Journal of Global Optimization manuscript No.
(will be inserted by the editor)

Global Search Perspectives for Multiobjective Optimization

Alberto Lovison

Received: date / Accepted: date

Abstract Extending the notion of global search to multiobjective optimization is far than straightforward, mainly for the reason that one almost always has to deal with infinite Pareto optima and correspondingly infinite optimal values. Adopting Stephen Smale's global analysis framework, we highlight the geometrical features of the set of Pareto optima and we are led to consistent notions of global convergence. We formulate then a multiobjective version of a celebrated result by Stephens and Baritompa, about the necessity of generating everywhere dense sample sequences, and describe a globally convergent algorithm in case the Lipschitz constant of the determinant of the Jacobian is known.

Keywords Multiobjective optimization · global optimization · nonconvexity and multiextremality · stability of mappings · continuation methods

Mathematics Subject Classification (2000) 90C26 · 90C29 · 58K05

1 Introduction

In a celebrated paper [35] Stephens and Baritompa formulate a very general result about the necessity of sampling everywhere densely the search domain when seeking the global optimum of a function, unless some global information about the function is available.

At the present day, at least in the author's knowledge, the analogous question has not yet been completely solved for the case of multiobjective optimization (MO).

A. Lovison
Dipartimento di Matematica – Università degli Studi di Padova
via Trieste, 63 – 35121 Padova ITALY
Tel.: +39-049-8271317
Fax: +39-049-8271499
E-mail: lovison@math.unipd.it

Nonetheless, there exists a plenty of literature dealing with general nonlinear multiobjective optimization, witnessing the interest in the scientific community for possible improvements in this direction (see, among others, [7, 11, 16, 22]).

Along this line of thought, we propose some extensions of the notions of seeing and localizing the global optimum for the MO case. As a consequence, we reformulate and prove the theorem of Stephens and Baritompa for multiple objectives and moreover we devise a multiobjective version of the Pijavskij–Shubert algorithm [25, 31], exhibiting global convergence when the Lipschitz constant of the determinant of the Jacobian is available. This algorithm is also an attempt to formulate a globally convergent version of the singular continuation strategy [18].

The results presented here are subject to some smoothness and non degeneracy conditions on the functions at hand. More precisely, the functions considered will be *good* functions in the sense of Whitney [40]. These conditions are a natural extension to the conditions implicitly required for scalar functions in order to have isolated critical points and make gradient based algorithms work properly. Nonetheless, these conditions are fulfilled for *almost all* functions, i.e., for an open and everywhere dense subset of C^∞, as illustrated in detail in Theorems 3 and 4. In perspective, we hope that the ideas presented in this work could be of inspiration for devising further MO optimization strategies with global character.

2 Notations, definitions and basic results

We recall some basic notations from differential geometry. Our basic reference is [8].

- Let $W \subseteq \mathbb{R}^\ell$ an n–dimensional compact manifold,
- $f : W \longrightarrow \mathbb{R}^m$ be a smooth function, $f(x) = (f_1(x), \ldots, f_m(x))$, $m \geqslant 1$.
- Let T_xW denote the tangent space to W in $x \in W$, and $T_x^\star W$ the cotangent space. $Df(x) : T_xW \to \mathbb{R}^m$ denote the tangent map to f in $x \in W$. In a system of local coordinates $(x_1, \ldots, x_n)$ in W,
$$\forall v \in T_xW, \qquad Df(x)v = \left(\sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x) v_j \right)_{i=1,\ldots,m}$$
- $Df_i(x) \in T_x^\star W$ denote the tangent maps of the components of the function f. In local coordinates Df_i coincide with the i–th row of of the Jacobian matrix $\left(\frac{\partial f_i}{\partial x_j}\right)_{i,j}$.

2.1 Global and local optima for a single objective

For the single objective case $m = 1$, i.e., standard single objective optimization problems,

- let us denote by $f^\star$ the absolute, or *global*, maximum value of the function f, $x^\star$ being a point in W realizing the maximum, in other words, $x^\star$ is an *optimum*, while $X_f^\star$ is the set of all optima $X_f^\star := \left\{ x^\star \,\middle|\, f(x^\star) = f^\star \right\}$.

- A local optimum is a point $x^\star_{loc} \in W$ such that there exists a neighborhood $U \subseteq W$, $x^\star_{loc} \in U$, such that $f(x) \leqslant f(x^\star_{loc})$, for every $x \in U$. Let also

$$X^\star_{f,loc} := \left\{ x^\star \in W \,\middle|\, x^\star \text{ is a local optimum for } f \right\}.$$

2.2 Optimality for multiple objectives

For multiple objectives, $m \geqslant 2$, we adopt the global analysis framework of Smale et al. [20,21,23,24,33,38,39].

- $x^\star \in W$ is *non dominated* if there is no $x \in W$ such that $f_i(x) \geqslant f_i(x^\star)$ for all $i = 1,\dots,m$ and $f_j(x) > f_j(x^\star)$ for some j. $x^\star$ is also called a *(global) Pareto optimum*.
- $x^\star \in W$ is a *local Pareto optimum* if there exists a neighborhood $V \ni x^\star$, where $x^\star$ is non dominated.
- We denote by θ_{op} the set of global Pareto optima and by $\theta_{loc} \supseteq \theta_{op}$ the set of local Pareto optima.
- A curve $\mathbb{R} \supseteq (a,b) \ni t \mapsto \gamma(t) \in W$ is said *admissible* if $\frac{d}{dt}(f_i \circ \gamma)(t) > 0$, $\forall i = 1,\dots,m$, $\forall t \in (a,b)$.
- A point $x \in W$ is said *Pareto critical* if there does not exist any admissible curve passing through it. The set of Pareto critical points is denoted by θ.
- If any admissible curve starting near a critical point x remains near x, x is said a *stable* Pareto critical point. The stable Pareto critical set is denoted by θ_s.
- Criticality is a *necessary* condition for $x \in W$ to be a local Pareto optimum, while stable criticality is a *sufficient* condition [39], i.e., $\theta \supseteq \theta_{loc} \supseteq \theta_s$.

