A SYNTHETIC APPROACH TO MULTIOBJECTIVE OPTIMIZATION

ALBERTO LOVISON *

Abstract. We propose a strategy for approximating Pareto optimal sets based on the global analysis framework proposed by Smale (Dynamical systems, Academic Press, New York (1973) 531–544). We speak about synthetic approach because the optimal set is natively approximated by means of a compound geometrical object, i.e., a simplicial complex, rather than by an unstructured scatter of individual optima. The method distinguishes the hierarchy between singular set, Pareto critical set and stable Pareto critical set. Furthermore, a quadratic convergence result in set wise sense is proven and tested over numerical examples.

Key words. Multiobjective optimization, Pareto critical set, Delaunay tessellations in general dimension, stability of mappings, transversality

AMS subject classifications. 90C29, 58K25

1. Introduction.

1.1. Multiobjective optimization and Pareto optimality. Multiobjective optimization is concerned with the problem of optimizing several functions (or objectives) simultaneously. A precise mathematical statement in an economical framework was first given by the Italian mathematician Vilfredo Pareto [30, 31]. In his honor, optimal sets for multiobjective optimization are named after.

In the single objective case, optima are defined as the points \( x \in W \subseteq \mathbb{R}^d \) where a given function \( u : W \rightarrow \mathbb{R} \) assumes its maximum, if they exist.

In multiobjective optimization we consider two or more functions, \( u_1, \ldots, u_m : W \rightarrow \mathbb{R} \), and in all the non trivial cases the optima for one function are distinct from the optima of the remaining. More interestingly, there usually exist an infinite number of points which cannot be overcome according to all functionals at the same time.

In an olympic analogy, it is clear that defining a competition mixing different athletic specialties, let us say, e.g., running and swimming, will in general elect a winner which will be neither the running nor the swimming champion. Such an athlete will prevail over the running champion at swimming, and beat the swimming champion at running.

A rigorous definition can be stated:

**Definition 1 (Pareto optimality).** Let \( W \subseteq \mathbb{R}^d \) and \( u_1, \ldots, u_m : W \rightarrow \mathbb{R} \). A point \( \bar{x} \in W \) is called a Pareto optimum if there is no \( x \in W \) such that \( u_i(x) \geq u_i(\bar{x}) \) for all \( i = 1, \ldots, m \) and \( u_j(x) > u_j(\bar{x}) \) for some \( j \).

In this paper we will introduce an apparently novel numerical strategy for approximating Pareto optima, theoretically based on the global analysis framework established by Stephen Smale and others in the early seventies and reprised in some recent work. One of the basic facts highlighted in Smale’s framework, is that providing second order differentiability and some generic transversality condition, Pareto optimal sets are portions of \((m-1)\)-dimensional manifolds. According to the strategy we want to propose, Pareto optimal sets will be approximated by means of simplicial complexes, i.e., meshes. Furthermore, adopting the
Hausdorff measure as a distance between sets, it is possible to prove quadratic speed convergence, analogously to what happens for Newton methods in single objective optimization problems. We speak about synthetic approach because we think that its main novelty is that the approximation to the Pareto optimal set is natively represented as an aggregate geometrical object (the mesh), rather than an unstructured scatter of individual optima.

There exists a current in the recent multiobjective optimization literature concerning the topic of the even distribution of Pareto points [5, 24, 26, 63] while an alternative philosophy [36–38] deals with producing local meshes approximating Pareto sets, relying on continuation (homotopy) strategies. In the recent paper [32], both topics are addressed. More general techniques aiming to approximate the entire optimal set are described in the recent papers [12, 22], in the survey [99], and in the references therein. Finally, in [10], a Newton method for multiobjective optimization is described, and in some of the examples discussed a Delaunay tessellation of the Pareto optimal set is built. To the author’s knowledge, those are the most similar strategies available in literature to the approach proposed here. We highlight some possibly original aspects in our approach:

1. the representation of the optimal set as a mesh, as in the homotopy techniques, is extended to the whole input space and to a number of functions larger than two;
2. the entire Pareto optimal set can be filled with evenly spread points;
3. non convexity is treated natively;
4. the method relies on Delaunay tessellations, providing a description of the Pareto optimal set as a whole, improving visualization and problem description;
5. quadratic convergence in set wise sense is provided;
6. the method sheds light on fine geometrical and topological features of Pareto sets.

Possible drawbacks of the proposed synthetic method are the computational complexity related to Delaunay tessellations in high dimensional spaces and the intrinsic mathematical difficulty brought in by the theory of the singularities of mappings. These and other issues are discussed and addressed at the end of this paper.

2. The global analysis framework. In the early seventies, the Nobel prize for economy Gerard Debreu stimulated Stephen Smale to investigate the problem of optimizing several functions in the dynamical systems arena. In the series of works that followed emerged interesting topological and geometrical features of the sets of Pareto optima. The notion of Pareto critical set $\theta$, generalizing the concept of critical point for scalar functions, was introduced and furthermore, local Pareto optima were characterized by means of first and second derivatives. Quoting Smale [50]:

“We study the local and global nature of $\theta$, as one uses freshman calculus to study the maximum of a single function.”

We recall now Smale’s definitions and results. Let $W \subseteq \mathbb{R}^d$ be a smooth $n$–dimensional manifold, $u : W \rightarrow \mathbb{R}^m$ a smooth vector function, with $m \leq n$. Let $Pos$ be the open positive cone in $\mathbb{R}^m$, $Pos := \{ y \in \mathbb{R}^m \mid y_j > 0, \forall j = 1, \ldots, m \}$, and $C_x$ the corresponding open cone in the tangent space $T_xW$, $C_x := Du^{-1}(Pos) \subseteq T_xW$.

