On High-order Model Regularization for Multiobjective Optimization *

L. Calderón M. A. Diniz-Ehrhardt J. M. Martínez[†]

December 17, 2018

Abstract

A p-order regularization method for finding weak stationary points of multiobjective optimization problems with constraints is introduced. Under Hölder conditions on the derivatives of the objective functions, complexity results are obtained that generalize properties recently proved for scalar optimization.

1 Introduction

Multiobjective optimization (MO) is frequently used for solving engineering and applied mathematics problems [1, 33, 35, 37, 42]. Given a set of objective functions $\{f_i\}_{i \in I}$ and a feasible set Ω, a solution of the MO problem is a feasible point x^* for which the proposition

$$f_i(x) < f_i(x^*) \text{ for all } x \in \Omega \text{ and } i \in I \tag{1}$$

is false. In the MO terminology the property above defines "weak Pareto" points. Stronger definitions of optimality have been given in several works (see [37]). The so called "Pareto front" defines the set of solutions of MO, among which engineers and practitioners usually choose the implementable alternative, taking into account additional, frequently subjective, criteria.

Algorithms for solving nonlinear MO problems are iterative. Given an iterate x^k, a popular approach consists of finding x^{k+1} such that $f_i(x^{k+1})$ is sufficiently small than $f_i(x^k)$ for all $i \in I$. This is done defining a model for each $f_i(x) - f_i(x^k)$ around x^k and minimizing, approximately, the maximum of those models with restricted steps. Fliege and Svaiter [20] seem to be the first in using this approach for solving MO problems. If the step is sufficiently reduced, suitable convergence theorems can be proved. Linear models define gradient-like methods and quadratic models define Newton or quasi-Newton methods. See [19, 20, 21, 25, 27, 28, 29, 30, 31, 36].

When the constraints of the MO problem define a polytope, or even a convex set, the process of finding approximate minimizers of the model onto the feasible set may be simple and

*This work was supported by Cepid-Cemeai-Fapesp, PRONEX-CNPq/FAPERJ E-26/111.449/2010-APQ1, FAPESP (grants 2010/10133-0, 2013/03447-6, 2013/05475-7, and 2013/07375-0), and CNPq.

[†]Department of Applied Mathematics, Institute of Mathematics, Statistics, and Scientific Computing (IMECC), State University of Campinas, 13083-859 Campinas SP, Brazil. e-mail: martinez@ime.unicamp.br

successive trial points may be obtained even using line searches. In the presence of generally nonconvex constraints line searches do not preserve feasibility and, thus, should be replaced with trust-region or regularization approaches. Fortunately, solving subproblems within certain nonconvex constraints is, many times, affordable. This is the case of constraints that define spheres, intersections of spheres with other simple sets, density matrices in electronic calculations [22, 23] and others [24]. Moreover, if the objective functions are very expensive in comparison to model evaluations, even iterative methods are acceptable for solving subproblems, provided that we have an adequate criterion for defining approximate solutions.

Our problem in this paper is solving MO problems with constraints. Roughly speaking, the only condition for the constraints is that they should be simple enough to make it possible the approximate solution of subproblems generating feasible solutions with mild optimality satisfaction. Given an iterate x^k our method will minimize, approximately, a model of the original problem plus a regularization term, subject to the constraints. We will prove that, if the regularization term is large enough, sufficient descent of each objective function occurs. In that case, the corresponding trial point is accepted as new iterate.

If the model of the objective function f_i around x^k agrees with $f_i(x)$ with error $O(\|x-\bar{x}\|^{p+1})$, we may employ regularization terms of the form $\sigma\|x - x^k\|^{p+1}$. Moreover, even in the case that the error is $O(\|x - \bar{x}\|^{p+\beta})$ with $\beta < 1$, provided that $p + \beta > 1$, the employment of $p + 1$-regularization is admissible. For example, errors of the form $O(\|x - \bar{x}\|^{p+\beta})$ occur when the model is the $p - th$ Taylor approximation of $f_i(x)$ and the $p-$th derivatives are not Lipschitz continuous but are Hölder continuous with parameter β. The algorithm introduced in this paper will be proved to produce approximate solutions with precision ε employing $O(\varepsilon^{-\frac{p+\beta}{p+\beta-1}})$ iterations and functional evaluations. For the scalar case, this property has been proved in [14]. Grapiglia and Nesterov [32] addressed the unconstrained scalar case with $p = 2$ whereas Cartis, Gould, and Toint [14] considered the scalar case with convex constraints with arbitrary p with similar complexity results. The idea of consider high-order Taylor models ($p > 2$) for unconstrained regularization methods comes from [3]. For unconstrained MO problems and MO problems with convex constraints, complexity results in the case $p = 1$ were given in [29, 30, 21].

