by Algorithm 2 with*

$$\|d(x_k)\| > \epsilon, \quad \text{for } k = 0, \ldots, T-1,$$

for some $\epsilon \in (0,1)$. *Then,*

$$T \leq 8 \max\left\{\frac{\kappa_\theta}{(1-\eta_{\max})}, 1\right\} (\varphi(F(x_0)) - \varphi_{\ell ow})\beta^{-1}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}. \tag{3.2}$$

Proof. In view of Corollary 2.8, it is enough to show that for all T we have

$$\sum_{k=0}^{T-1} \nu_k \leq \frac{\eta_{\max}}{1-\eta_{\max}}(\varphi(F(x_0)) - \varphi_{\ell ow}).$$

Indeed, realize that

$$\begin{aligned}
Q_0 &= 1 \\
Q_1 &= \eta_0 + 1 \\
Q_2 &= \eta_1 Q_1 + 1 = \eta_1 + \eta_1\eta_0 + 1 \\
&\vdots \qquad\qquad \vdots \\
Q_k &= 1 + \sum_{j=1}^{k}\prod_{i=1}^{j} \eta_{k-i}
\end{aligned}$$

for $k > 0$. Since $\eta_k \leq \eta_{\max}$,

$$\begin{aligned}
Q_k &\leq 1 + \sum_{j=1}^{k}\prod_{i=1}^{j} \eta_{\max} \leq \sum_{j=0}^{k-1} \eta_{\max}^j \\
&\leq \sum_{j=0}^{\infty} \eta_{\max}^j = \frac{1}{1-\eta_{\max}}.
\end{aligned}$$

Besides that,

$$\begin{aligned}
C_k &= \frac{\eta_{k-1}Q_{k-1}C_{k-1} + \varphi(F(x_k))}{Q_k} = \frac{\eta_{k-1}Q_{k-1}C_{k-1}}{Q_k} + \frac{\varphi(F(x_k))}{Q_k} \\
&= \frac{(Q_k - 1)C_{k-1}}{Q_k} + \frac{\varphi(F(x_k))}{Q_k}.
\end{aligned}$$

Because $C_k = \nu_k + \varphi(F(x_k))$,

$$\begin{aligned}
\nu_k + \varphi(F(x_k)) &= \frac{(Q_k - 1)C_{k-1}}{Q_k} + \frac{\varphi(F(x_k))}{Q_k} \\
&= \frac{(Q_k - 1)(\nu_{k-1} + \varphi(F(x_{k-1})))}{Q_k} + \frac{\varphi(F(x_k))}{Q_k} \\
&= \nu_{k-1} + \varphi(F(x_{k-1})) - \frac{\nu_{k-1}}{Q_k} - \frac{\varphi(F(x_{k-1}))}{Q_k} + \frac{\varphi(F(x_k))}{Q_k}.
\end{aligned}$$

Then,

$$\frac{1}{Q_k}(v_{k-1}+\varphi(F(x_{k-1})-\varphi(F(x_k)))=\nu_{k-1}-\nu_k+\varphi(F(x_{k-1}))-\varphi(F(x_k))).$$

This equality with the facts that exist $\varphi_{\ell ow}$: $\varphi_{\ell ow}\leq\varphi(F(x))$ for all x, $\nu_0=0$ and $\nu_k\geq 0$ for all $k\geq 1$ implies

$$\begin{aligned}\sum_{k=1}^{T-1}\frac{1}{Q_k}(v_{k-1}+\varphi(F(x_{k-1})-\varphi(F(x_k))) =&\sum_{k=1}^{T-1}(\nu_{k-1}-\nu_k)+\sum_{k=1}^{T-1}\varphi(F(x_{k-1}))-\varphi(F(x_k)))\\ =&\nu_0-\nu_{T-1}+\varphi(F(x_0))-\varphi(F(x_{T-1})))\\ \leq&\varphi(F(x_0))-\varphi_{\ell ow}.\end{aligned}\tag{3.3}$$

Combining (3.3) with the fact that $Q_k\leq\dfrac{1}{1-\eta_{\max}}$ for all k, we get

$$\begin{aligned}\sum_{k=0}^{T-1}\nu_k &= \sum_{k=1}^{T-1}\nu_k=\sum_{k=1}^{T-1}C_k-\varphi(F(x_k))\\ &= \sum_{k=1}^{T-1}\frac{(Q_k-1)C_{k-1}}{Q_k}+\frac{\varphi(F(x_k))}{Q_k}-\varphi(F(x_k))\\ &= \sum_{k=1}^{T-1}\frac{(Q_k-1)C_{k-1}}{Q_k}+\frac{(1-Q_k)\varphi(F(x_k))}{Q_k}\\ &= \sum_{k=1}^{T-1}\frac{Q_k-1}{Q_k}(v_{k-1}+\varphi(F(x_{k-1}))-\varphi(F(x_k)))\\ &\leq \left(\frac{1}{1-\eta_{max}}-1\right)\sum_{k=1}^{T-1}\frac{1}{Q_k}(v_{k-1}+\varphi(F(x_{k-1})-\varphi(F(x_k)))\\ &= \frac{\eta_{\max}}{1-\eta_{max}}(\varphi(F(x_0))-\varphi_{\ell ow}).\end{aligned}$$

As T is arbitrary and $\eta_{\max}\leq 1$, we have

$$\sum_{k=0}^{\infty}\nu_k\leq\frac{1}{1-\eta_{max}}(\varphi(F(x_0))-\varphi_{\ell ow})<\infty$$

and together with (2.12) we obtain (3.2). □

Besides the generalization of known nonmonotone line-searches for single-objective optimization, our framework allows the development of new methods specially designed to multi-objective optimization. For example, given $M>0$, we can consider a version of