**Definition 2** (Pareto critical set $\theta$). The set

$$\theta := \left\{ x \in W \mid C_x = \emptyset \right\},$$

(2.1)
SYNTHETIC MULTI OBJECTIVE OPTIMIZATION

is called the Pareto critical set.

We characterize $\theta$ in terms of the Jacobian of $u$.

**Proposition 3** (First order proposition). Let $x \in W$. Then, $x \in \theta$ if and only if:

(a) $\{Du_j(x)\}_{j=1,\ldots,m}$ do not belong to a unique open half space of the cotangent space $T_x^* W$.

(b) $\exists \lambda_j \geq 0, j = 1, \ldots, m$, not all zero such that $\sum_j \lambda_j Du_j(x) = 0$.

In order to discriminate the nature of the Pareto critical points we introduce a notion of stability and point out its relation with the second derivatives of $u$.

**Definition 4.** A curve $(a,b) \ni t \mapsto \varphi(t) \in W$ is said to be admissible if

$$\frac{d}{dt} u_i(\varphi(t)) > 0, \quad t \in (a,b), \quad \forall i = 1, \ldots, m. \quad (2.2)$$

**Definition 5.** A Pareto critical point $x$ is said to be stable, $x \in \theta_S$, if, given a neighborhood $V_x$ of $x$ in $W$, there exists a neighborhood $U_x$ of $x$ in $V_x$, such that every admissible curve $\varphi : [a,b] \to W$, with $\varphi(a) \in U_x$ satisfies $\text{Image}(\varphi) \subset V_x$. It descends from the definition that Pareto optima are stable Pareto critical points, while conversely stable Pareto critical points are local Pareto optima.

Second derivatives are not defined invariantly, but if we think about them as a symmetric bilinear form restricted to the kernel of the differential $Du(x)$ assuming values on the cokernel $\mathbb{R}^m/\text{Image}(Du(x))$, then this form is invariantly defined. It is called “2nd intrinsic derivative” (see [23,35]).

The case of greatest importance is where corank $Du(x)$ is 1 (i.e., rank $Du(x)$ is $m-1$). In this case the second intrinsic derivative assumes values in a 1-dimensional vector space. If we consider $x \in \theta$, we have $\text{Image}(Du(x)) \cap \text{Pos} = \emptyset$, thus $\mathbb{R}^m/\text{Image}(Du(x))$ has a canonical positive ray. We call the 2nd intrinsic derivative, in this case, generalized Hessian $H_x$. It makes sense to say that $H_x$ is negative definite or positive definite, as well as to define an index, as the index of the symmetric form $H_x$. We set

$$\partial \theta = \left\{ x \in \theta \mid \text{Image}(Du(x)) \cap \{ \text{Cl(Pos)} \setminus \{0\} \} \neq \emptyset \right\} \quad (2.3)$$

where $\text{Cl(Pos)}$ is the closure of Pos.

**Proposition 6** (2nd order Proposition). Let $u : W \to \mathbb{R}^m$ a smooth map with $x \in \theta, x \not\in \partial \theta$ and corank $Du(x) = 1$. Then

(a) if the generalized Hessian $H_x$ is negative definite, then $x \in \theta_S$.

(b) Let $\lambda_j \geq 0, j = 1, \ldots, m$ be as in the 1st order proposition; then (up to a positive scalar)

$$H_x = \sum_{j=1}^m \lambda_j D^2 u_j(x), \quad \text{on ker } Du(x). \quad (2.4)$$

The proposition is proved in [53], while a discussion of the genericity of the hypotheses on the rank assumption is given in [50].

**2.1. The structure of Pareto sets.** We recall the notion of Thom’s stratification (see [59,62]).

**Definition 7.** Let $A \subset W$ be a closed subset. A stratification $S$ of $A$ is a finite collection of connected submanifolds of $W$ satisfying the following properties:

(1) $\cup_{S \in S} S = A.$
(2) If $S \in \mathcal{S}$ then $\partial S = \text{Cl}(S) \setminus S$ is a union of elements of $\mathcal{S}$ of lower dimension.

(3) If $S \in \mathcal{S}$ and $U$ is a submanifold of $W$ transversal to $S$ at $x \in S$ then $U$ is transversal to all elements of $\mathcal{S}$ in a neighborhood of $x$.

The following theorem has been proved in [6]. Consider the space $C^\infty(W, \mathbb{R}^m)$ endowed with the $C^\infty$ topology. $W$ is a compact manifold with dimension $n \geq m$.

**Theorem 8** ($\theta$ is a stratified set of dimension $m - 1$). There is an open and dense set $\mathcal{G} \subset C^\infty(W, \mathbb{R}^m)$ such that if $u \in \mathcal{G}$ then $\theta$ is a stratified set of dimension $m - 1$.

Remark 9. If $m > n$, it is possible to prove that, for a generic mapping $u$, $\theta$ is a stratified set of dimension $n$.

From the point of view of the numerical applications, we state that in the generic case the strata of the Pareto critical set $\theta$ can be discretized by means of a collection of $(m-1)$-dimensional meshes. Obviously we would like to refine this procedure to $\theta_S$. Unfortunately, the following conjecture has been proved only for $m = 2, 3$ (see [7, 64]).

**Conjecture 10.** There is an open and dense set $\mathcal{G} \subset C^\infty(W, \mathbb{R}^m)$ such that if $u \in \mathcal{G}$ then $\theta$ is a stratified set and $\theta_S$ is a union of strata.

Remark 11. The stable Pareto critical set $\theta_S$ is formed by all the local Pareto optimal points. The global Pareto optimal points cannot be distinguished from local optima by means of differential features as in the statements presented above. Global Pareto optima can only be filtered out a posteriori.