Notation

$\|\cdot\|$ denotes the Euclidean norm.

If $v, w \in \mathbb{R}^n$, $\min\{v, w\}$ denotes the vector with components $\min\{v_1, w_1\}, \ldots, \min\{v_n, w_n\}$.

If $v \in \mathbb{R}^n$, v_+ is the vector with components $\max\{v_1, 0\}, \ldots, \max\{v_n, 0\}$.

If $v, w \in \mathbb{R}^n$, we denote $v \leq w$ if $v_i \leq w_i$ for all $i = 1, \ldots, n$. Analogously, $v < w$ means that $v_i < w_i$ for all $i = 1, \ldots, n$.

2 Main results

Assume that $F : \mathbb{R}^n \to \mathbb{R}^r$, $g_i : \mathbb{R}^n \to \mathbb{R}^n$ for all $i = 1, \ldots, r$, $h_E : \mathbb{R}^n \to \mathbb{R}^m$, and $h_I : \mathbb{R}^n \to \mathbb{R}^q$. Our problem will be

$$\text{Minimize } F(x) \text{ subject to } h_E(x) = 0 \text{ and } h_I(x) \leq 0. \tag{2}$$

The points that satisfy the constraints $h_E(x) = 0$ and $h_I(x) \leq 0$ will be called feasible. The components of F will be denoted $f_1, \ldots, f_r$. We consider that x^* is a solution of (2) if, for all feasible $x \in \mathbb{R}^n$, the statement $F(x) < F(x^*)$ is false.

Throughout this paper we will assume that $p \in \{1, 2, 3, \ldots\}$, $L > 0$, $\delta > 0$, and $\beta \in [0, 1]$ with $p + \beta > 1$. We will also assume that g_i denotes the gradient of f_i, while $h'_E(x)$ and $h'_I(x)$ are the Jacobians of $h_E(x)$ and $h_I(x)$, respectively.

For all $\bar{x} \in \mathbb{R}^n$, $i \in \{1, \ldots, r\}$, we define $M_i^{\bar{x}} : \mathbb{R}^n \to \mathbb{R}$ (intended to be a "model" of $f_i(x)$ around $\bar{x}$). The derivatives of $M_i^{\bar{x}}$ will be assumed to exist for all $x \in \mathbb{R}^n$. We say that $\bar{x}$ and x satisfy the assumptions (3) and (4) when

$$\|g_i(x) - \nabla M_i^{\bar{x}}(x)\| \leq L\|x - \bar{x}\|^{p+\beta-1} \tag{3}$$

and

$$M_i^{\bar{x}}(\bar{x}) = f_i(\bar{x}) \text{ and } f_i(x) \leq M_i^{\bar{x}}(x) + L\|x - \bar{x}\|^{p+\beta} \tag{4}$$

for all $i = 1, \ldots, r$.

If $M_i^{\bar{x}}(x)$ is the Taylor polynomial of order p around $\bar{x}$ and the $p-$th derivatives satisfy a Hölder condition defined by β and L the assumptions (3) and (4) are satisfied for all $\bar{x}, x \in \mathbb{R}^n$ (see, for example, [2]).

Algorithm 2.1

Assume that $x^0 \in \mathbb{R}^n$ is feasible, $\alpha \in (0, 1)$, $\varepsilon \in (0, 1)$, $\delta > 0$, $\theta > 0$, and $\sigma_{min} > 0$,
Initialize $k \leftarrow 0$ and $\sigma_0 = \sigma_{min}$.

Step 1. Set $\sigma \leftarrow \sigma_k$.