Algorithm 1 with $\nu_0 = 0$ and

$$\nu_k = \frac{\max\{0, \min\{\omega_k, M\}\}}{k^\rho} \tag{3.4}$$

for all $k \geq 1$, where $\rho > 0$ and

$$\omega_k = \text{mean }(F(x_{k-1})) - \text{mean }(F(x_k)). \tag{3.5}$$

The idea behind this choice of ν_k is to allow the acceptance of a larger stepsize at the kth iteration when the $(k-1)$th iteration was "successful" in the sense that

$$\text{mean }(F(x_k)) < \text{mean }(F(x_{k-1})).$$

Notice that ν_k in (3.4) satisfies

$$\nu_k \leq \frac{M}{k^\rho}, \ \forall k \geq 1. \tag{3.6}$$

The next theorem provides complexity bounds for all versions of Algorithm 1 when $\{\nu_k\}$ satisfies (3.6).

Theorem 3.2. *Suppose that* ***A1-A2*** *hold and let* $\{x_k\}_{k=0}^{T-1}$ *be generated by Algorithm 1 with*

$$\|d(x_k)\| > \epsilon, \quad \textit{for } k = 0, \ldots, T-1$$

for some $\epsilon \in (0,1)$*. If* $\nu_0 = 0$ *and* ν_k *satisfies* (3.6)*, for all* $k \geq 1$ *then*

$$T \leq \begin{cases} \max\left\{\dfrac{8M}{\rho-1}\beta^{-1}\kappa_\theta, 8(\varphi(F(x_0)) - \varphi_{\ell ow})\beta^{-1}\right\}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)} & \textit{if } \rho > 1 \\ \max\left\{\left(\dfrac{16M\kappa_\theta}{\beta}\right)^{\frac{\rho+1}{\rho}}, 1 + \left(\dfrac{16M\kappa_\theta}{\beta}\right)^{\frac{1}{\rho}}, 8(\varphi(F(x_0)) - \varphi_{\ell ow})\right\}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)\left(\frac{\rho+1}{\rho}\right)} & \textit{if } \rho \in (0,1]. \end{cases} \tag{3.7}$$

Proof. Suppose that $\rho > 1$. Then, in view of (3.6), for all T, we have

$$\sum_{k=0}^{T-1} \nu_k \leq \sum_{k=1}^{T-1} \frac{M}{k^\rho} = M \sum_{k=1}^{T-1} \frac{1}{k^\rho} \leq \frac{M}{\rho - 1}.$$

Thus, Corollary 2.8 provides the corresponding bound in (3.7). Now, suppose that $\rho \in (0,1]$. By (3.6), we have $\nu_k \to 0$. We also have $\nu_k \leq M$ for all $k \geq 0$ and, given δ, if $k \geq \left(\frac{M}{\delta}\right)^{\frac{1}{\rho}}$ then $\nu_k \leq \delta$ So denoting $C = M$ and $k_0(\delta) = \left(\dfrac{M}{\delta}\right)^{\frac{1}{\rho}}$ the corresponding bound in (3.7) follows from Corollary 2.9, with $\delta = \dfrac{\beta}{8\kappa_\theta}\epsilon^{\left(\frac{1}{\theta_{\min}}+1\right)}$. Indeed, just notice that

$$
\begin{aligned}
k_0\left(\frac{\delta}{2}\right) &= k_0\left(\frac{\beta}{16\kappa_\theta}\epsilon^{\left(\frac{1}{\theta_{\min}}+1\right)}\right) = \left(\frac{16M\kappa_\theta}{\beta}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}\right)^{\frac{1}{\rho}} \\
&= \left(\frac{16M\kappa_\theta}{\beta}\right)^{\frac{1}{\rho}}\epsilon^{-\left(\frac{1}{\theta_{\min}\rho}+\frac{1}{\rho}\right)}
\end{aligned}
$$

and so

$$
\begin{aligned}
k_0\left(\frac{\delta}{2}\right)\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)} &= \left(\frac{16M\kappa_\theta}{\beta}\right)^{\frac{1}{\rho}}\epsilon^{-\left(\frac{1}{\theta_{\min}\rho}+\frac{1}{\rho}+\frac{1}{\theta_{\min}}+1\right)} \\
&= \left(\frac{16M\kappa_\theta}{\beta}\right)^{\frac{1}{\rho}}\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)\left(\frac{\rho+1}{\rho}\right)}.
\end{aligned}
$$

□

Chapter 4

Numerical Tests

Implementation Details

To evaluate the performance of the methods discussed in Chapters 2 and 3, numerical experiments were a conducted on a set of 44 problems based-constrained multiobjective problems obtained with the functions and bounds describe in [26]. In these problems, most values for n and m vary between 2 and 10 and one of them has $n = 100$.

For each problem, 300 starting points were generated with a uniform mesh of the feasible set

$$\Omega = \{x \in \mathbb{R}^n : a \leq x \leq b\}.$$

The following `Julia` implementations were tested:

- M1: the monotone Algorithm 1 by setting $\nu_k = 0$ for all k.
- NM1: the nonmonotone algorithm obtained from Algorithm 1 by setting $\nu_0 = 0$ and

  $$\nu_k = \frac{\max\{0, \min\{\omega_k, M\}\}}{k^\rho}, \ \forall k \geq 1,$$

  where $\omega_k = \text{mean}(F(x_{k-1})) - \text{mean}\ (F(x_k))$, $M = 5$ and $\rho = 0.5$.
- NM2: the nonmonotone Algorithm 2 with $\eta_k = \dfrac{0.85}{k}, \forall k \geq 1$.
- FS: the monotone algorithm proposed by [6] with parameters $\sigma = 0.25$, $\eta_k = \dfrac{0.85}{k}$,and $\beta_k = 1$.