3. First order search algorithm. We have designed an algorithm (Algorithm 1) for approximating the Pareto critical set $\theta$ on the basis of Proposition 3. The procedure reminds contour plot algorithms for plotting level sets of functions. We start by considering a set of data points $D = \{P_1, \ldots, P_N\}$, then we build a Delaunay tessellation having $D$ as nodes.

**Algorithm 1** First order algorithm for approximating the Pareto critical set $\theta$

1. Consider a set of data points $D = \{P_1, \ldots, P_N\}$;
2. evaluate the gradients of the $u_j$ on the data points;
3. build a Delaunay tessellation on the nodes $D$;
4. for all Delaunay simplex $\Delta = \langle P_{i_1}, \ldots, P_{i_{n+1}} \rangle$ in the tessellation do
5. compute the $(m-1)$-polytope $\Sigma$ where the 1st order approximation of the Jacobian of $u$ vanishes;
6. extract the sub-polytope $\hat{\theta}$ where the vanishing linear combination $\lambda_1 Du_1 + \cdots + \lambda_m Du_m = 0$ has non negative coefficients;
7. end for
8. compose a simplicial complex glueing together adjacent polytopes $\hat{\theta}$.

3.1. Analysis of simplexes. We cycle through the tessellation simplexes $\Delta = \langle P_{i_1}, \ldots, P_{i_{n+1}} \rangle$ and approximate the portion of the Pareto critical set $\theta$ possibly contained in $\Delta$. To determine the linear approximation $\hat{\theta}_s$ of the stable Pareto critical portion $\theta_s \cap \Delta$ we recall that $\theta$ is contained in the singular set $\Sigma$, i.e., the set where

\[ \text{In the implementation considered in what follows we employed the qhull software [3], based on the computation of convex hulls, and, in the two dimensional examples, we employed the triangle software [47, 48]. For iterative schemes, a efficient alternative is offered by the Bowyer–Watson algorithm [4, 69], which is incremental.} \]
the rank of the differential $Du(x)$ is less than maximal:

$$\theta_s \subseteq \theta \subseteq \Sigma \subseteq W, \quad (\Rightarrow \hat{\theta}_s \subseteq \hat{\theta} \subseteq \hat{\Sigma} \subseteq \Delta) \quad (3.1)$$

Adjacent approximate portions $\hat{\theta}_s$ are eventually sewed together.

3.1.1. Singular set $\hat{\Sigma}$. We fix a cell $\Delta := \langle P_1, \ldots, P_{n+1} \rangle$. The Jacobian is an $n \times m$ matrix which rank is non maximal on the singular set $\Sigma$. The rank of $J_u$ drops when the rows are linearly dependent, e.g., when a suitable subset of the $m$–order minors are degenerate. We consider for instance the following submatrices:

$$M_1 = \begin{pmatrix}
\frac{\partial u_1}{\partial x_1} & \cdots & \frac{\partial u_1}{\partial x_m} \\
\vdots & \ddots & \vdots \\
\frac{\partial u_m}{\partial x_1} & \cdots & \frac{\partial u_m}{\partial x_m}
\end{pmatrix}, \quad M_2 = \begin{pmatrix}
\frac{\partial u_1}{\partial x_2} & \cdots & \frac{\partial u_1}{\partial x_{m+1}} \\
\vdots & \ddots & \vdots \\
\frac{\partial u_m}{\partial x_2} & \cdots & \frac{\partial u_m}{\partial x_{m+1}}
\end{pmatrix}, \quad \ldots, \quad M_{n-m+1} = \begin{pmatrix}
\frac{\partial u_1}{\partial x_{n-m+1}} & \cdots & \frac{\partial u_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial u_m}{\partial x_{n-m+1}} & \cdots & \frac{\partial u_m}{\partial x_n}
\end{pmatrix}. \quad (3.2)$$

We denote the number of minors by $r := n-m+1$ and set $\omega_j(x) := \det M_j(x)$, for $j = 1, \ldots, r$ and consider all the $(r+1)$–faces of the cell $\Delta$, i.e., for every $\{i_1, \ldots, i_{r+1}\} \subseteq \{1, \ldots, n+1\}$, with $i_1 < i_2 < \cdots < i_{r+1}$, we consider the simplex $\langle P_{i_1}, \ldots, P_{i_{r+1}} \rangle$. The solution $(\mu_1, \ldots, \mu_{r+1})$ of the system:

$$\begin{cases}
\mu_1 \omega_1(P_{i_1}) + \cdots + \mu_{r+1} \omega_1(P_{i_{r+1}}) = 0 \\
\vdots & \vdots \vdots \\
\mu_1 \omega_{r+1}(P_{i_1}) + \cdots + \mu_{r+1} \omega_{r+1}(P_{i_{r+1}}) = 0 \\
\mu_1 + \cdots + \mu_{r+1} = 1
\end{cases} \quad (3.3)$$

leads to a singular vertex $Q := \mu_1 P_{i_1} + \cdots + \mu_{r+1} P_{i_{r+1}}$ of $\hat{\Sigma}$ if all $\mu_j > 0$, i.e., if $Q$ is contained in the $(r+1)$–face of $\Delta$ considered.

The (possibly empty) singular set $\hat{\Sigma}$ is an $(m-1)$–polytope defined as the convex hull of the singular vertices $Q$.

3.1.2. Critical set $\hat{\theta}$. In the previous subsection we have detected the singular set $\Sigma$, on the basis of the fact that on the singular set the gradients are linearly dependent. On the other hand, on the critical set $\theta$ there exists a positive linear combination of the gradients giving zero. Thus we proceed by estimating the coefficients $\lambda_j$ of the vanishing linear convex combination of the gradients, and cutting out the critical set $\theta$ from $\Sigma$ by intersection with the half spaces where the linear interpolations of the $\lambda$s are positive.