Step 2. Find $x \in \mathbb{R}^n$, $\gamma \in \mathbb{R}^r$, $\lambda \in \mathbb{R}^m$, and $\mu \in \mathbb{R}^q_+$ such that, if $\bar{x} = x^k$,

$$M_i^{\bar{x}}(x) + \sigma\|x - \bar{x}\|^{p+1} \leq f_i(\bar{x}) \quad \text{for all} \quad i = 1, \ldots, r, \tag{5}$$

$$\gamma \geq 0, \sum_{i=1}^{r} \gamma_i = 1, \tag{6}$$

and

$$\|\sum_{i=1}^{r} \gamma_i \nabla[M_i^{\bar{x}}(x) + \sigma\|x - \bar{x}\|^{p+1}] + h'_E(x)^T \lambda + h'_I(x)^T \mu\| \leq \theta\|x - \bar{x}\|^p, \tag{7}$$

where

$$\|\min\{\mu, -h_I(x)\}\| \leq \delta, \tag{8}$$

$$\|h_E(x)\| = 0, \text{ and } \|h_I(x)_+\| = 0. \tag{9}$$

Step 3. If $\| \sum_{i=1}^r \gamma_i g_i(x) + h_E'(x)^T \lambda + h_I'(x)^T \mu \| \le \varepsilon$, stop.

Step 4. Test the sufficient descent condition

$$f_i(x) \le f_i(x^k) - \frac{\alpha}{(2p+4)^{\frac{p+1}{p}}} \frac{\varepsilon^{\frac{p+1}{p}}}{\sigma^{\frac{1}{p}}} \text{ for all } i = 1, \dots, r. \tag{10}$$

If (10) does not hold, set $\sigma \leftarrow 2\sigma$ and go to Step 2. Else, continue at Step 5.

Step 5. Define $x^{k+1} = x$, $k \leftarrow k+1$, $\sigma_k = \sigma$, and go to Step 1.

Let us discuss the plausibility of the conditions (5)–(9). If $\bar{x}$ is not a solution of (2), there exists a feasible $x \in \mathbb{R}^n$ such that $f_i(x) < f_i(\bar{x})$ for all $i = 1, \dots, r$. Since $M_i^{\bar{x}}(x)$ is a model for $f_i(x)$, it is reasonable to ask for a feasible x such that the components of $M^{\bar{x}}(x)$ decrease as much as possible. This leads to minimize $\max\{M_1^{\bar{x}}(x) - M_1^{\bar{x}}(\bar{x}), \dots, M_r^{\bar{x}}(x) - M_r^{\bar{x}}(\bar{x})\}$ subject to feasibility. Equivalently, we may wish to solve, approximately,

$$\text{Minimize } z \tag{11}$$

subject to

$$z \ge M_i^{\bar{x}}(x) - M_i^{\bar{x}}(\bar{x}), i = 1, \dots, r, \tag{12}$$

$$h_E(x) = 0, \text{ and } h_I(x) \le 0. \tag{13}$$

If the constraints (12)–(13) satisfy a constraint qualification, a solution of (11)–(13) satisfy the KKT conditions. But the KKT conditions of (11)–(13) are the conditions (6)–(9) with $\sigma = 0$, $\theta = 0$ and $\delta = 0$. Therefore, (6)–(9) are approximate KKT conditions for the minimization of $\max\{M_1^{\bar{x}}(x) - M_1^{\bar{x}}(\bar{x}), \dots M_r^{\bar{x}}(x) - M_r^{\bar{x}}(\bar{x})\} + \sigma\|x - \bar{x}\|^{p+1}$ subject to $h_E(x) = 0$ and $h_I(x) \le 0$. Since $M^{\bar{x}}(x)$ is a model for $F(x)$, it can be expected that, restricting the step by means of regularization, the condition (5) will also hold.

Theorem 2.1 *Assume that x^k is an iterate computed by Algorithm 2.1 and (3), (4), (5), (6), (7), (8), and (9) are satisfied by $\bar{x} = x^k$, x, γ, λ, and μ. Then,*

$$\| \sum_{i=1}^r \gamma_i g_i(x) + h_E'(x)^T \lambda + h_I'(x)^T \mu \| \le (\theta + (p+1)\sigma)\|x - \bar{x}\|^p + L\|x - \bar{x}\|^{p+\beta-1}. \tag{14}$$

Moreover, if

$$\| \sum_{i=1}^r \gamma_i g_i(x) + h_E'(x)^T \lambda + h_I'(x)^T \mu \| i \ge \varepsilon \tag{15}$$

and

$$\sigma \ge \max\left\{\theta, \varepsilon^{\frac{\beta-1}{p+\beta-1}} \max\left\{\frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}}((2p+4)^{(\beta-1)/(p+\beta-1)}}\right\}\right\} \tag{16}$$

we have that

$$\sigma\|x-\bar{x}\|^p \geq \frac{\varepsilon}{2p+4}, \tag{17}$$

$$f_i(x) \leq f_i(\bar{x}) - \alpha\sigma\|x-\bar{x}\|^{p+1} \tag{18}$$

and

$$f_i(x) \leq f_i(\bar{x}) - \frac{\alpha}{(2p+4)^{\frac{p+1}{p}}}\frac{\varepsilon^{\frac{p+1}{p}}}{\sigma^{\frac{1}{p}}} \tag{19}$$

for all $i = 1, \ldots, r$.