In implementations M1, NM1 and NM2 s $\alpha_0 = 1$ and $\beta = 0.5$ were considered because preliminary numerical tests suggest it is a good choice. We used the stopping rules:

$$\|d(x_k)\| \leq 10^{-4}, \tag{4.1}$$

$$k = 1000 \tag{4.2}$$

and

$$\ell = 20. \tag{4.3}$$

All the experiments were performed with `Julia` 1.5.2 on a PC with i7-87000U microprocessor 3.20GHz. We use Gurobi [17] to obtain the solution of the subproblem (2.3).

Performance Evaluation

The efficiency of the codes is compared using performance profile [4]. Specifically, let S be a set of solvers and $\mathcal{P}$ be a set of problems, and denote by $t_{p,s} > 0$ the performance of the solver $s \in S$ applied to problem $p \in \mathcal{P}$. Assume that lower values of $t_{p,s}$ mean the better performances. Consider the performance. Then the performance profile of solver s is the graph of the function $\gamma_s : [1, \infty) \to [0, 1]$ given by

$$\gamma_s(\tau) = \frac{1}{|P|}\left|\left\{p \in P : \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}} \leq \tau\right\}\right..$$

Note that, $\gamma_s(1)$ is the percentage of problems for which solver s wins over the rest of the solvers.

The natural choice for $t_{p,s}$ is the number of iterations that solver needs to generate a point x_k satisfying (4.1) for problem p. However, in the context of multiobjective optimization other performance measures are also relevant. For example, it is particularly interesting to quantify the ability of the algorithms to generate Pareto front, that is, the set of all efficient solutions. By applying solver s to problem p using a set of N starting points, one obtains a set of N "final points" in Ω. Removing from this set all the dominated points[1] let $PF_{p,s}$ be the corresponding approximation to the Pareto front obtained for problem p using solver s. Then, removing from $\cup_{s \in S} PF_{p,s}$ all the dominated points we can get an approximate Pareto front PF_p by combining the points obtained by all solvers. Using these sets we can defined the following performance measures:

[1]We say that a point x_1 is dominated by a point x_2 when $f_i(x_2) \leq f_i(x_1)$, $\forall i \in I$ and $f_j(x_2) < f_j(x_1)$ for some $j \in I$.

- Purity Metric [32]: This metric measures the number of nondominated points belonging to PF_p. More specifically, $\bar{t}_{p,s} := \dfrac{|PF_{p,s}| \cap |PF_p|}{|PF_p|}$ and $t_{p,s} := \dfrac{1}{\bar{t}_{p,s}}$. Note that when a solver s is not able to obtain a single nondominated point in PF_p, $\bar{t}_{p,s} = 0$, which means that $t_{p,s} = \infty$ and the solver s cannot solve problem p. When using the Purity Metric, as suggested in [1] we compare the algorithms in pairs.

- Spread Metrics: According to [1], this type of metric aims to measure the capacity of a particular solver s to get points well distributed along the Pareto Frontiers. Given a solver s and a problem p, consider $x_1, \ldots, x_N$ approximate Pareto points. For each $j = \{1, \ldots, m\}$ (objective function) sort by increasing order ($f_j(x_i) \leq f_j(x_{i+1})$) and find x_0 and $x_{N+1} \in PF_p$: x_0 and x_{N+1} are the points corresponding to the lowest and highest values of f_j, respectively.

  For each $j = 1, \ldots, m,\ i = 0, \ldots N$ we consider $\delta_{i,j} = f_j(x_{i+1}) - f_j(x_i)$ and for each j, $\bar{\delta}_j$ is the average of the distances $\delta_{i,j}$

$$\Gamma_{p,s} = \max_{j=1,\ldots,m} \{ \max_{i=0,\ldots,N} \delta_{i,j} \}$$

  and

$$\Delta_{p,s} = \max_{j=1,\ldots,m} \left(\frac{\delta_{0,j} + \delta_{N,j} + \sum_{i=1}^{N} |\delta_{i,j} - \bar{\delta}_j|}{\delta_{0,j} + \delta_{N,j} + (N-1)\bar{\delta}_j} \right),$$

  we name Γ and Δ metric, respectively, when $t_{p,s} = \Gamma_{p,s}$ and $t_{p,s} = \Delta_{p,s}$.

Comparing M1 and NM1

As can be seen in the Figure 4.1, comparing the number of iterations as performance profile, $M1$ and $NM1$ have similar robustness (83% and 86%) while $NM1$ the most efficient.