We recall the characterization of critical and stable critical points in terms of first and second derivatives according to Smale [33,34].[1]

PROPOSITION 1 (FIRST ORDER PROPOSITION) *Let $x \in W$. Then $x \in \theta$ if and only if one of the following equivalent hypotheses hold:*

(a) Let $Pos = \left\{ y \in \mathbb{R}^m \,\middle|\, y_j > 0, \forall j = 1,\dots,m \right\}$, *then* $(Df(x))^{-1}(Pos) = \emptyset$.
(b) $Df_i(x)$ *do not all belong to some open half–space in the cotangent space* $T^\star_x W$.
(c) $\exists \lambda_i \geqslant 0$, *not all zero, with* $\sum_i \lambda_i Df_i(x) = 0$.

We recall some useful definitions for dealing with second order conditions.

- $\partial\theta = \left\{ x \in \theta \,\middle|\, ImDf(x) \cap \overline{Pos} \setminus \{0\} \neq \emptyset \right\}$, where $\overline{Pos}$ is the closure of Pos in $\mathbb{R}^\ell$.
- The second derivative is not defined invariantly, although it becomes invariant when restricted to the kernel of $Df(x)$ and its values are taken on the cokernel, i.e., $\mathrm{coker}(Df(x)) := \mathbb{R}^m / ImDf(x)$. We call this second intrinsic derivative the *generalized Hessian* $H_f(x)$:

$$H_f(x) : \ker Df(x) \longrightarrow \mathrm{coker}(Df(x)).$$

[1] *"We study the local and global nature of θ, as one uses freshmen calculus to study the maximum of a single function."*

The following proposition extends the non degenerate Hessian theorem.

PROPOSITION 2 (SECOND ORDER PROPOSITION) *Let* $x \in \theta \setminus \partial\theta$ *and* $\operatorname{corank} Df(x) = 1$ *(*rank assumption*). Then*

(a) if the generalized Hessian $H_f(x)$ *is negative definite, then* $x \in \theta_s$*.*
(b) Let $\lambda_i \geqslant 0$ *as in the first order proposition, then up to a positive scalar:*

$$H_f(x) = \sum_{i=1}^{m} \lambda_i D^2 f_i(x), \qquad (\textit{on } \ker Df(x)).$$

2.3 Structural stability of Pareto sets

We conclude with some observations about differential and topological aspects of the critical set. These remarks appear of primal importance for multiobjective optimization, because they introduce and prove that the Pareto set, under some generic condition, has a precise geometrical structure: it is part of a manifold, and therefore it can be parametrized. Moreover they prove the *structural stability* of this structure: small perturbation of the functions, or also the fact that only approximate values for the functions are known, will not alter dramatically the Pareto set [32].

– The critical set θ is a subset of the singular set $\Sigma := \left\{ x \in W \,\middle|\, \ker Df(x) \neq \{0\} \right\}$.
– Depending on the degeneracy order of f, Σ could be a smooth submanifold of W with any dimension $r \leqslant n$, or even Σ could not be a manifold at all. Nevertheless, for $n \geqslant 2m-4$, Thom's transversality theorem holds [5,17,37] and there is an open and dense set $G_0 \subseteq C^\infty(W, \mathbb{R}^m)$, whose functions f are *good*, i.e., there exists a set $n-m+1$ independent equations defining Σ_f, which is therefore a smooth submanifold of W with dimension $m-1$.
– Two submanifolds U and V in W are said *transversal*, $U \pitchfork V$ if

$$\forall x \in U \cap V, \qquad T_x U \oplus T_x V := \left\{ u+v \,\middle|\, (u,v) \in T_x U \times T_x V \right\} = T_x W.$$

– Let A be a closed subset of W. A *stratification* $\mathscr{S}$ of A is a finite collection of connected submanifolds of W such that
  1. $\cup_{S \in \mathscr{S}} S = A$
  2. If $S \in \mathscr{S}$, then $\partial S = \overline{S} \setminus S$ is a union of elements of $\mathscr{S}$ of lower dimension.
  3. If $S \in \mathscr{S}$ and U is a submanifold of W transversal to S in x, then U is transversal to all elements of $\mathscr{S}$ in a neighborhood of x.

  The elements of $\mathscr{S}$ are called a *strata* and A is said a *stratified set*.

If we adopt the C^∞ topology on the space $C^\infty(W, \mathbb{R}^m)$ and assume that $m \leqslant n$, we have

THEOREM 3 (DE MELO [20]) *There is an open and dense subset* $\mathscr{G} \subset C^\infty(W, \mathbb{R}^m)$ *such that if* $f \in \mathscr{G}$ *then* θ *is a stratified set of dimension* $m-1$*.*

THEOREM 4 (DE MELO [21]) *If* $m \leqslant 3$*, there exists an open and dense set of* $C^\infty(W, \mathbb{R}^m)$ *whose mappings are* θ*–stable.*

2.4 A pictorial example

Consider the following vector function:

$$\begin{cases} f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \\ f_1(x,y) := -(x-6)^2 - (y+0.5)^2, \\ f_2(x,y) := -x^2 - y^2 - 4\left(e^{-(x+2)^2-y^2} + e^{-(x-2)^2-y^2}\right). \end{cases} \tag{1}$$

The singular set is the solution set of the equation:

$$\Sigma = \left\{ (x,y) \in \mathbb{R}^2 \middle| \det \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} = 0 \right\},$$

while the critical set is the subset of Σ where the gradients Df_1 and Df_2 are opposed to each other:

$$\theta = \left\{ (x,y) \in \Sigma \middle| Df_1(x,y) \cdot Df_2(x,y) \leqslant 0 \right\}.$$

The local Pareto optima coincides here with the stable Pareto critical set, which is the subset of θ where the generalized Hessian is negative definite:

$$\theta_s = \left\{ (x,y) \in \theta \middle| \left\langle \lambda_1 D^2 f_1(x,y) + \lambda_2 D^2 f_2(x,y) v, v \right\rangle \leqslant 0, \quad \forall v \in \ker J_f \right\}.$$

In Figure 1 is illustrated a piecewise linear approximation to the singular set (the union of gray, orange and red curves) the critical set (orange and red) and the stable critical set (red).

2.5 Convergence of optimization algorithms.

Before introducing deterministic sequential algorithms we need to introduce the concept of *local information*. Let $\mathscr{X}_{finite}$ the set of all finite sequences of points in W.