Proof By (3), for all $i = 1, \ldots, r$,

$$\|\gamma_i g_i(x) - \gamma_i \nabla M_i^{\bar{x}}(x)\| \leq \gamma_i L\|x-\bar{x}\|^{p+\beta-1}.$$

Then, by (6),

$$\sum_{i=1}^{r} \|\gamma_i g_i(x) - \gamma_i \nabla M_i^{\bar{x}}(x)\| \leq L\|x-\bar{x}\|^{p+\beta-1}.$$

Then,

$$\|\sum_{i=1}^{r} \gamma_i g_i(x) - \sum_{i=1}^{r} \gamma_i \nabla M_i^{\bar{x}}(x)\| \leq L\|x-\bar{x}\|^{p+\beta-1}.$$

So,

$$\|\sum_{i=1}^{r} \gamma_i g_i(x) + h'_E(x)^T\lambda + h'_I(x)^T\mu\| \leq \|\sum_{i=1}^{r} \gamma_i \nabla M_i^{\bar{x}}(x) + h'_E(x)^T\lambda + h'_I(x)^T\mu\| + L\|x-\bar{x}\|^{p+\beta-1}. \tag{20}$$

Then, by (7) and (20),

$$\|\sum_{i=1}^{r} \gamma_i g_i(x) + h'_E(x)^T\lambda + h'_I(x)^T\mu\|$$

$$\leq \|[\sum_{i=1}^{r} \gamma_i \nabla M_i^{\bar{x}}(x) + \nabla\sigma\|x-\bar{x}\|^{p+1}] + h'_E(x)^T\lambda + h'_I(x)^T\mu\| + \|\nabla[\sigma\|x-\bar{x}\|^{p+1}]\| + L\|x-\bar{x}\|^{p+\beta-1}$$

$$\leq \|[\sum_{i=1}^{r} \gamma_i \nabla M_i^{\bar{x}}(x) + \nabla\sigma\|x-\bar{x}\|^{p+1}] + h'_E(x)^T\lambda + h'_I(x)^T\mu\| + (p+1)\sigma\|x-\bar{x}\|^p + L\|x-\bar{x}\|^{p+\beta-1}$$

$$\leq \theta\|x-\bar{x}\|^p + (p+1)\sigma\|x-\bar{x}\|^p + L\|x-\bar{x}\|^{p+\beta-1}.$$

Therefore, (14) is proved.

By (14) and (15), we have that

$$\varepsilon \leq \theta\|x-\bar{x}\|^p + (p+1)\sigma\|x-\bar{x}\|^p + L\|x-\bar{x}\|^{p+\beta-1}.$$

Then, by (16), since $\sigma \geq \theta$,

$$(p+2)\sigma\|x-\bar{x}\|^p + L\|x-\bar{x}\|^{p+\beta-1} \geq \varepsilon. \tag{21}$$

By (16) and (21), after some algebraic manipulations, we obtain (17). (The detailed deduction of this fact follows mimicking the paragraph in the proof of Lemma 2.1 of [38] in which inequality (17) of [38]) is deduced.)

Let us now prove (18). By (4) and (5), for all $i = 1, \ldots, r$,

$$f_i(x) \leq M_i^{\bar{x}}(x) + L\|x-\bar{x}\|^{p+\beta}$$

$$\leq M_i^{\bar{x}}(x) + \sigma\|x-\bar{x}\|^{p+1} - \sigma\|x-\bar{x}\|^{p+1} + L\|x-\bar{x}\|^{p+\beta} \leq f_i(\bar{x}) - \sigma\|x-\bar{x}\|^{p+1} + L\|x-\bar{x}\|^{p+\beta}.$$

Thus, for proving (18) it is enough to prove that

$$\sigma\|x-\bar{x}\|^{p+1} - L\|x-\bar{x}\|^{p+\beta} \geq \alpha\sigma\|x-\bar{x}\|^{p+1}. \tag{22}$$

The proof of (22) follows from the algebraic manipulations employed in [38] for proving inequality (22) of that paper.

Therefore, by (16), the proof of (18) is complete.