More precisely, we solve the system:

$$\begin{cases}
\lambda_1 Du_1(P) + \cdots + \lambda_m Du_m(P) = 0, \\
\lambda_1 + \cdots + \lambda_m = 1
\end{cases} \quad (3.4)$$

for $\lambda_1, \ldots, \lambda_m$.

The Jacobian of $u$ has rank $m-1$ in almost all the points of the singular set (generic hypothesis), thus the system $[3.4]$ has rank $m$, and by the implicit function
Theorem $\lambda_j$s are smooth functions of $P$. As a result the level sets \( \{\lambda_j(P) = 0\} \), which define the boundary of $\theta$, are smooth manifolds. At the first order we are working with, the requests $\lambda_j(P) > 0$ cut out half spaces in $\hat{\Sigma}$, defining possibly a critical sub polytope $\hat{\theta}$ in $\Delta$.

We notice that we do not know the actual values of $Du$ on the singular vertices, i.e., the nodes of $\hat{\Sigma}$. Nevertheless, we can estimate them by linearly interpolating the values of $Du$ on the data nodes $P_1, \ldots, P_{r+1}$ defining the vertex $Q$ in $\hat{\Sigma}$. By taking the coefficients $\mu_1, \ldots, \mu_{r+1}$ solving the system (3.3), we are guaranteed that the $\hat{Du}_j(Q) := \mu_1 Du_j(P_1) + \cdots + \mu_{r+1} Du_j(P_{r+1})$ are linearly dependent, and we are legitimate to solve for the vanishing linear combination $\lambda_1 \hat{Du}_1(Q) + \cdots + \lambda_m \hat{Du}_m(Q) = 0$.

3.1.3. Convergence analysis towards $\theta$. Let us consider for this section a single simplex $\Delta$. Intuitively, it is clear that the approximation $\hat{\Sigma}$ of $\Sigma$ obtained by linear interpolation is quadratically good because of Taylor's theorem. We state more precisely this result in the set–wise context we have adopted. The distance between the sets $A$ and $B$ can be measured in terms of Hausdorff distance:

$$d_H(A, B) := \max \left( \sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{y \in B} \inf_{x \in A} d(x, y) \right).$$ (3.5)

**Theorem 12** (Quadratic precision for $\Sigma$). Let $P_0, \ldots, P_n$ be in general position, and such that $Du$ has maximum rank. We denote by $\Delta = \langle P_0, \ldots, P_n \rangle$ the $n$–simplex which vertices are those points. Let $\omega_1(x), \ldots, \omega_r(x)$ a selection of independent minors of $Du$, and $\hat{\omega}_j(x)$ be the 1st order interpolation of the values of $\omega_j$ on the nodes $P_i$. Assume $0$ is a regular value for $\omega_1, \ldots, \omega_r$, that the zero levels of the $\omega_j$ are transversal and that $\omega_j(P_i) \neq 0$, for all $i, j$. Then

$$\Sigma = \{\omega_1(x) = 0\} \cap \cdots \cap \{\omega_r(x) = 0\},$$ (3.6)

$$\hat{\Sigma} = \{\hat{\omega}_1(x) = 0\} \cap \cdots \cap \{\hat{\omega}_r(x) = 0\},$$ (3.7)

and there exists a constant $C$:

$$d_H(\Sigma, \hat{\Sigma}) \leq C \delta^2,$$ (3.8)

where $\delta > 0$ is the diameter of the simplex $\Delta$.

**Proof.** First of all, we notice that the $\omega_k(x)$ are polynomials of the first derivatives of $u$, thus are smooth in our hypotheses. Inductively, consider $r = 1$ and denote $\omega = \omega_1$. By Taylor's theorem,

$$\omega(x) = \hat{\omega}(x) + O \left( |x - P_0|^2 \right), \quad \text{i.e.,}$$

$$|\omega(x) - \hat{\omega}(x)| \leq C \delta^2,$$ (3.9)

for a suitable $C > 0$. Assume, without loss of generality, $\omega > 0$ on $P_0, \ldots, P_k$ and $\omega < 0$ on $P_{k+1}, \ldots, P_n$. Let $\varepsilon := C \delta^2$. (See panel (a) of Figure 3.1). Thus the zero levels of $\omega$ and $\hat{\omega}$ are comprised between the $\pm \varepsilon$ levels of $\hat{\omega}$, i.e.,

$$\left\{ x \in \Delta \left| \omega(x) = 0 \right. \right\} \subseteq \left\{ x \in \Delta \left| -\varepsilon \leq \hat{\omega}(x) \leq \varepsilon \right. \right\}.$$ (3.10)
By compactness of $\Delta$, here exist $x_0 \in \{\hat{\omega} = 0\}$, $x_\varepsilon \in \{\hat{\omega} = \varepsilon\}$, such that,

$$d_H (\{\hat{\omega} = 0\}, \{\hat{\omega} = \varepsilon\}) = |x_0 - x_\varepsilon|$$

(3.11)

and it holds that

$$\hat{\omega}(x_\varepsilon) - \hat{\omega}(x_0) = |\hat{\omega}'(x_0) \cdot (x_\varepsilon - x_0)| = \left| \frac{\partial \hat{\omega}}{\partial w}(x_0) \right| |x_\varepsilon - x_0|,$$

(3.12)

where $w = \frac{x_\varepsilon - x_0}{|x_\varepsilon - x_0|}$. By means of an elementary linear algebra argument we have also that

$$\left| \frac{\partial \omega}{\partial w}(x_0) \right| \geq \min_{i=1, \ldots, k} \left| \frac{\hat{\omega}(P_i) - \hat{\omega}(P_{i'})}{P_i - P_{i'}} \right| =: B > 0,$$