DEFINITION 5 Local information *for a family $\mathscr{F}$ is a function LI defined on $\mathscr{F} \times \mathscr{X}_{finite}$ satisfying, $\forall f,g \in \mathscr{F}$, $\forall N \subseteq W$, open in W containing X, if $f\big|_N = g\big|_N$ then $LI(f,X) = LI(g,X)$.*

The range of a local information is left intentionally unspecified:

> *[...] because it could be of very different nature. Local information include any function depending on function values or any limiting information at a finite number of sample points. Limiting information include for instance partial or total derivatives, or even the fractal limiting dimension at a point. Any formula depending on such examples is itself local information, as for instance the maximum sample point, the maximum slope between sample points or the interpolating polynomial through sample points are local information[35].*

– An *algorithm* is a finite sequence of well–defined instructions, which, when running on a function f, produces the sample sequence $X_f := \{x_1, \ldots, x_k, \ldots\} \subseteq W$.

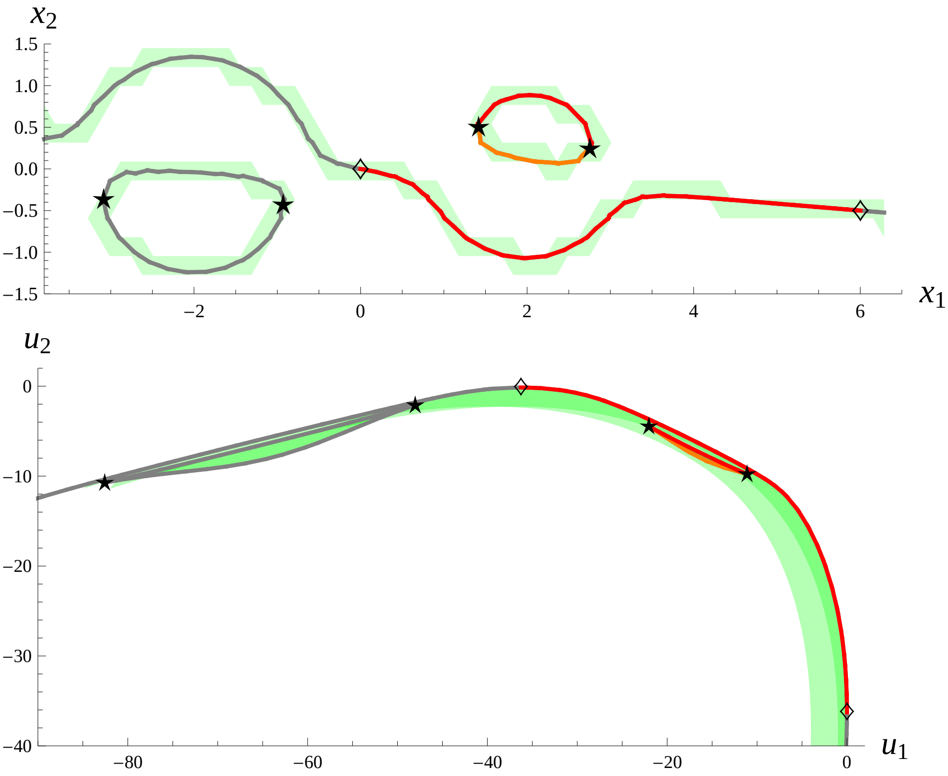

Fig. 1 Upper panel: singular set Σ (gray, orange and red curves), critical set θ (orange and red) and stable critical set θ_s (red) for the function in example 1. Lower panel: image of the function f with the images of the relevant sets Σ, θ and θ_s highlighted.

- A *deterministic sequential algorithm* on a class of functions $\mathscr{F}$ is an algorithm for which there is a local information function LI such that, for all $f \in \mathscr{F}$, when running on f, x_{k+1} depends only on $LI(f, X_k)$.
- We denote by X_f the full infinite sequence produced by an algorithm when given a function f, by $X_{f,k}$ or X_k the partial k–sequence. $\overline{X_f}$ is the closure of X_f while $X'_f = \overline{X_f} \setminus X_f$, is the set of limit points of X_f.
- An algorithm *sees* the global minimum of the function f if $X^\star_f \cap X'_f \neq \emptyset$.
- An algorithm *localizes* the global minimum if $X'_f = X^\star_f$ (or *weakly localizes* if $X'_f \subseteq X^\star_f$).

It seems useful to give further precise description of a class of algorithms for detecting structured subsets rather than scatters of points.

- A *set–wise sequential algorithm* is a deterministic algorithm which, besides the sample sequence $X_f = \{x_1, x_2, \dots\}$, generates a sequence of subsets $\{S_1, \dots, S_k, \dots\}$, $S_k \subseteq W$, intended to give an approximation of the Pareto set.

– More or less explicitly, any multiobjective optimization strategy is a set wise sequential algorithm. The sequence of sets approximating the Pareto set is given by the non dominated sets of the partial sequences:

$$S_k := nd(\{x1,\ldots,x_k\}).$$

– Simplicial methods like [9,18] at each new iteration produce a simplicial complex as approximation of the Pareto optimal set. Such geometrical objects are piecewise linear surrogates for smooth manifolds with corners, such as the stratified sets in the generic conditions discussed in Section 2.2.

2.6 Convergence in multiobjective optimization

To be convergent, an algorithm should produce a sequence $S_1, S_2, \ldots$, converging in some sense to the set of optima θ_{op}. Expressing a crude translation of the concepts of *seeing* and *localizing* the optima is poorly useful, because apart from degenerate cases, the set of Pareto optimal values does not consist in a single (vector) value $f^\star \in \mathbb{R}^m$. In the generic case, the set of Pareto optimal values is infinite, as well as, of course, the set of Pareto optima θ_{op}. Limits have to be considered in a set–wise sense, and therefore we need a concept of distance between sets.

– Let $A, B \subseteq W$. The *Hausdorff distance* between A and B is defined as

$$d_{\mathscr{H}}(A,B) := \max\left\{\max_{x\in A}\min_{y\in B} d(x,y), \max_{y\in B}\min_{x\in A} d(x,y)\right\}. \tag{2}$$

– We say that algorithm $\mathscr{A}$ *sees* the global Pareto optima θ_{op} if

$$\lim_{k\to\infty}\min_{x\in S_k, y\in\theta_{op}} d(x,y) = 0. \qquad (\mathscr{A} \text{ sees } \theta_{op}) \tag{3}$$

(In a sense the limit set $\lim_k S_k \cap \theta_{op} \neq \emptyset$, i.e., at least the Pareto set generated by the algorithm touches a portion of the global Pareto set, i.e., it generalizes the statement $X'_f \cap X^\star_f \neq \emptyset$.)