Now let us prove (19). We proceed as in [38]. By (17) we have that

$$\|x-\bar{x}\| \geq \frac{\varepsilon^{1/p}}{(2p+4)^{1/p}\sigma^{1/p}}.$$

Therefore,

$$\alpha\sigma\|x-\bar{x}\|^{p+1} \geq \frac{\alpha}{(2p+4)^{\frac{p+1}{p}}} \frac{\varepsilon^{(p+1)/p}}{\sigma^{1/p}}.$$

As a consequence, (19) follows from (18). This completes the proof. □

Theorem 2.1 shows that, given an iterate x^k computed by Algorithm 2.1, either the stopping criterion is satisfied at some trial point, or the iteration k finishes computing x^{k+1} that satisfies the sufficient descent condition(10), which states that all the functions f_i decrease uniformly.

Theorem 2.2 *Assume that the hypotheses of Theorem 2.1 are satisfied. Then, the iterate x^{k+1} is well defined and, defining*

$$c_p = \min\left\{ \frac{1}{(2p+4)^{\frac{p+1}{p}}} \frac{1}{(2\theta)^{\frac{1}{p}}}, \frac{1}{(2p+4)^{\frac{p+1}{p}} \left\{ 2\max\left\{ \frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}} ((2p+4)^{(\beta-1)/(p+\beta-1)})} \right\} \right\}^{\frac{1}{p}}} \right\}, \tag{23}$$

we have that, for all $i = 1, \ldots, r$,

$$f_i(x^{k+1}) \leq f_i(x^k) - \alpha c_p \varepsilon^{\frac{p+\beta}{p+\beta-1}}. \tag{24}$$

Proof We follow closely the proof of Theorem 2.1 of [38]. Define

$$\sigma_{max} = 2\max\left\{\theta, \varepsilon^{\frac{\beta-1}{p+\beta-1}} \max\left\{ \frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}}((2p+4)^{(\beta-1)/(p+\beta-1)})} \right\}\right\} \tag{25}$$

By Theorem 2.1 we have that $\sigma_k \leq \sigma_{max}$ for all k. Therefore, by (19), at each iteration we have:

$$f_i(x^{k+1}) \leq f_i(x^k) - \frac{\alpha}{(2p+4)^{\frac{p+1}{p}}} \frac{\varepsilon^{\frac{p+1}{p}}}{\sigma_k^{\frac{1}{p}}} \leq f_i(x^k) - \frac{\alpha}{(2p+4)^{\frac{p+1}{p}}} \frac{\varepsilon^{\frac{p+1}{p}}}{\sigma_{max}^{\frac{1}{p}}}.$$

Therefore, by (25), at each iteration we have that either

$$f_i(x^{k+1}) \leq f_i(x^k) - \frac{\alpha}{(2p+4)^{\frac{p+1}{p}}} \frac{\varepsilon^{\frac{p+1}{p}}}{(2\theta)^{\frac{1}{p}}} \tag{26}$$

or

$$f_i(x^{k+1}) \leq f_i(x^k) - \alpha \frac{\varepsilon^{\frac{p+1}{p}}}{(2p+4)^{\frac{p+1}{p}} \left\{ 2\varepsilon^{\frac{\beta-1}{p+\beta-1}} \max\left\{ \frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}}((2p+4)^{(\beta-1)/(p+\beta-1)})} \right\}\right\}^{\frac{1}{p}}}. \tag{27}$$

In the case of (26) the thesis holds by the definition of c_p and $\varepsilon \leq 1$. If (27) takes place, we have that:

$$\begin{aligned} f_i(x^{k+1}) &\leq f_i(x^k) - \alpha \frac{\varepsilon^{\frac{p+1}{p} + \frac{1-\beta}{p(p+\beta-1)}}}{(2p+4)^{\frac{p+1}{p}} \left\{ 2\max\left\{ \frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}}((2p+4)^{(\beta-1)/(p+\beta-1)})} \right\}\right\}^{\frac{1}{p}}} \\ &\leq f_i(x^k) - \alpha c_p \varepsilon^{\frac{p+\beta}{p+\beta-1}}. \end{aligned}$$