(3.13)

so we can conclude

$$|x_0 - x_\varepsilon| \leq \varepsilon \frac{B}{B} = C \varepsilon^2 = C \delta^2$$

(3.14)

and eventually

$$d_H (\{\omega = 0\}, \{\hat{\omega} = 0\}) \leq C \delta^2.$$

(3.15)

Consider now $r > 1$, and assume inductively that the Hausdorff distance between the intersection of the zero levels of $r - 1$ transversal functions and the intersection of the zero level of the respective linear interpolations on an $n$–simplex is quadratically smaller than the simplex diameter. Thus we have

$$\Sigma_- = \{\omega_1(x) = 0\} \cap \cdots \cap \{\omega_{r-1}(x) = 0\},$$

(3.16)

$$\hat{\Sigma}_- = \{\hat{\omega}_1(x) = 0\} \cap \cdots \cap \{\hat{\omega}_{r-1}(x) = 0\},$$

(3.17)

$$d_H (\Sigma_-, \hat{\Sigma}_-) \leq C \delta^2.$$

(3.18)
If we consider one more function $\omega_r(x)$ on the linear space $\hat{\Sigma}_-$, we are in the previous case, so there exists $A > 0$,

$$d_H \left( \hat{\Sigma}_- \cap \{ \omega_r(x) = 0 \}, \hat{\Sigma}_- \cap \{ \hat{\omega}_r(x) = 0 \} \right) \leq A \delta^2. \quad (3.19)$$

By transversality of the $\omega_1, \ldots, \omega_r$, the fact holding for the linear space $\hat{\Sigma}_-$ holds also for the compact manifold with boundary $\Sigma_-$ and the function $\omega_r$ (see Lemma 13 for the details). Thus there exists a $B > 0$ such that

$$d_H \left( \Sigma_- \cap \{ \omega_r(x) = 0 \}, \Sigma_- \cap \{ \hat{\omega}_r(x) = 0 \} \right) \leq B \delta^2. \quad (3.20)$$

On the other hand, for the intersection of the zero levels of the transversal functions $\omega_1, \ldots, \omega_{r-1}$ on the linear space $\{ \hat{\omega}_r(x) = 0 \}$, by the inductive hypothesis there exists $C > 0$ such that

$$d_H \left( \{ \omega_r(x) = 0 \} \cap \Sigma_-, \{ \omega_r(x) = 0 \} \cap \hat{\Sigma}_- \right) \leq C \delta^2, \quad (3.21)$$

so the thesis is proved by the triangle inequality.

**Lemma 13.** Let $\Sigma$ a manifold with boundary diffeomorphic to an $n$–simplex $\Delta$, and $\omega : \Sigma \to \mathbb{R}$ differentiable and without critical points inside $\Sigma$. We have $\omega(x) = \hat{\omega}(x) + O(\delta^2)$, where $\hat{\omega}$ is an affine approximation and $\delta$ is the simplex diameter. Thus we have that

$$d_H (\{ \omega(x) = c \} \cup \{ \hat{\omega}(x) = c \}) \leq C \delta^2, \quad \text{for all } c \in \mathbb{R}. \quad (3.22)$$

**Proof.** Let $\Delta \xrightarrow{\varphi} \Sigma$ be a diffeomorphism, with $\xi > |\varphi'| > \eta > 0$. Thus we have, for all $y \in \Delta$,

$$\omega \circ \varphi(y) = \hat{\omega} \circ \varphi(y) + O(\delta^2).$$

For any $y^*$ in the zero level of $\omega \circ \varphi$ we can find a line segment $[y_1, y_2]$, with $y_1$ being one of the nodes of $\Delta$ where $\omega \circ \varphi$ is negative and $y_2$ is a point on a face of $\Delta$ where on the forming nodes $\omega \circ \varphi$ is positive. By continuity there exists a point $\tilde{y}$ on the line $[y_1, y_2]$ where $\hat{\omega} \circ \varphi$ is zero. Thus

$$\omega \circ \varphi(y^*) - \omega \circ \varphi(\tilde{y}) = \hat{\omega} \circ \varphi(y^*) - \hat{\omega} \circ \varphi(\tilde{y}) + O(\delta^2) = \hat{\omega}' \circ \partial \varphi \frac{\partial \varphi}{\partial w} |y^* - \tilde{y}| + O(\delta^2), \quad (3.23)$$

which gives

$$|y^* - \tilde{y}| \leq C \delta^2. \quad (3.24)$$

**Note 14.** The hypotheses of Theorem 12 are generic in the sense that they hold for a open and dense set of functions in the $C^\infty$ topology. In particular, 0 is assumed to be a regular value for $\omega_1, \ldots, \omega_r$ because the set of the singular values has zero measure (Sard’s Theorem). See [1, 13, 21, 23, 27].

**Theorem 15.** In the simplex $\Delta = (P_0, \ldots, P_n)$, if $\theta$ is the Pareto critical set and $\hat{\theta}$ is its linear approximation, there exists $C > 0$ such that

$$d_H \left( \theta, \hat{\theta} \right) \leq C \delta^2. \quad (3.25)$$
Proof. The \( \lambda_j \)s computed as described in Algorithm 1 are first order approximations to smooth functions, apart from a measure zero set of points. Thus the conclusions of Theorem 12 apply as well to the intersection of \( \Sigma \) with the half spaces \( \lambda_j(P) \geq 0 \).