– We say that $\mathscr{A}$ *weakly localizes* the global Pareto optima θ_{op} if

$$\lim_{k\to\infty}\max_{t\in\theta_{op}} d(t, S_k) = 0. \qquad (\mathscr{A} \text{ weakly localizes } \theta_{op}) \tag{4}$$

(In a sense the limit set will contain all portions of the global Pareto set $\theta_{op} \subseteq \lim_k S_k$. The limit set is possibly larger than the Pareto set.)

– We say that $\mathscr{A}$ *strictly localizes* the global Pareto optima θ_{op} if

$$\lim_{k\to\infty} d_{\mathscr{H}}(S_k, \theta_{op}) = 0. \qquad (\mathscr{A} \text{ strictly localizes } \theta_{op}) \tag{5}$$

The Pareto set generated by the algorithm *coincides* with the true Pareto set. Saying that $\mathscr{A}$ sees, weakly localizes or strictly localizes the local Pareto optima θ_{loc}, the stable Pareto critical set θ_s, the critical set θ or the singular set Σ has an obvious meaning.

– Dealing with algorithms which merely see the global optimum, or that localize non strictly the set of Pareto optima seems not completely satisfactory from the global multiobjective optimization point of view. For instance, an algorithm optimizing only to one component of the vector function f would give a non dominated point, and it would *see* the Pareto optimum.

3 Global optimization requires global information

The last notions we need in order to deal with convergence issues are the families of functions *sufficiently rich*, i.e., functions that can be deformed as desired, while maintaining smoothness requirements.

DEFINITION 6 *A non empty class of (vector) functions $\mathscr{F} \subseteq C(W, \mathbb{R}^m)$ is* sufficiently rich, *if $\forall y \in \mathbb{R}^m$, $x \in W$, $f \in \mathscr{F}$ and N open in W containing x, there exists $g \in \mathscr{F}$ such that $f\Big|_{W\setminus N} = g\Big|_{W\setminus N}$ and $g(x) = y$.*

Standard classes of functions which are sufficiently rich are for instance C^∞ functions, C^r functions, $r \in \mathbb{N}$, continuous functions with a unique global optimum, Lipschitz continuous functions, and functions with Lipschitz continuous derivatives, among many others. We propose an explicit extension of the deterministic theorem by Stephens and Baritompa [35].

THEOREM 7 (STEPHENS AND BARITOMPA 1998) *A deterministic sequential sampling algorithm on a sufficiently rich class of functions $\mathscr{F}$ sees the global optimum of every function $g \in \mathscr{F}$ if and only if $\overline{X_f} = W$ for every function $f \in \mathscr{F}$.*

A possible extension to multiple objectives can be stated as:

THEOREM 8 *A deterministic sequential sampling algorithm on a sufficiently rich class of vector functions $\mathscr{F}$ sees the global Pareto optima for every function $g \in \mathscr{F}$ if and only if $\overline{X_f} = W$ for every $f \in \mathscr{F}$.*

Proof If $\overline{X_f} = W$ for every $f \in \mathscr{F}$ the extraction of the nondominated set from the sample sequence X_k will obviously strictly localize the Pareto set.

Conversely, if $\overline{X_f} \subsetneq W$, it is possible to consider a ball B_δ with non zero radius δ centered in a point x_0 of $W \setminus \overline{X_f}$. It is possible to redefine f as a new function $g \in \mathscr{F}$ such that each of the components of g is deformed smoothly only inside B_δ such that $g_i(x_0) > max_{x\in W} f_i(x)$, for all $i = 1, \ldots, m$. (See e.g., Figure 2). When running the algorithm on g rather than on f exactly the same X_f will be produced, therefore it is impossible that the branch of global optima contained in the ball B_δ would be detected. □

4 A globally convergent algorithm in two dimensions

As a *pars construens* of this work, we propose a globally convergent algorithm in two dimension, for a smooth vector function f with prescribed regularity. In particular we

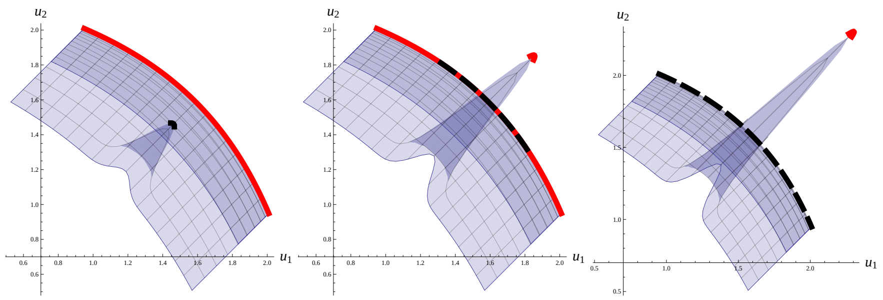

Fig. 2 Modificatins in an arbitrarily small subset of the domain can completely alter the Pareto optimal set of a function.

assume that f is *good* in the sense of Whitney [40], i.e., we assume that 0 is a regular value for the determinant of the Jacobian $\det(Df)$. This means that we have either $\det Df \neq 0$ or $D(\det Df) \neq 0$. As a consequence of the implicit function theorem, we have that the singular set $\Sigma = \{p | \det Df(p) = 0\}$ is a collection of smooth curves in W. As discussed above in Section 2.3, good functions form an open and dense set in C^∞. We recall that for extending the present methods to the Pareto critical set and to the stable Pareto critical set, one needs to resort to theorems 3 and 4.

Furthermore, we need a piece of global information in order to define a non space–filling global search strategy:

$$\begin{array}{ll} f: W \subseteq \mathbb{R}^2 \longrightarrow \mathbb{R}^2 & f \in C^\infty,\ f \text{ is } \textit{good}, \quad W \text{ is an equilateral triangle}, \\ \exists L > 0, \forall p_0, p_1 \in W, & |\det Df(p_0) - \det Df(p_1)| < L\,|p_0 - p_1|\,. \end{array} \tag{6}$$

This assumption that W is a triangle can be relaxed to deal with more general cases by assuming that the domain W is tessellated by a finite number of triangles, or also iteratively approximated by a suitable tessellation. We describe here only the procedure for tracing the full singular set Σ. The refinement to the Pareto optimal set is discussed in [18].

4.1 Algorithm discussion

We discuss two features of the algoritm: first we prove that eliminated triangles do not contain portions of the singular set. So the singular set will be certainly weakly localized. Next we will prove that sooner or later the triangles not crossing the singular set will be discarded. This is enough to establish that the singular set will be strictly localized by the sequence of the S_k.