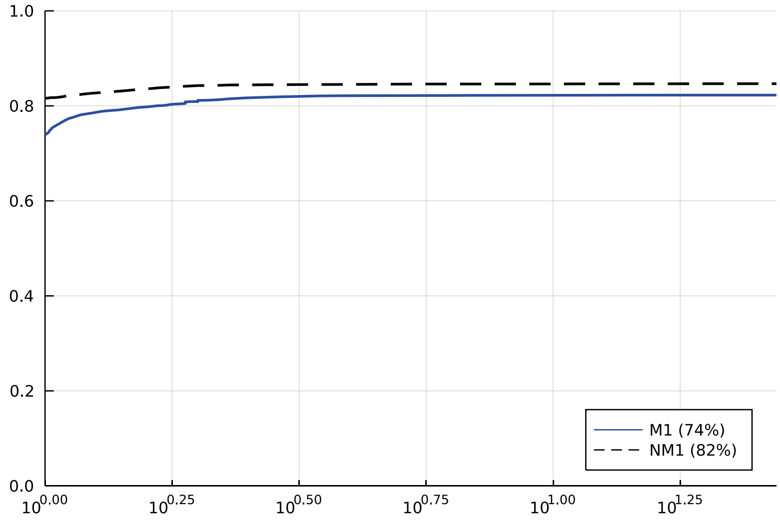


Figure 4.1: Performance profile using number of iterations comparing M1 and NM1

Regarding the spread metrics, Figures 4.2 and 4.3 show that, for the spreads performance profile Δ and Γ NM1 is more efficient than M1.

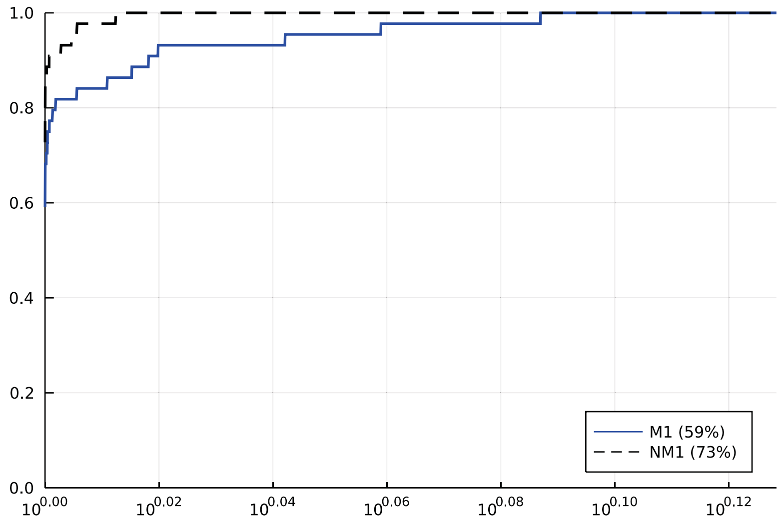


Figure 4.2: Spread Δ performance profiles comparing M1 and NM1.

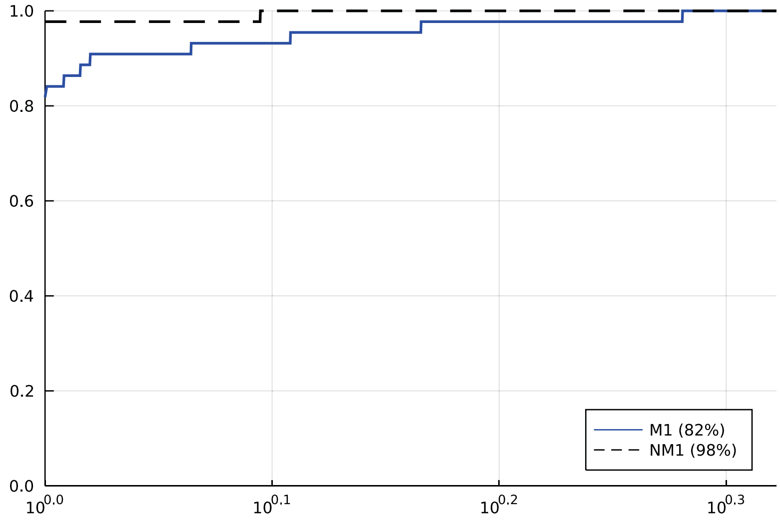


Figure 4.3: Spread Γ performance profiles comparing M1 and NM1.

Finally, with to respect to the Purity Metric, we can see in Figure 4.4 that NM1 was much more efficient than M1.

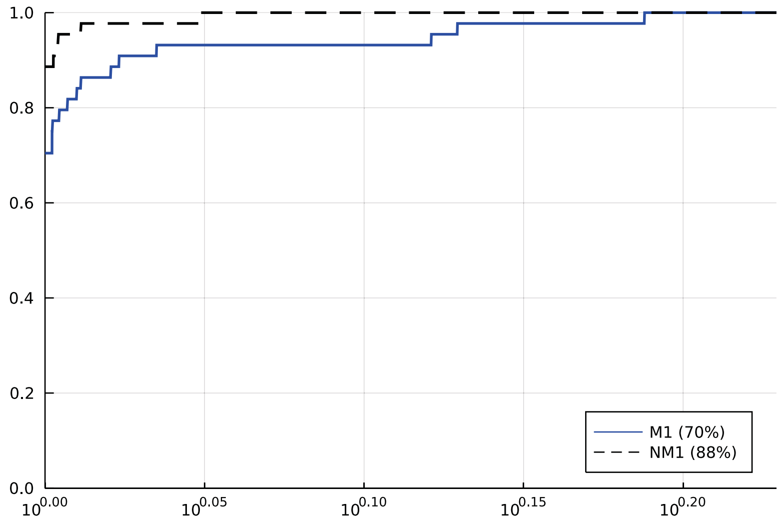


Figure 4.4: Purity performance profiles comparing M1 and NM1.

Comparing $NM1$ and $NM2$

As can be seen in the Figure 4.5, NM1 outperformed NM2 when we comparing number of iterations.