4. Second order algorithm. In Algorithm 2 we describe how to extract the stable critical set \( \theta_s \), i.e., the set of locally Pareto optimal points, from the critical set \( \theta \) determined in the first order algorithm.

\begin{algorithm}
\caption{Second order algorithm for the stable Pareto critical set \( \theta_s \)}
\begin{algorithmic}[1]
\State Consider a set of data points \( D = \{P_1, \ldots, P_N\} \) and proceed as in Algorithm 1.
\ForAll{Delaunay simplex \( \Delta = \langle P_{i_0}, \ldots, P_{i_n} \rangle \) in the tessellation}
\State Compute the matrix of the second derivatives \( D^2u \) on the nodes \( P_{i_0}, \ldots, P_{i_n} \).
\State On the vertices \( Q \) of \( \hat{\theta} \), linearly interpolate the second derivatives \( \hat{D}^2u(Q) \).
\State Compute a basis \( w_1, \ldots, w_{n-m+1} \) for \( \ker Du(Q) \), and set \( \hat{H}(Q) := w^\top \cdot \left( \lambda_1(Q)\hat{D}^2u_1(Q) + \cdots + \lambda_m(Q)\hat{D}^2u_m(Q) \right) \cdot w. \)
\State Compute the eigenvalues \( \sigma_1, \ldots, \sigma_{n-m+1} \) of \( \hat{H}(Q) \).
\State Cut out from \( \hat{\theta} \) the sub polytope \( \hat{\theta}_s \) where \( \sigma_k \leq 0 \) for all \( k = 1, \ldots, n-m+1 \).
\EndFor
\State Compose a simplicial complex glueing together adjacent polytopes \( \hat{\theta}_s \).
\end{algorithmic}
\end{algorithm}

The second derivatives could be also approximated computing the finite differences of the values of the gradients on the nodes of the \( n \)-simplex. Indeed, setting

\[ v_i = P_i - P_0, \quad i = 1, \ldots, n, \]

we have

\[ D^2u = \left( \frac{\partial^2 u}{\partial x_i \partial x_j} \right)_{i,j} \approx \sum_k \left( \frac{\partial^2 u}{\partial x_k \partial x_j} \right) \cdot \frac{\partial v_k}{\partial x_i} \approx \sum_k \left( \nabla u(P_k) - \nabla u(P_0) \right)_j \cdot (P_k - P_0)_i. \]  

(4.1)

Using this formula, the quadratic precision cannot be guaranteed for locating boundary points of the stable critical set. Furthermore, because the boundary faces belong to different simplexes, the estimated boundary points for \( \theta_s \) would jump from simplex to simplex.

On the other hand, the formula will be correct for discriminating the nature of inner stable critical points, without extra computations. Boundary simplexes can thus analyzed with second derivatives, allowing the computation of the boundary of \( \hat{\theta}_s \).

5. Applications.

5.1. Two dimensional examples. A series of examples in two dimensions is presented below. Functions gradients are evaluated on a grid of regular triangles and the critical set \( \theta \) is estimated according to the first order algorithm.

Boundary points are marked with black diamonds. The generalized Hessian is estimated on the nodes of the critical set, computing second derivatives in the triangles where its index changes, allowing to estimate the position of the points separating stable from unstable branches. These points are cusps (see [64]), and are marked by a
black star. Stable branches are colored in red, unstable branches in orange and finally non critical branches are gray.

Example 1. Consider two negative definite quadratic polynomials. The critical stable set is a line joining the two individual critical points. Other singular branches occur in outer regions of the domain.

$$u_1(x, y) = -1.05x^2 - 0.98y^2,$$
$$u_2(x, y) = -0.99(x - 3)^2 - 1.03(y - 2.5)^2.$$  \hfill (5.1)

See Figure 5.1.

Example 2. This example is taken from [53].

$$u_1(x, y) = -y,$$
$$u_2(x, y) = \frac{y - x^3}{x + 1}.$$  \hfill (5.2)

The critical set is a single curve split in a stable and an unstable branch, while the separating point is a cusp. See Figure 5.2.

Example 3. In the following mapping there are two second order polynomials, one negative definite and the other indefinite. The outcome is an (unbounded) global Pareto front and a local unbounded front terminating in a cusp.

$$u_1(x, y) = -x^2 - y^2,$$
$$u_2(x, y) = -(x - 0)^2 + (y + 0.3)^2.$$  \hfill (5.3)

See Figure 5.3.

Example 4. The following mapping is composed by a quadratic polynomial and a bimodal function. The resulting singular set is composed by an unbounded branch and two loops. One of the loops is critical and forms a local Pareto front delimited by two cusps, while the other loop is non critical.

$$u_1(x, y) = -x^2 - y^2 - 4(\exp(-(x + 2)^2 - y^2) + \exp(-(x - 2)^2 - y^2)),$$
$$u_2(x, y) = -(x - 0)^2 - (y + 0.5)^2.$$  \hfill (5.4)
5.2. Three functions in three dimensions examples. Example 5. The simplest nontrivial non degenerate example we can build in the three dimensional case is composed by three negative definite 2nd order polynomial functions $f_j(x), j = 1, 2, 3$. 

See Figure [5.4]
Additionally, we introduce a small non polynomial perturbation.

\[
f_j(x) = (x - C_j)^\top \begin{pmatrix} -\alpha_{j,1} & 0 & 0 \\ 0 & -\alpha_{j,2} & 0 \\ 0 & 0 & -\alpha_{j,3} \end{pmatrix} (x - C_j), \quad j = 1, 2, 3,
\]

\[
\begin{pmatrix} u_1(x) \\ u_2(x) \\ u_3(x) \end{pmatrix} := \begin{pmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_2 \sin \left( \frac{\pi}{\gamma_2} (x + y) \right) \\ \beta_3 \cos \left( \frac{\pi}{\gamma_3} (x - y) \right) \end{pmatrix}.
\]

(5.5)

Where \(x = (x_1, x_2, x_3)^\top \in \mathbb{R}^3\), \(\alpha_{j,i} > 0\), \(i, j = 1, 2, 3\), \(C_1, C_2, C_3 \in \mathbb{R}^3\) are distinct, non collinear points, while \(\beta_j, \gamma_j\) are real numbers. In the generic case the singular set is an hypersurface of \(\mathbb{R}^3\), while the critical set \(\theta\), which is stable, is diffeomorphic to a triangle, i.e., \(\theta\) is a compact connected manifold with boundary and three corners, corresponding to the minima of the three functions \(u_1, u_2, u_3\). See Figure 5.5(a).