Next we will discuss a more subtle question related to the last step of the algorithm, which actually defines an alternative sequence of approximating subsets S'_k. Each of the S'_k is a polygonal curve approximating Σ, obtained by considering the triangles on which vertices $\det(Df)$ assumes values of both signs. By inverse linear interpolation, one gets two points of intersection along two edges of the triangle, and

Algorithm 1 *Algorithm for approximating the singular set Σ with a prescribed accuracy ε*

Let $W = S_0 = T_0$, an equilateral triangle, with side length $\ell = \ell_0 > 0$.
Let $\varepsilon > 0$, $\varepsilon \ll \ell_0$ the prescribed accuracy for the determination of Σ.
Evaluate $d_{0,i} := \det(Df(x_i))$ for every vertex x_i of T_0, $i = 1,2,3$.
Set $Q = \{T_0\}$.
repeat
  Elimination of triangles not containing Σ:
  for each triangle T_α in Q, $\alpha = 1,\ldots,N$, **do**
    if $|d_{\alpha,i}| > L \times \ell$, for all $i = 1,2,3$ **then**
      discard T_α from Q.
    end if
  end for
  Refinement of remaining triangles in Q. Set $Q' = \{\emptyset\}$.
  for each triangle T_α in Q, $\alpha = 1,\ldots,N'$, **do**
    Remove T_α from Q.
    Split T_α in four triangles by halving the edges.
    Evaluate $\det(Df(x_j))$ in every new node x_j.
    Put the newly defined four triangles with half edge length in the new queue Q'.
  end for
  Set $Q = S_{k+1} \to Q'$ and $\ell_{k+1} \to \ell_k/2$.
until $\ell_k \leqslant \varepsilon$ or $Q = \emptyset$
S_{k+1} is an ε–approximation of Σ.
By application of *singular continuation*, i.e., inverse linear interpolation of the zero level of $\det(Df(x))$, one gets a polygonal chain $S'_{k+1} \subseteq S_{k+1}$.

the segment is a surrogate for the portion of Σ intersecting the triangle. Actually, this is the core of the singular continuation strategy [18].

We claim that this piece–wise linear surrogate, S'_k, is homeomorphic to Σ for every triangle size ℓ_k smaller than a certain value ℓ_c.

In particular, there does not exist arbitrarily loops, or arbitrarily sharp turns in Σ, which could be missed by a finite size triangulation, but under a certain size the shape of Σ is essentially captured by its linear surrogate S'_k, and moreover the surrogate is also quadratically precise, i.e.,

$$d_{\mathscr{H}}\left(\Sigma, S'_k\right) < C\ell^2. \tag{7}$$

This Newton–like convergence speed is proved in detail in [18].

PROPOSITION 9 *If assumption (6) holds, the sequence of subsets S_k defined in Algorithm 1 strictly localizes the singular set Σ.*

Proof Let $\ell_k = \ell_0(\frac{1}{2})^k$ be the triangles size at the kth iteration, with $\ell_0 > 0$ the size of the starting triangle. Denote by $\left\{T_\alpha^{\ell_k}\right\}_{\alpha \in I_{\ell_k}}$ the sequence of the retained triangles at the kth iteration, with I_{ℓ_k} the set of indices. We are going to prove that

$$\bigcap_{k=1}^{\infty} S_k = \bigcap_{k=1}^{\infty} \bigcup_{\alpha \in I_{\ell_k}} T_\alpha^{\ell_k} = \Sigma. \tag{8}$$

For doing so it is sufficient to prove that, for each fixed iteration ℓ_k, the singular set is contained in $S_k = \cup_\alpha T_\alpha^{\ell_k}$, i.e., the union of the retained triangles, and that every empty triangle not already discarded will be sooner or later eliminated from the sequence.

Let T_α be one of the discarded triangles, i.e.,

$$|d_{\alpha,i}| > L \times \ell_k, \qquad \text{for all } i = 1,2,3.$$

Assume also there is a point $x \in T_\alpha$, $\det Df(x) = 0$. Then $|d_{\alpha,1}| = |d_{\alpha,1} - \det Df(x)| \leqslant L|d_{\alpha,1} - x| < L \times \ell_k$, a contradiction.

On the other hand, let T be a triangle not discarded at the moment, but not containing any portion of the singular set. Let P the point of the triangle boundary ∂T realizing the minimum value of $|\det Df|$, $D := |\det Df(P)| = \min_{x\in\partial T} |\det Df(x)|$. Then it is clear that at the first iteration of the algorithm when $\ell_h < \frac{D}{L}$, each of the possibly survived sub–triangles obtained from the refinement of T will be discarded. □

THEOREM 10 *Let $f : W \to \mathbb{R}^2$ be* good*, i.e., $f \in C^\infty$ and 0 is a regular value for* $\det(Df(x))$. *If ℓ_k is sufficiently small, the poligonal curve S'_k obtained by inverse linear interpolating the zero set of* $\det(Df)$ *is homeomorphic to Σ.*

Proof For simplicity, we denote by $J : \mathbb{R}^2 \to \mathbb{R}$ the function whose zero set is Σ, i.e.,

$$\begin{gathered} J(x,y) := \det \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}, \qquad \text{i.e., } \Sigma = \left\{ p \in W \,\middle|\, J(p) = 0 \right\}, \\ L_{DJ} := \max_{p\in W} \left\| D^2 J(p) \right\| = \max_{\substack{p\in W,\\ v\in\mathbb{R}^2, |v|=1}} \left| D^2 J(p)\cdot v \right|, \qquad m_{DJ} := \min_{p\in\Sigma} |DJ(p)|, \end{gathered} \tag{9}$$

i.e., L_{DJ} is the Lipschitz constant of DJ. Moreover, because at least one among $\frac{\partial}{\partial x}J$ or $\frac{\partial}{\partial y}J$ is non zero, the implicit function theorem applies and $\Sigma = \{J = 0\}$ is composed of smooth curves and finally $m_{DJ} > 0$, also because of compactness.