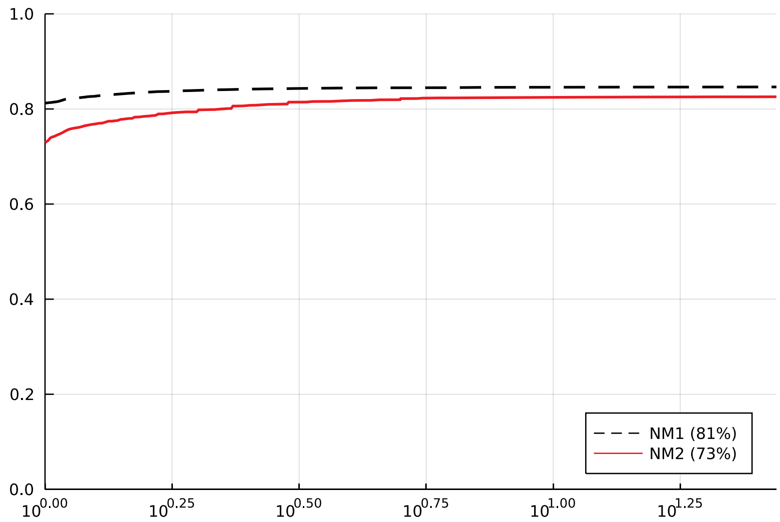


Figure 4.5: Performance profile using number of iterations comparing NM1 and NM2.

Regarding the spread metrics, Figures 4.6 and 4.7 show that for spread performance profile Δ, $NM2$ and $NM1$ are very similar. On the other hand, for spread performance profile Γ, NM1 was more efficient than $NM2$.

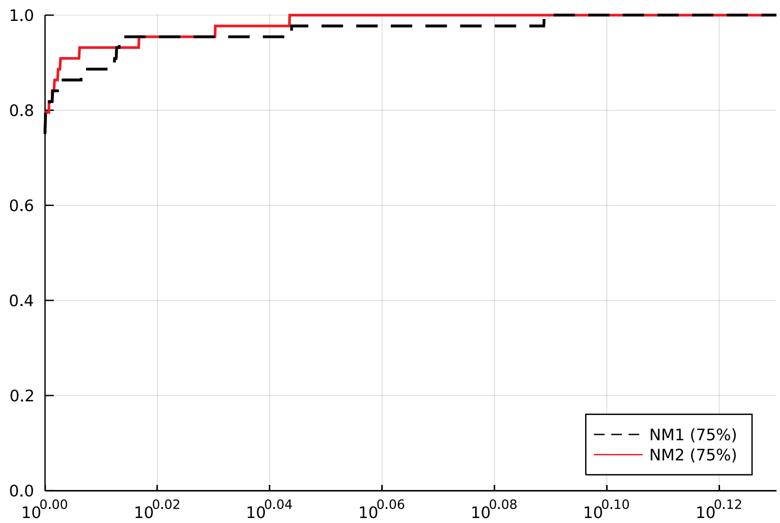


Figure 4.6: Spread Δ performance profiles comparing NM1 and NM2

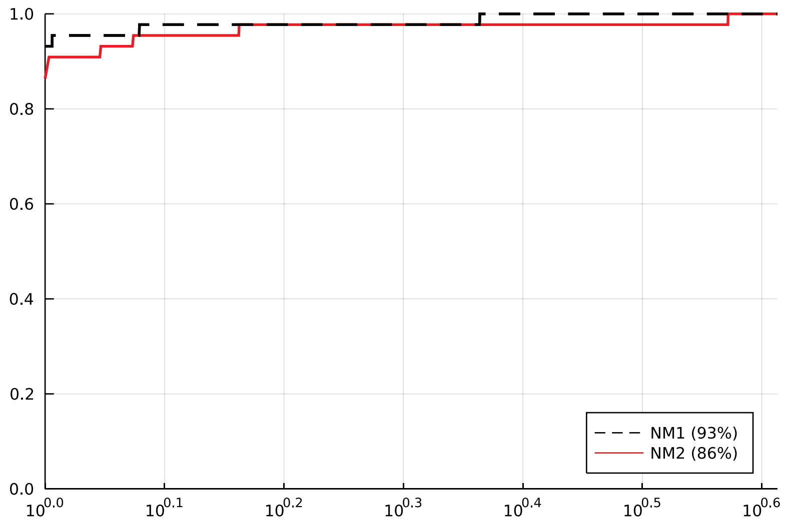


Figure 4.7: Spread Γ performance profiles comparing NM1 and NM2

Finally, with to respect to the Purity Metric, Figure 4.8 shows that NM1 outperformed NM2.

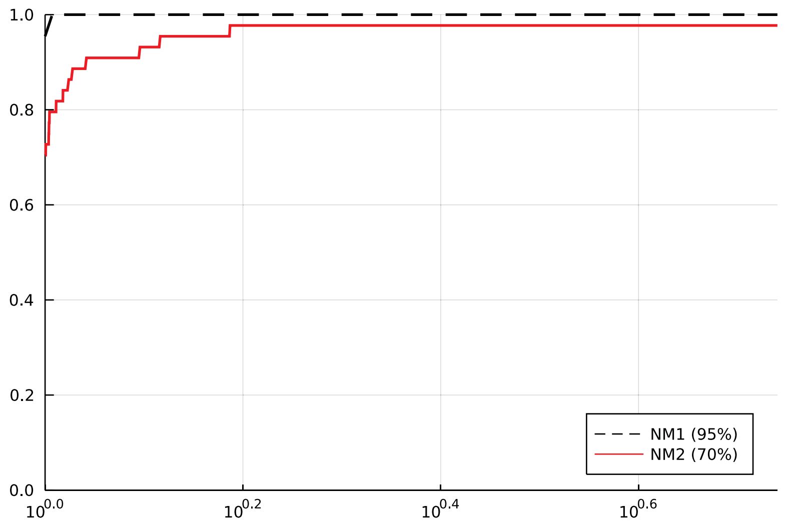


Figure 4.8: Purity performance profile comparing NM1 and NM2.