This completes the proof. □

Theorem 2.3 *Assume that the hypotheses of Theorem 2.1 hold for all x^k computed by Algorithm 2.1. Let $f_{target} \in \mathbb{R}$ be arbitrary and $i \in \{1, \ldots, r\}$. Let c_p be defined by (23). Then, after, at most,*

$$(f_i(x^0) - f_{target}) \frac{\varepsilon^{-\frac{p+\beta}{p+\beta-1}}}{\alpha c_p} \tag{28}$$

iterations, Algorithm 2.1 computes a trial point $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^m$, and $\mu \in \mathbb{R}^q_+$ verifying $f_i(x) \leq f_{target}$ or

$$\|g(x) + h_E'(x)^T\lambda + h_I'(x)^T\mu\| \leq \varepsilon, \tag{29}$$

$$\|h_E(x)\| = 0, \|h_I(x)_+\| = 0, \text{ and } \|\min\{\mu, -h_I(x)\}\| \leq \delta. \tag{30}$$

Proof The desired result on the number of iterations follows from (8), (9), Theorem 2.1, and the stopping criterion at Step 3 of the algorithm. Moreover, observe that, excluding the gradient evaluations for testing the stopping criterion at Step 3, Algorithm 2.1 requires to evaluate derivatives of f_i only at the iterates x^k (and not at the rejected trial points x). Therefore, the second part of the thesis is also proved. □

The following corollary states that, if at least one of the functions $f_1, \dots f_r$ is bounded below, Algorithm 2.1 stops employing $O(\varepsilon^{-\frac{p+\beta}{p+\beta-1}})$ iterations and gradient evaluations.

Corollary 2.1 *Assume that the hypotheses of Theorem 2.1 hold for all x^k computed by Algorithm 2.1. If $i \in \{1, \dots, r\}$ and $f_{target} \in \mathbb{R}$ are such that $f_i(x) > f_{target}$ for all $x \in \mathbb{R}^n$ such that $h_E(x) = 0$ and $h_I(x) \leq 0$, Algorithm 2.1 stops employing, at most the number of iterations given in (28).*

Proof The desired results follows straightforwardly from Theorem 2.3. □

Theorem 2.3 and Corollary 2.1 concern the number of iterations performed by Algorithm 2.1. The computer work associated to each iteration involves the evaluation of the gradients at x^k (for testing the stopping criterion at Step 3) and, perhaps, derivatives of higher order. Independently of this work, at each iteration we need to evaluate $f_1, \dots, f_r$ for different trial points x. This means that, in addition to the computer work related with building the model $M^{x^k}(x)$, we must consider the function evaluations at the rejected trial points. This corresponds to the number of times in which the regularization parameter σ needs to be increased. By Theorem 2.1, increasing σ will not necessary if $\sigma \geq \sigma_{max}$. As consequence, an upper bound for the number of functional evaluations is given in the following theorem.

Theorem 2.4 *Assume that the hypotheses of Theorem 2.3 hold. Then, the number of evaluations of F employed by Algorithm 2.1 before obtaining $f_i(x^k) < f_{target}$ is bounded above by*

$$(f_i(x^0) - f_{target}) \frac{\varepsilon^{-\frac{p+\beta}{p+\beta-1}}}{\alpha c_p} + \left[\max\left\{ \log_2(\theta), \left(\frac{1-\beta}{p+\beta-1} \log_2(\varepsilon^{-1}) + c_\ell \right) \right\} \right] - \log_2(\sigma_{min}) + 1. \quad (31)$$

where

$$c_\ell = \log_2(\max\left\{ \frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}} ((2p+4)^{(\beta-1)/(p+\beta-1)}} \right\}).$$

Proof By Theorem 2.2, the evaluation of f_i and its derivatives at each iterate x^k is responsible for the first term of (31). The remaining functional evaluations are at trial points at which the descent criterion (10) does not hold. Each time (10) fails, the regularization parameter σ is doubled. However, by Theorem 2.1, if $\sigma \geq \sigma_{max}$, (10) necessarily takes place. This means that the number of times at which σ is doubled is bounded above by $\log_2(\sigma_{max}/\sigma_{min})$. Therefore, by (25), the second term of (31) is

$$\log_2 \left[2 \max\left\{ \theta, \varepsilon^{\frac{\beta-1}{p+\beta-1}} \max\left\{ \frac{2^{\frac{-\beta+1}{p+\beta-1}}}{p+2} L^{\frac{p}{p+\beta-1}}, \frac{L^{\frac{p}{p+\beta-1}}}{(1-\alpha)^{\frac{p}{p+\beta-1}} ((2p+4)^{(\beta-1)/(p+\beta-1)}} \right\} \right\} \right] - \log_2(\sigma_{min})$$

$$= \left[\max \left\{ \log_2(\theta), \left(\frac{\beta - 1}{p + \beta - 1} \log_2(\varepsilon) + c_\ell \right) \right\} \right] - \log_2(\sigma_{min}) + 1.$$