Example 6. We break the convexity of the previous example by adding a secondary maximum to the first function. We define a further negative definite, 2nd order polynomial \(f_4(x)\) and set \(u(x)\) as:

\[
\begin{pmatrix} u_1(x) \\ u_2(x) \\ u_3(x) \end{pmatrix} := \begin{pmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{pmatrix} + \begin{pmatrix} \beta_1 \exp \left( \frac{1}{\gamma_1} f_4(x) \right) \\ \beta_2 \sin \left( \frac{\pi}{\gamma_2} (x + y) \right) \\ \beta_3 \cos \left( \frac{\pi}{\gamma_3} (x - y) \right) \end{pmatrix}.
\]

(5.6)

The main portion of the Pareto set is slightly deformed while a new branch appears. In Figure 5.5(b) is shown the resulting Pareto critical set \(\theta\) obtained by iterative application of Algorithm 1 as described in the following Section.

6. Iterative schemes. The previously presented approach defines an approximation of the Pareto optimal set given any distribution of points in the domain. Here
we propose and discuss an iterative scheme. At every step a selection of points from the approximated Pareto optimal set is added to the dataset $\mathcal{D}$, the gradients in the new points are evaluated, the tessellation is updated and a refined approximation of the Pareto set is built. The desired effect is obviously to get closer and closer to the actual optimal set, but an efficient strategy should produce an as uniform as possible discretization of the optimal set. A naif approach would suggest to insert in the set of the candidates for evaluation all of the nodes of the complexes, i.e., all the stable admissible vertices computed and all of the boundary points, both for criticality and stability. Nevertheless, a glance at the examples of the previous section reveals that the sizes of the optimal complexes cover a wide distribution, in particular the patches result very small if $\Sigma$ passes close to tessellation nodes. Moreover, little experience shows that large patches are reduced sensibly slowly if none of their internal points is introduced. With this criteria in mind we introduce an iterative scheme for the case of two functions.

6.1. Two functions iterative scheme. In the two functions case the Pareto optimal set is a one dimensional manifold with boundary, i.e., a collection of curved intervals. The discrete approximation is a collection of polygonal curves. For every interval a sequence of candidate points equally spaced along the polygonal curve is extracted. The number of points is chosen equal to the number of segments, so that approximately every triangle containing optimal points is split as close as possible to the optimal set.

6.2. Higher number of functions. It seems reasonable to take into account of the stratified structure of $\theta$ in the design of an iterative strategy. In fact, strata should be filled as uniformly as possible, where the uniformity is determined according to the $k$-dimensional measure, if $k$ is the dimension of the stratum. So, taking for instance the situation of example 5, corners’ approximations are re-evaluated at each iteration, uniformly spaced points are taken along boundary lines, exactly as in the two functions case, while internal points should be distributed proportionally to the

\[2\] The question of producing evenly distributed Pareto fronts has been pursued by I. Das and J. Dennis in \cite{5} and is widely discussed in \cite{24, 26, 63} and in the references therein. The present approach focuses on the evenly sampled Pareto sets in the input space, rather than in the output space.
area of the triangles and polygons composing \( \hat{\theta} \). This is more difficult to be defined precisely. Indeed, the problem of uniformly filling a general \( n \)-dimensional region is a long–time crucial issue for statistical applications \cite{41}. Furthermore, in our problem we have to fill uniformly a general \( n \)-dimensional manifold, thus we have somehow to take into account of the effects of the curvature on the measure of the volumes.

Taking inspiration from a Design of Experiments strategy called maximin distance design \cite{14} we proceed as described in Algorithm 3.

**Algorithm 3** Uniform filling a simplicial complex

1: Tessellate in simplexes the polytopes of the mesh
2: Build the adjacency lists of the simplexes
3: Evaluate the volume of each simplex
4: For every simplex define the accumulated volume as the sum of its volume and the volume of the adjacent simplexes
5: Pick the simplex with the maximum accumulated volume
6: Add to the candidates stack the center of mass of this maximal simplex
7: repeat
8: Recompute the accumulated volumes excluding the already picked simplexes
9: until the desired number of candidate points is collected

6.3. Stopping criteria. Analogously to gradient based methods of single function optimization (nonlinear conjugate gradient, Newton and Newton–like methods), a stopping criterion could be based on the magnitude of the minors \( M_1, \ldots, M_r \) computed in the points of the last iteration. The magnitude of the minors is analogous to the magnitude of the gradients for single objective optimization.

In fact, we could define a different iterative strategy taking the rule of subdividing only stable critical triangles contained in simplexes where the minors are larger than a prescribed threshold.

6.4. Application. We show the behavior of the iterative scheme described above applied to the mapping in Example 4. At each iteration we generate a number of evenly spaced points along the approximate stable Pareto critical set. In order to exhibit the claimed quadratic convergence, it is necessary to sample the approximated optimal set by quadratically finer intervals, i.e., comparable to the precision gained. As a result the density of points will grow exponentially w.r.t. the number of iterations.

Such a density of points rapidly deteriorates the mesh quality, i.e., skinny triangles suddenly appear leading to numerical instability. Thus, at each iteration, a number of extra nodes (namely, the circumcenters of the most skinny triangles) should be introduced in the mesh in order to produce a nicely grading. At this extent we have coupled our method with Ruppert’s algorithm, as implemented in the triangular mesh refinement software TRIANGLE by J. R. Shewchuk \cite{47,48}.