Comparing $NM1$ and FS

As can be seen in the Figure 4.9 that, FS outperforms NM1.

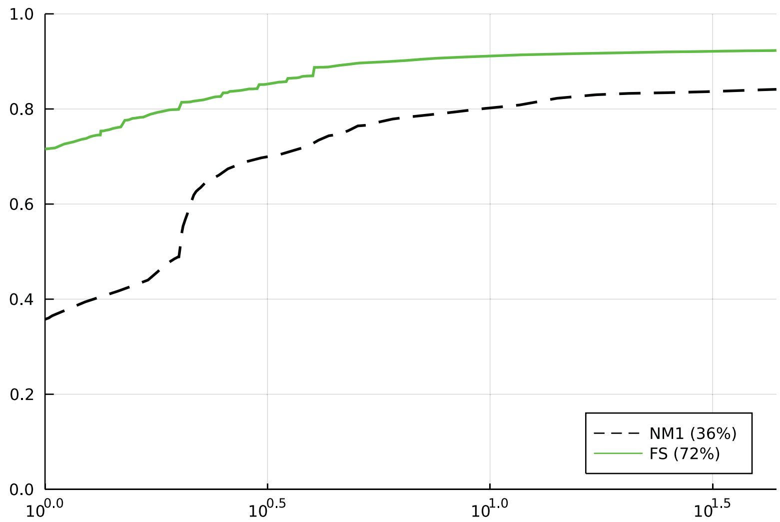


Figure 4.9: Performance profile using number of iterations comparing NM1 and FS.

Regarding the spread metrics, Figures 4.10 and 4.11 show that for spreads performance profile Δ and Γ, FS was the most efficient.

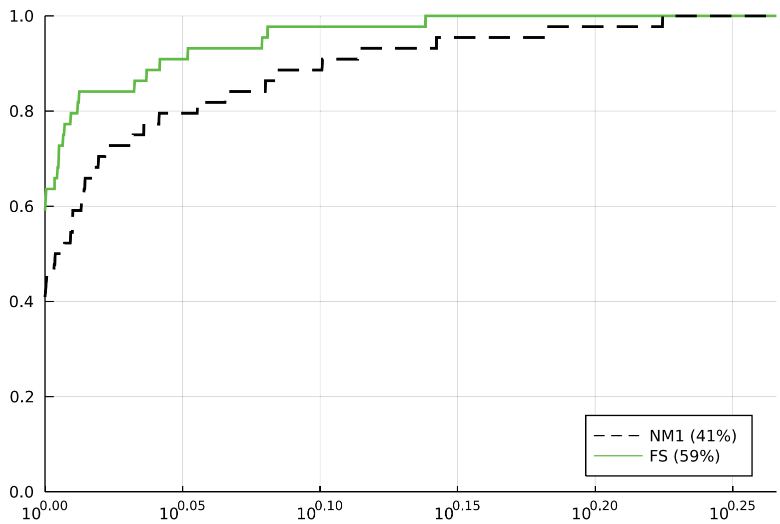


Figure 4.10: Spread Δ performance profiles comparing NM1 and FS

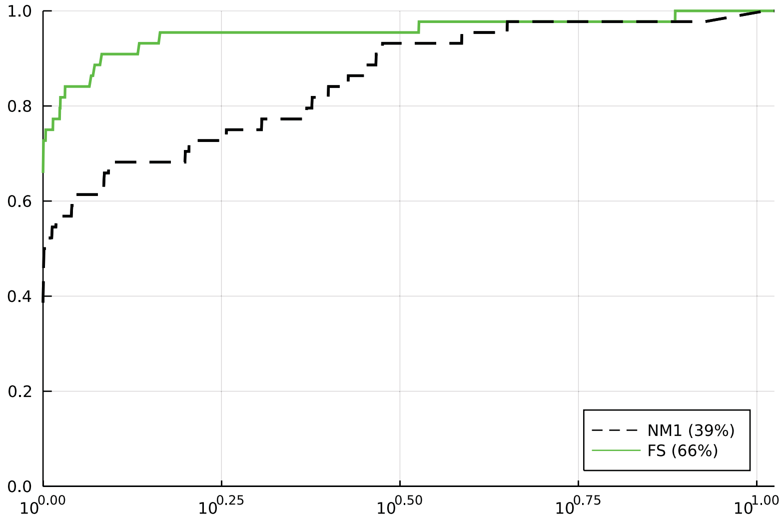


Figure 4.11: Spread Γ performance profiles comparing NM1 and FS

Finally, with to respect to the Purity Metric, Figure 4.12 shows FS is more efficient than NM1.