This completes the proof. □

3 Conclusions

Multiobjective optimization techniques aim to produce a sufficient number of alternative solutions, among which the decision makers will choose the most adequate one employing, probably, qualitative criteria. Scalarization techniques achieve this objective based on the observation that each point of the Pareto front is the solution of a scalar (single objective) optimization problem, that depends of a set of parameters. So, different choices of parameters produce different efficient solutions and, given a particular choice, the corresponding solution comes from solving a standard scalar optimization problem. See [1, 9, 17, 18, 26, 33, 34, 35, 37, 41, 42] and many others.

In the approach of [19, 20, 25, 27, 29, 30, 31, 36] and the present paper, the diversity of points in the efficient set is obtained as a consequence of the diversity of initial approximations. Probably, different alternative solutions could be also obtained by means of a clever use of ad hoc constraints. Another situation in which this approach could be useful is when one tries to optimize some additional function $\varphi(x)$ onto the efficient set [7, 8, 16]. In this case, constraints of type $\varphi(x) \leq c$ could help to approximate acceptable solutions without increasing the complexity of a single multiobjective optimization.

After the seminal work of Nesterov and Polyak [40], complexity results for unconstrained and constrained scalar optimization problems using regularization were proved in many papers. See [4, 5, 10, 11, 12, 13, 14, 15, 32, 39], among others. Frequently, these results motivate the introduction of implementable and efficient algorithms. It can be expected that, in the next years, the complexity tendency will be extended to non-standard optimization problems, as equilibrium, bilevel, and multilevel optimization.

References

[1] V. Belton and T. J. Stewart, *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic Publishers, Boston, 2002.

[2] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, Massachussets, USA, 1999.

[3] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint, *Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models*, Mathematical Programming 163 (2017), pp. 359-368.

[4] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and Ph. L. Toint, *Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models*, SIAM Journal on Optimization 26 (2016), pp. 951-967.

[5] E. G. Birgin and J. M. Martínez, *The use of quadratic regularization with cubic descent for unconstrained optimization*, SIAM Journal on Optimization 27 (2017), pp. 1049-1074.

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

[7] H. Bonnel and C. Y. Kaya, *Optimization over the efficient set of multi-objective convex optimal control problems*, J. Optim. Theory Appl. 147 (2010), pp. 93–112.

[8] L. F. Bueno, G. Haeser, and J. M. Martínez, *An inexact restoration approach to optimization problems with multiobjective constraints under weighted-sum scalarization*, Optimization Letters 10 (2016), pp. 1315-1325.

[9] R. S. Burachik, C. Y. Kaya, and M. M. Rizvi, *A new scalarization technique to approximate Pareto fronts on problems with disconnected feasible sets*, J. Optim. Theory Appl. 162(2) (2014), pp. 428–446.

[10] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization*, SIAM Journal on Optimization 20 (2010), pp. 2833–2852.

[11] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *Adaptive cubic regularization methods for unconstrained optimization. Part I: motivation, convergence and numerical results*, Mathematical Programming 127 (2011), pp. 245–295.

[12] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *Adaptive cubic regularization methods for unconstrained optimization. Part II: worst-case function and derivative complexity*, Mathematical Programming 130 (2011), pp. 295–319.

[13] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *On the complexity of finding first-order critical points in constrained nonlinear optimization*, Mathematical Programming 144 (2014), pp. 93–106.

[14] C. Cartis, N. I. M. Gould, and Ph. L. Toint, *Universal regularization methods - varying the power, the smoothness and the accuracy*, to appear in SIAM J. Optim.

[15] F. E. Curtis, D. P. Robinson, and M. Samadi, *A trust-region algorithm with a worst-case iteration complexity of* $O(\varepsilon^{-3/2})$, Mathematical Programming 162 (2017), pp. 1-32.

[16] J. P. Dauer and T. A. Fosnaugh, *Optimization over the efficient set*, J. Global Optim. 7 (1995), pp. 261–277.

[17] J. Dutta and C. Y. Kaya, *A new scalarization and numerical method for constructing weak Pareto front of multi-objective optimization problems*, Optim. 60 (2011), pp. 1091–1104.

[18] G. Eichfelder, *Scalarization Methods in Multiobjective Optimization*, Springer, Berlin, Heidelberg, 2008.