Already at the fifth iteration the triangulation starts to suffer from numerical instability, thus we consider \( \hat{\theta}_S^{(4)} \) generated at the fourth iteration as the optimum and evaluate the Hausdorff distances between \( \hat{\theta}_S^{(i)} \) and \( \hat{\theta}_S^{(4)} \), for \( i = 1, \ldots, 3 \). As in can be seen in Figure 6.2, panel (a), the Hausdorff distances between the approximated Pareto sets and the numerical optimum converges superlinearly. For reference also the convergence behavior of the maximum and the mean minors magnitude are reported.
In Figure 6.1 is illustrated how the triangulation and the representation of the Pareto set evolves from one iteration to the subsequent.

In Figure 6.3, the three dimensional problem of Example 5 is tackled by the procedure described in Algorithm 3. The algorithm has been applied by introducing only a small number (~10) of new points on the sites with the largest magnitude of the minors. In such a way it was possible to iterate 70 times the scheme reaching a very small magnitude for the minors.

Because of the mentioned exponentially growing number of samples necessary to exhibit quadratic convergence speed for the iterative scheme, the experiment described for the two dimensional case becomes prohibitive in three dimensions.

On the other hand, the quadratic precision can be verified as well by means of progressively finer regular meshes. The distance between the approximated Pareto set and the numerical optimum should reduce as the square of the mesh size, or, in other words, their ratio should asymptotically converge towards a constant. Indeed, this is precisely the behavior observed in the experiment, as reported in Figure 6.2 panel (b).

7. Conclusions and perspectives. In this paper we have proposed an approximation scheme for representing in a synthetic fashion the solution sets of multiobjective optimization problems. The emphasis is posed on the unified and structured representation of the solution set, which seems not yet tackled in the previous literature, at least to the knowledge of the author. Furthermore, a quadratic convergence result in a set wise sense is proven and tested numerically. The applicability and possible extensions of the algorithms described in this paper are conditioned by the issues enumerated below.

7.1. The curse of dimensionality. The first problem one encounters when trying to apply this algorithms to industrial strength problems is the limitations to the input dimension. The whole procedure is based on a Delaunay tessellation of the input domain, which complexity grows exponentially with dimension. As pointed out for instance in the qhull documentation [3], building the convex hull of a 9-hypercube is computationally exhaustive. Analogous limitations are encountered in global optimization, where the search for optima in high dimensional domains cannot realistically be performed on real case problems. Indeed, typically, global search algorithms are rarely tested and compared over dimensions larger than 5 (see [11, 15, 17, 20, 33, 34, 43, 46, 58, 71]. This problem is structural and cannot be resolved by augmenting the computational resources. Therefore, the presented algorithms are best suited for low dimensional problems.

In fact, the curse of dimensionality is a strong motivation for reflecting carefully on the necessity of introducing extra input variables when tackling new problems and designing experiments.

As possible exit strategy could be screening the input variables [42, 55, 57]. This practice can be surprisingly successful, because usually sparsity of effects occurs, revealing a pronounced hierarchy among input variables, leading to sensible simplification of the problem formulation.

3The sparsity of effects is an empirical law stating that in a generic physical experiment one usually observes that the 80% of the effects are due to the 20% of the factors. Related phenomena are that the first order contributions are the most important, while higher order contributions decay
Fig. 6.1. Iterative scheme for the mapping in Example 4.
7.2. Structural instability. Another serious issue coming from singularity theory is the possible occurrence of structurally unstable singularities of mappings. As claimed by J. Mather [23] and reported by Arnol’d [1, 2], a mapping \( u : \mathbb{R}^n \rightarrow \mathbb{R}^m \), with \( m \geq 7 \) can present structurally unstable singularities. In the lower dimensional cases, the stability results support the convergence of numerical techniques to analytical solutions. In particular, all of the mappings in a sufficiently small neighborhood of a stable map have diffeomorphic level sets. As a result, incremental approximation to the singular set \( \Sigma \) are diffeomorphic, if the domain is sampled enough densely. If the number of functions is less than 7, almost all mappings are structurally stable, thus one can reasonably expect that real world functions are stable. This cannot be guaranteed for higher dimensions, where the set of structurally unstable mappings has non zero measure.

7.3. Future work. In industrial applications, when the objective functions at hand are computationally expensive and the derivatives are not easily computed, we figure that the applicability of the algorithm proposed here will be significantly extended having resort to surrogate models. There exists an extensive literature developed in recent year on this subject (see [15, 16, 40, 41] and the references therein), also with specific applications to multiobjective optimization [18, 19].

The procedures of this paper can be adapted applying the Algorithms 1 and 2.
to a surrogate model $\tilde{u}$ fitted to the values of the true functions $u$ computed on the given data points. On the outgoing candidate points, new evaluations of $u$ are to be computed, and a new surrogate model fitted to the increased dataset. This reduces the computational effort for computing derivatives and furthermore prevents premature stopping of the optimization process due to accidental failure of function evaluation at some data point.

Moreover, globalized version of the iterative strategies, guaranteeing that all possible branches of Pareto sets are discovered, and the handling of constraints, will be the subject of future research.

REFERENCES
SYNTHETIC MULTI OBJECTIVE OPTIMIZATION

[26] E. Miglierina and E. Molho, Convergence of minimal sets in convex vector optimization,
[30] V. Pareto, Cours d'économie politique/Professe à l'Université de Lausanne, Rouge, Lausanne, 1896-1897.
[31] -----, Manuale di economia politica con una introduzione alla scienza sociale, no. 13 in Piccola biblioteca scientifica, Società editrice libraria, Milan, 1906.
[53] I. M. Sobol’, Global sensitivity indices for nonlinear mathematical models and their monte