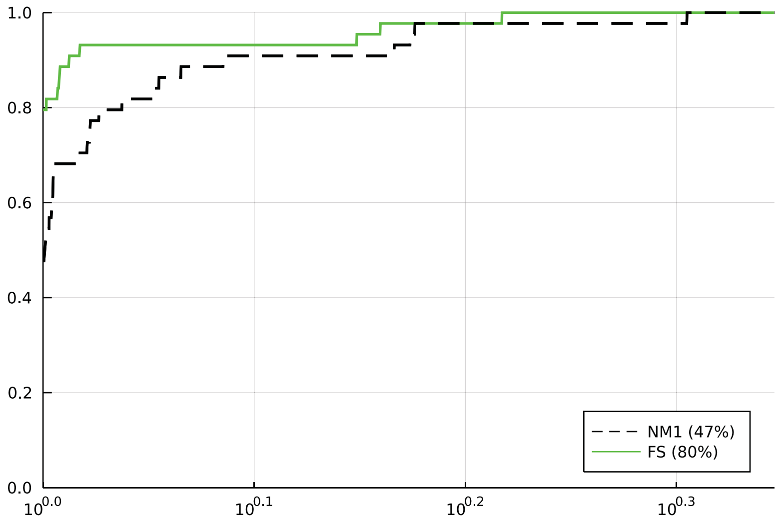


Figure 4.12: Purity performance profile comparing NM1 and FS.

Conclusions

In this dissertation we propose a class of nonmonotone line-search methods to solve multi-objective problems of the form

$$\begin{array}{ll} \min & F(x) = (f_1(x), \ldots, f_m(x))^T \\ \text{s.t} & x \in \Omega, \end{array}$$

where $f_i \in C^{1,\theta_i}_{L_i}(\Omega)$ for $i = 1, \ldots, m$, and $\Omega \in \mathbb{R}^n$ is a closed convex set. The methods are called nonmonotone because they allow the acceptance of x_{k+1} such that $\varphi(F(x_{k+1})) \geq \varphi(F(x_k))$, where $\varphi(v) = \max_{i=1,\ldots,m}\{v_i\}$ is the function using for scalarizing the problem. The level of nonmonotonicity allowed at each iteration is specified by a sequence $\{\nu_k\}_{k\geq 0}$ of non negative numbers. Different choices for this sequence give different nonmonotone methods. The worst-case complexity of the general method was analyzed under very weak assumptions on this sequence.

When the sequence $\{\nu_k\}_{k\geq 0}$ is summable, it was shown that the corresponding methods take at most $\mathcal{O}\left(\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)}\right)$ iterations to generate a $\epsilon-$approximate Pareto critical point, where $\theta_{\min} = \min_{i=1,\ldots,m}\{\theta_i\}$. In particular, if all the component functions have Lipschitz continuous gradients on Ω (i.e., $f_i \in C^{1,1}_{L_i}(\Omega)$ for $i = 1, \ldots, m$) then the complexity bound reduces to $\mathcal{O}(\epsilon^{-2})$, which agrees in order with the bounds obtained in [19] and [42], for monotone methods to unconstrained multi-objective optimization. We also obtained a complexity bound of $\mathcal{O}\left(\epsilon^{-\left(\frac{1}{\theta_{\min}}+1\right)\left(\frac{\rho+1}{\rho}\right)}\right)$ when $\nu_k \leq Mk^{-\rho}$ for all $k \geq 1$, with $\rho \in (0, 1]$.

Exploiting the flexibility of our framework, we designed two new nonmonotone line-search methods for multi-objective optimization. Numerical experiments suggest that one of the methods (NM1) is very competitive in comparison with the monotone version of Algorithm 1 M1, and also in comparison with other nonmonotone solver NM2. Notably, in our numerical tests, NM1 outperformed M1 in all comparisons. This result illustrates the potential benefits of the nonmonotone technique and also the use of our class of methods as a framework for design new methods for multi-objective optimization.

Bibliography

[1] J.Madeira A. Custódio, I. Vaz, and L. Vicente. "Direct Multisearch for Multiobjective Optimization". In: *SIAM Journal on Optimization* 21 (July 2011), pp. 1109–1140.

[2] N. Gould C. Cartis and Ph. Toint. "Adaptive cubic regularisation methods for unconstrained optimization. Part II: Worst-case function- and derivative-evaluation complexity". In: *Math. Program.* 130 (Dec. 2011), pp. 295–319.

[3] C.A.C. Coelho. "A comprehensive survey of evolutionary-based multiobjective optimization techniques". In: *Knowledge and Information Systems* 1 (199), pp. 269–308.

[4] E. Dolan and J. Moré. "Benchmarking Optimization Software with Performance Profiles". In: *Mathematical Programming* 91 (Mar. 2001).

[5] L.M. Drummond and A. Iusem. "A Projected Gradient Method for Vector Optimization Problems". In: *Computational Optimization and Applications* 28 (Apr. 2004), pp. 5–29.

[6] N. Fazzio and M. Schuverdt. "Convergence analysis of a nonmonotone projected gradient method for multiobjective optimization problems". In: *Optimization Letters* 13 (Sept. 2019), pp. 1365–1379.

[7] J. Fliege and B. Svaiter. "Steepest descent methods for multicriteria optimization". In: *Math Methods Oper Res* 51 (Aug. 2000), pp. 479–494.

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

[9] D.M. Frangopol. and M. Liu. "Maitenance and management of civil infraestructure based on condition, safety, optimization, and life-cycle cost." In: *Structure and Infraestructure Engineering* 3 (2007), pp. 29–41.

[10] E. Fukuda and L. Drummond. "A survey on multiobjective descent methods". In: *Pesquisa Operacional* 34 (Sept. 2014), pp. 585–620.

[11] E. Fukuda and L. Drummond. "Inexact projected gradient method for vector optimization". In: *Computational Optimization and Applications* 54 (Apr. 2013), pp. 473–493.

[12] E. Fukuda and L. Drummond. "On the convergence of the projected gradient method for vector optimization". In: *Optimization* 60 (Dec. 2011), pp. 1009–1021.

[13] J. Cruz Neto G. Bento and A. Soubeyran. "A Proximal Point-Type Method for Multicriteria Optimization". In: *Set-Valued and Variational Analysis* 22 (Sept. 2014), pp. 557–573.