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

[20] Fliege and B. F. Svaiter, *Steepest descent methods for multiobjective optimization*, Math. Methods Oper. Res. 51 (2000), pp. 479–494.

[21] J. Fliege, A. I. F. Vaz, and L. N. Vicente, *Complexity of gradient descent for multiobjective optimization*, to appear in Optimization Methods and Software.

[22] J. B. Francisco, J. M. Martínez, and L. Martínez, *Globally convergent Trust-Region methods for Self-Consistent Field electronic structure calculations*, Journal of Chemical Physics 121 (2004), pp. 10863-10878.

[23] J. B. Francisco, J. M. Martínez, and L. Martínez, *Density-Based Globally Convergent Trust-Region Method for Self-Consistent Field Electronic Structure Calculations*, Journal of Mathematical Chemistry 40 (2006), pp. 349-377.

[24] J. B. Francisco and F. S. Viloche-Bazán, *Nonmonotone algorithm for minimization on closed sets with application to minimization on Stiefel manifolds*, Journal of Computational and Applied Mathematics 236 (2012), pp. 2717-2727.

[25] E. H. Fukuda and L. M. Graña Drummond, *On the convergence of the projected gradient method for vector optimization*, Optimization 60 (2011), pp. 1009–102.

[26] A. M. Geoffrion, *Proper efficiency and the Theory of Vector Maximization*, J. Math. Anal. Appl. 22 (1968), pp. 387–407.

[27] L. M. Graña Drummond and A. N. Iusem, *A projected gradient method for vector optimization problems*, Comput. Optim. Appl. 28 (2004), pp. 5–29.

[28] L. M. Graña Drummond, F. M. P. Raupp, and B. F. Svaiter, *A quadratically convergent Newton method for vector optimization*, Optimization 63 (2014), pp. 651–677.

[29] G. N. Grapiglia, *Três contribuições em Otimização Não Linear e Não Convexa*, Ph.D. thesis, Universidade Federal do Paraná, 2014. In Portuguese.

[30] G. N. Grapiglia, *On the worst-case complexity of projected gradient methods for convexly constrained multiobjective optimization*, manuscript cited in [21], 2016.

[31] G. N. Grapiglia, J-Y. Yuan, and Y-X Yuan, *On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization*, Math. Program. 152 (2015), pp. 491–520.

[32] G. N. Grapiglia and Yu. Nesterov, *Regularized newton methods for minimizing functions with Hölder continuous hessians*, SIAM Journal on Optimization 27 (2017), pp. 478–506.

[33] J. Haslinger and R. A. E. Mäkinen, *Introduction to Shape Optimization*, SIAM Publications, Philadelphia, 2003.

[34] J. Jahn, *Vector Optimization*, Springer-Verlag Berlin Heidelberg, 2011.

[35] D. T. Luc, *Theory of Vector Optimization*, Lecture Notes in Economy and Mathematical Systems, Vol. 319, Springer, Berlin, 1989.

[36] L. R. Lucambio Pérez and L. F. Prudente, *Nonlinear conjugate gradient methods for vector optimization*, to appear in SIAM J. Optim.

[37] K. M. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, 1999.

[38] J. M. Martínez, *On high-order model regularization for constrained optimization*, SIAM J. Optim. 27 (2017), pp. 2447-2458.

[39] J. M. Martínez and M. Raydan, *Cubic-regularization counterpart of a variable-norm trust-region method for unconstrained minimization*, Journal of Global Optimization 68 (2017), pp. 367–385.

[40] Y. Nesterov and B. T. Polyak, *Cubic regularization of Newton's method and its global performance*, Mathematical Programming 108 (2006), pp. 177–205.

[41] A. Pascoletti and P. Serafini, *Scalarizing vector optimization problems*, J. Optim. Theory Appl. 42 (1984), pp. 499–524.

[42] T. Stewart, O. Bandte, H. Braun, N. Chakraborti, M. Ehrgott, M. Göbelt, Y. Jin, H. Nakayama, S. Poles, and D. Di Stefano, *Real-World Applications of Multiobjective Optimization*, in *Multiobjective Optimization*, J. Branke, K. Deb, K. Miettinen, and R.Sowiski R., eds., Lecture Notes in Computer Science, Vol. 5252. Springer, Berlin, Heidelberg, 2008.

[43] Y. Yamamoto, *Optimization over the efficient set: overview*, J. Glob. Optim.22 (2002), pp. 285–317.