In fact, we we are going to prove something stronger than the topological equivalence. We start by showing that Σ intersects "nicely" the triangles edges, for all sizes ℓ smaller than a critical value $\ell_c > 0$. Let T be one of the triangles of the tessellation of W. We have that,

(i) it is not restrictive to assume that J never vanishes on the nodes of the tessellation.
(ii) If $\Sigma \cap T \neq \emptyset$, then $DJ \neq 0$ inside T.
(iii) If Σ has a loop, the loop contains at least a vertex of a triangle.
(iv) Let $[p_0, p_1]$ one of the edges of T where $J(p_0)J(p_1) < 0$. Then Σ intersects only once $[p_0, p_1]$ in a point a. The remaining edge where J changes sign on the extrema carries another intersection point b.
(v) The curved interval of Σ joining a and b and intersecting T is diffeomorphic to the line segment $[a,b]$.

We have that (i) holds for any finite size ℓ, because the set of vertices is finite and Σ has measure zero. Therefore, in case of intersection, we could slightly modify the position of the starting vertices. Next, let $DJ(y)=0$ and $x\in\Sigma$. Then we have

$$m_{DJ}\leqslant|DJ(x)|=|DJ(x)-DJ(y)|\leqslant L_{DJ}\,|x-y|\,,\;\Rightarrow\;|x-y|>\frac{m_{DJ}}{L_{DJ}},$$

and (ii) holds whenever ℓ is smaller than $\frac{m_{DJ}}{L_{DJ}}$.

Consider now a small loop of Σ. Along such a loop there are infinite pairs p_0,p_1 such that $DJ(p_0)$ is parallel but opposite to $DJ(p_1)$. Therefore,

$$2m_{DJ}\leqslant|DJ(p_0)-DJ(p_1)|\leqslant L_{DJ}\,|p_0-p_1|\,,$$

which implies that $|p_0-p_1|>\frac{2m_{DJ}}{L_{DJ}}$. Thus, if $\ell<\frac{2m_{DJ}}{L_{DJ}}$ no such loop can be contained in a unique triangle. See the blue curve in panel (a) of Figure 3. Then also (iii) holds.

Now let p_0,p_1,p_2 the vertices of T with, e.g., $J(p_0)>0$ while $J(p_1)<0$ and $J(p_2)<0$. Then Σ intersects $[p_0,p_1]$ in a unique point a and $[p_0,p_2]$ in b. It is possible that Σ invades the adjacent triangle T' crossing back and forth along the remaining edge $[p_1,p_2]$. However, it is not possible that Σ, stretching over T', could cross another edge of T'. In such a case, as depicted with the red line in panel (a) of Figure 3, the gradients of J in at least two points of Σ will form an angle of at least $\frac{\pi}{3}$. Then these two intersection must be at least m_{DJ}/L_{DJ} far apart, because $L_{DJ}\,|x_0-x_1|\geqslant|DJ(x_0)-DJ(x_1)|\geqslant m_{DJ}$.

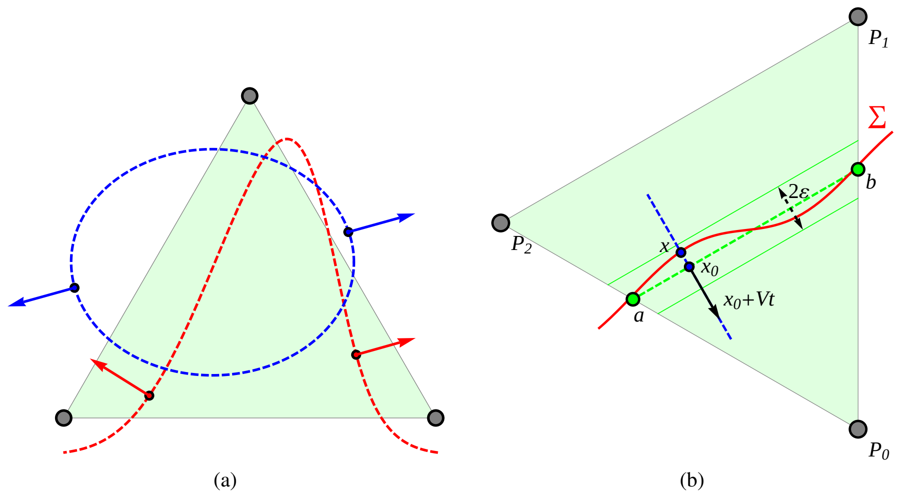

Fig. 3 (a): possible problematic behavior of Σ when intersecting the triangles in the tessellation. (b): the construction of the diffeomorphism $\sigma(\cdot)$ between $\Sigma_{a,b}$ and the segment $[a,b]$.

We now build explicitly the diffeomorphism among $[a,b]$ and the corresponding interval of Σ:

$$\sigma_T:[a,b]\longrightarrow\Sigma_{a,b}. \tag{10}$$

The construction of $\sigma(\cdot)$ is illustrated in Figure 3. We define a vector V "almost equal" to the gradient of J over T, by setting

$$V := \lambda \cdot \frac{(b-a)^{\perp}}{|b-a|}, \qquad \lambda = \pm \langle |DJ(x)| \rangle_{x \in T}$$

where the sign $\pm$ is chosen such that $V \cdot DJ > 0$ in T. We have that $DJ(x) = V + O(\ell)$ for all $x \in T$. Now we will see that $\Sigma_{a,b}$ is contained in a tubular neighborhood of $[a,b]$ with quadratically small thickness, i.e., $d_{\mathscr{H}}([a,b],\Sigma_{a,b}) < \varepsilon$, with $\varepsilon \sim \ell^2$. Indeed, notice that $\widehat{J}(x) := V \cdot (x-a)$ is a linear approximation of J, i.e., $J(x) = \widehat{J}(x) + O(\ell^2)$. Let $x \in \Sigma$, thus $J(x) = 0$ and $|V \cdot (x-a)| \leqslant C\ell^2$ and therefore $|x - x_0| \leqslant \frac{C}{|V|}\ell^2$, where x_0 is the foot of the normal to $[a,b]$ passing through x. Then we can conclude that $\Sigma_{a,b}$ is contained in an ε–tubular neighborhood of $[a,b]$, $\varepsilon := \frac{C}{|V|}\ell^2$.

Next, let us consider the normal to $[a,b]$ passing by $x_0 \in [a,b]$. We have

$$J\left(x_0 + \frac{V}{|V|^2}t\right) = J(x_0) + DJ(x_0)\frac{V}{|V|^2}t + O(t^2) = J(x_0) + t + O(t^2)$$

therefore, being $J(x) \simeq \varepsilon$, and $t^2 \ll t$, we have only one value $\tilde{t}_{x_0}$ of t in $[-\varepsilon,\varepsilon]$ for which J vanishes.