[14] P. Lotito G. Carrizo and M. Maciel. "Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem". In: *Mathematical Programming* 159 (Oct. 2015).

[15] G.N. Grapiglia and E.Sachs. "A generalized worst-case complexity analysis for nonmonotone line searches". In: *Numerical Algorithms* 87 (June 2021), pp. 779–796.

[16] A. Guerraggio and D. Luc. "Optimality Conditions for C1,1 Vector Optimization Problems". In: *Journal of Optimization Theory and Applications* 109 (June 2001), pp. 615–629.

[17] Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual*. 2021. URL: `https://www.gurobi.com`.

[18] A. Iusem H. Bonnel and B. Svaiter. "Proximal Methods in Vector Optimization". In: *SIAM Journal on Optimization* 15 (Jan. 2005), pp. 953–970.

[19] I. Vaz J. Fliege and L. Vicente. "Complexity of gradient descent for multiobjective optimization". In: *Optimization Methods and Software* 34 (Aug. 2018), pp. 1–11.

[20] L.M. Graña-Drummond J. Fliege and B.F. Svaiter. "Newton's method for multiobjective optimization." In: *SIAM Journal on Optimization* 20 (2009), pp. 602–626.

[21] D. Kell J. Handl and J. Knowles. "Multiobjective optimization in bioinformatics and computational biology. IEEE/ACM". In: *Transactions on Computational Biology and Bioinformatics* 4 (May 2007), pp. 279–92.

[22] L.R. Lucambio Pérez J.Y. Bello Cruz and J.G Melo. "Convergence of the projected gradient method for quasiconvex multiobjective optimization." In: *Nonlinear Analysis* 74 (2011), pp. 5268–5273.

[23] P.R. Oliveira K.D.V. Villacorta and A. Soubeyran. "A trust-region method for unconstrained multiobjective problems with applications in satis cing processes". In: *Journal of Optimization Theory and Applications* 160 (2014), pp. 865–889.

[24] M. A. Diniz-Ehrhardt L. Calderón and J. M. Martínez. "On high-order model regularization for multiobjective optimization". In: *Optimization Methods and Software* 0.0 (2020), pp. 1–17.

[25] A.N. Iusem L.M Graña-Drummond. "A projected gradient method for vector optimization problems." In: *Computational Optimization and Applications* 28 (2005), pp. 5–29.

[26] F. S. Lima M. L. N. Gonçalves and L. F. Prudente. "Globally convergent Newton-type methods for multiobjective optimization". In: *Optimization Online* (2021).

[27] S.A. Burns Y.K M. Liu and Wen. "Multiobjective optimization for performance-based seismic design of stell moment frame structures." In: *Earthquake Engineering and Structural Dynamics* 34 (2005), pp. 289–306.

[28] R. Marler and J. Arora. "The weighted sum method for multi-objective optimization: New insights". In: *Structural and Multidisciplinary Optimization* 41 (June 2010), pp. 853–862.

[29] R.T. Marler and J.S. Arora. "Survey of multiobjective optimization methods for engineering." In: *Structural and Multidiciplinary Optmization* 26 (2004), pp. 369–395.

[30] K.M. Miettinen. *Nonlinear multiobjective optimization*. Kluwer Academic Publishers, 1998.

[31] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[32] S.K. Pal S. Bandyopadhyay and B. Aruna. "Multi-objective GAs, quantitative indices and pattern classification". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B:Cybernetics* 34 (Jan. 2004), pp. 2009–2088.

[33] E. Sachs and S.M. Sachs. "Nonmonotone line searches for optimization algorithms". In: *Control and Cybernetics* 40 (Jan. 2011), pp. 1059–1075.

[34] G. Schneider and U. Fechner. "Computer-based de novo design of drug-like molecules." In: *Nature Reviews Drug Discovery* 4 (2005), pp. 649–663.

[35] M. Tavana. "A Subjective Assessment of Alternative Mission Architectures for the Human Exploration of Mars at NASA Using Multicriteria Decision Making". In: *Computers & Operations Research* 31 (June 2004), pp. 1147–1164.

[36] K. Ueda and N. Yamashita. "On a Global Complexity Bound of the Levenberg-Marquardt Method". In: *Journal of Optimization Theory and Applications* 147 (Dec. 2010), pp. 443–453.

[37] J.W. Hartmann V. Coverstone and W.J. Mason. "Optimal multi-objective low-thrust spacecraft trajectories". In: *Computer Methods in Applied Mechanics and Engineering* 186 (June 2000), pp. 387–402.

[38] L. Vicente. "Worst case complexity of direct search". In: *EURO Journal on Computational Optimization* 1 (May 2013), pp. 143–153.

[39] M. Yashtini. “On the global convergence rate of the gradient descent method for functions with Hölder continuous gradients”. In: *Optimization Letters* 10 (Aug. 2015).

[40] Y. Yu. “Multiobjective decision theory for computational optimization in radiation therapy”. In: *Medical physics* 24 (1997), pp. 445–1454.

[41] G.N. Grapiglia J. Yuan and Y. Yuan. “A derivative-free trust-region algorithm for composite nonsmooth optimization”. In: *Computational and Applied Mathematics* 35 (Dec. 2014), pp. 1–25.

[42] G.N. Grapiglia J. Yuan and Y. Yuan. “On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization”. In: *Mathematical Programming* 152 (July 2014).

[43] H. Zhang and W. W. Hager. “A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization”. In: *SIAM J. Optim.* 14 (2004), pp. 1043–1056.