We write $\sigma(x_0) = x_0 + V\tilde{t}_{x_0}$, with possibly $\tilde{t}_{x_0} = 0$, e.g., for $x_0 = a$ or b. $\sigma(\cdot)$ is a diffeomorphism because the linear perpendicular projection from T to $[a,b]$ is C^{∞}, and because the correspondence among $\Sigma_{a,b}$ and $[a,b]$ is one–to–one. Finally, we notice that S'_k does not coincide exactly with $[a,b]$ inside T, because the linear interpolation of the values of J on the vertices of T is not supposed to vanish on a or b. However, it is clear that S'_k and $[a,b]$ are ε–close and that they are trivially diffeomorphic. □

In Figure 4 we illustrate the sequence $\{S_k\}$ obtained by applying Algorithm 1 to the function in (1). The procedure has been iterated nine times, while the Lipschitz constant adopted were $L_H = 45.0$.

5 Conclusions

We have discussed the notion of global convergence in the case of multiple objectives, by using the global analysis framework of Smale, and we restated the everywhere dense sampling necessity discussed by Stephens and Baritompa. We also presented a Lipschitz global bi–objective optimization algorithm in the spirit of the Pijavskij–Shubert algorithm.

Extensions are possible along two directions. First of all it seems attractive adapting the strategy for higher dimensions. However, this seems far than trivial, because spaces with dimension larger or equal than three cannot be tessellated by means of regular simplexes as in dimension two. On the one hand, adopting non regular simplexes could lead to skinny simplexes, which do not behave well for the numerical stability of algorithms [30]. On the other hand, the adoption of different geometrical

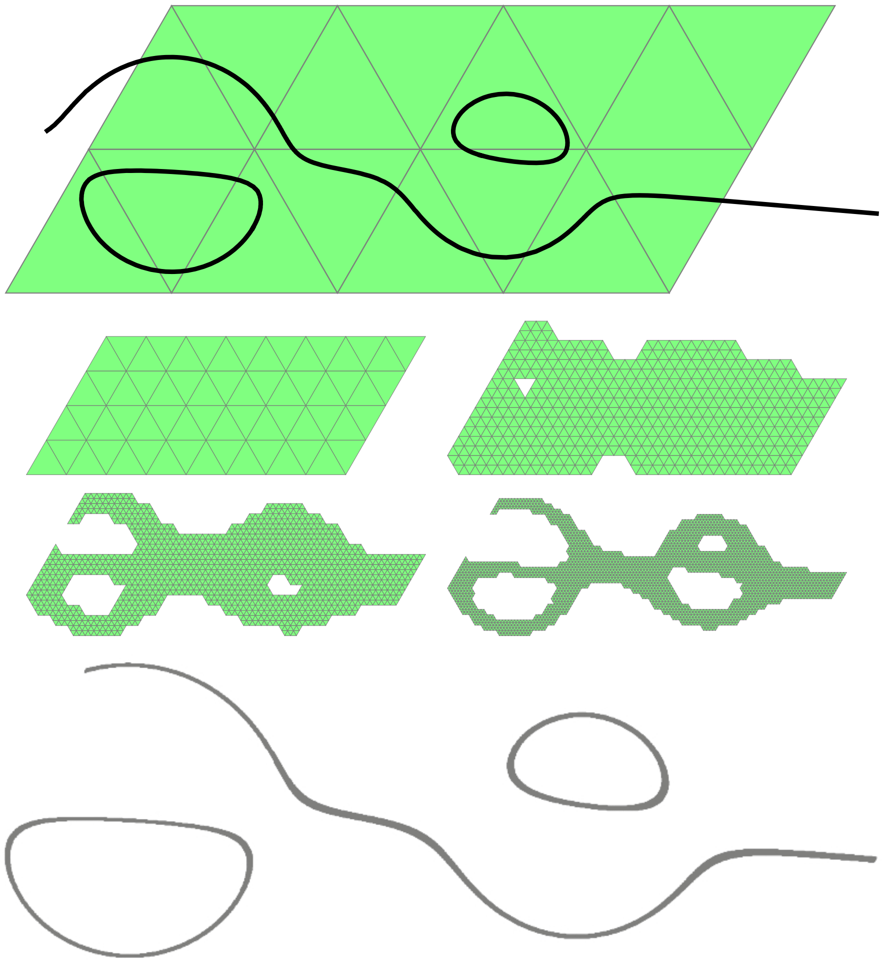

Fig. 4 Application of algorithm 1 to the function (1) for nine iterations and a Lipschitz constant of 45.0. We reported the starting setting and iterations 1, 3, 4, 5 and 9.

entities (e.g., hypercubes) appears not suitable for the linear interpolation adopted in [18] and in general continuation strategies [1–4]. Nevertheless, hypercubes has been employed successfully for isolating the critical set and for combining this strategy with local search techniques [27].

Secondly, it seems interesting to remove the assumption on the Lipschitz constant and devise extensions of efficient global optimization strategies such as [12–15, 19, 26, 28, 29, 36, 41] to multiple objectives.

Finally, one could try to incorporate the point of view of global optimization in the standard setting of Karush–Kuhn–Tucker conditions, as considered in [6, 10, 27].

References

1. Allgower, E.L., Georg, K.: Simplicial and continuation methods for approximating fixed points and solutions to systems of equations. SIAM Rev. **22**(1), 28–85 (1980)
2. Allgower, E.L., Georg, K.: Piecewise linear methods for nonlinear equations and optimization. Journal of Computational and Applied Mathematics **124**(1-2), 245–261 (2000)
3. Allgower, E.L., Gnutzmann, S.: An algorithm for piecewise linear approximation of implicitly defined two-dimensional surfaces. SIAM J. Numer. Anal. **24**(2), 452–469 (1987)
4. Allgower, E.L., Sommese, A.J.: Piecewise linear approximation of smooth compact fibers. J. Complexity **18**(2), 547–556 (2002)
5. Arnol′d, V.I.: Singularities of smooth mappings. Russian Mathematical Surveys **23**(1), 1–43 (1968). URL http://stacks.iop.org/0036-0279/23/1
6. Askar, S., Tiwari, A.: Multi-objective optimisation problems: A symbolic algorithm for performance measurement of evolutionary computing techniques. In: M. Ehrgott (ed.) EMO 2009, *Lecture Notes in Computer Science*, vol. 5467, pp. 169–182. Springer-Verlag (2009)
7. Benson, H., Sayin, S.: Towards finding global representations of the efficient set in multiple objective mathematical programming. Naval Research Logistics **44**(1), 47–67 (1997)
8. Guillemin, V., Pollack, A.: Differential topology. Prentice-Hall (1974). URL http://www.ulb.tu-darmstadt.de/tocs/59605480.pdf
9. Hartikainen, M., Miettinen, K., Wiecek, M.M.: Constructing a pareto front approximation for decision making. Math. Methods Oper. Res. **73**(2), 209–234 (2011). DOI 10.1007/s00186-010-0343-0
10. Hillermeier, C.: Generalized homotopy approach to multiobjective optimization. J. Optim. Theory Appl. **110**(3), 557–583 (2001)
11. Hillermeier, C.: Nonlinear multiobjective optimization, *International Series of Numerical Mathematics*, vol. 135. Birkhäuser Verlag, Basel (2001). A generalized homotopy approach
12. Jones, D.R.: A taxonomy of global optimization methods based on response surfaces. J. Global Optim. **21**(4), 345–383 (2001)
13. Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the lipschitz constant. J. Optim. Theory Appl. **79**(1), 157–181 (1993)
14. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. J. Global Optim. **13**(4), 455–492 (1998)
15. Khompatraporn, C., Pintér, J.D., Zabinsky, Z.B.: Comparative assessment of algorithms and software for global optimization. J. Global Optim. **31**(4), 613–633 (2005)
16. Knowles, J.: ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Trans. Evolutionary Computation **10**(1), 50–66 (2006)
17. Levine, H.: Singularities of differentiable mappings. In: C. Wall (ed.) Proceedings of Liverpool Singularities — Symposium I, *Lecture Notes in Mathematics*, vol. 192, pp. 1–21. Springer Berlin / Heidelberg (1971). DOI 10.1007/BFb0066810. URL http://dx.doi.org/10.1007/BFb0066810
18. Lovison, A.: Singular continuation: Generating piecewise linear approximations to pareto sets via global analysis. SIAM J. Optim. **21**(2), 463–490 (2011). DOI 10.1137/100784746. URL http://dx.doi.org/doi/10.1137/100784746
19. Lovison, A., Rigoni, E.: Adaptive sampling with a Lipschitz criterion for accurate metamodeling. Communications in Applied and Industrial Mathematics **1**(2), 110–126 (2010)
20. de Melo, W.: On the structure of the Pareto set of generic mappings. Bol. Soc. Brasil. Mat. **7**(2), 121–126 (1976)
21. de Melo, W.: Stability and optimization of several functions. Topology **15**(1), 1–12 (1976)
22. Miettinen, K.: Nonlinear multiobjective optimization. International Series in Operations Research & Management Science, 12. Kluwer Academic Publishers, Boston, MA (1999)
23. Miglierina, E., Molho, E.: Scalarization and stability in vector optimization. J. Optim. Theory Appl. **114**(3), 657–670 (2002)
24. Miglierina, E., Molho, E., Rocca, M.: Critical points index for vector functions and vector optimization. J. Optim. Theory Appl. **138**(3), 479–496 (2008). DOI 10.1007/s10957-008-9383-5
25. Pijavskiĭ, S.A.: An algorithm for finding the absolute minimum of functions. In: Theory of Optimal Solutions (Seminar, Kiev, 1967), No. 2 (Russian), pp. 13–24. Akad. Nauk Ukrain. SSR, Kiev (1967)
26. Pintér, J.D.: Nonlinear optimization with GAMS/LGO. J. Global Optim. **38**(1), 79–101 (2007). DOI 10.1007/s10898-006-9084-2. URL http://dx.doi.org/10.1007/s10898-006-9084-2

27. Schütze, O., Dell'Aere, A., Dellnitz, M.: On continuation methods for the numerical treatment of multi-objective optimization problems. In: J. Branke, K. Deb, K. Miettinen, R.E. Steuer (eds.) Practical Approaches to Multi-Objective Optimization, no. 04461 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, Dagstuhl, Germany (2005). URL http://drops.dagstuhl.de/opus/volltexte/2005/349
28. Sergeyev, Y.D.: An information global optimization algorithm with local tuning. SIAM J. Optim. **5**(4), 858–870 (1995). DOI 10.1137/0805041. URL http://dx.doi.org/10.1137/0805041
29. Sergeyev, Y.D., Kvasov, D.E.: Global search based on efficient diagonal partitions and a set of lipschitz constants. SIAM J. Optim. **16**(3), 910–937 (2006)
30. Shewchuk, J.R.: Delaunay refinement algorithms for triangular mesh generation. Computational Geometry: Theory and Applications **22**(1-3), 21–74 (2002)
31. Shubert, B.O.: A sequential method seeking the global maximum of a function. SIAM J. Numer. Anal. **9**, 379–388 (1972)
32. Smale, S.: What is global analysis? Amer. Math. Monthly **76**, 4–9 (1969)
33. Smale, S.: Global analysis and economics. I. Pareto optimum and a generalization of Morse theory. In: Dynamical systems (Proc. Sympos., Univ. Bahia, Salvador, 1971), pp. 531–544. Academic Press, New York (1973)
34. Smale, S.: Optimizing several functions. In: Manifolds–Tokyo 1973 (Proc. Internat. Conf., Tokyo, 1973), pp. 69–75. Univ. Tokyo Press, Tokyo (1975)
35. Stephens, C., Baritompa, W.: Global optimization requires global information. J. Optim. Theory Appl. **96**, 575–588 (1998). DOI 10.1023/A:1022612511618. URL http://dx.doi.org/10.1023/A:1022612511618
36. Strongin, R.G., Sergeyev, Y.D.: Global optimization: fractal approach and non-redundant parallelism. J. Global Optim. **27**(1), 25–50 (2003)
37. Thom, R.: Les singularités des applications différentiables. Ann. Inst. Fourier (1956)
38. Wan, Y.H.: Morse theory for two functions. Topology **14**(3), 217–228 (1975)
39. Wan, Y.H.: On local Pareto Optima. J. Math. Econom. **2**(1), 35–42 (1975)
40. Whitney, H.: On Singularities of Mappings of Euclidean Spaces. I. Mappings of the Plane into the Plane. The Annals of Mathematics **62**(3), 374–410 (1955). URL http://www.jstor.org/discover/10.2307/1970070
41. Zhang, B., Wood, G.R., Baritompa, W.P.: Multidimensional bisection: the performance and the context. J. Global Optim. **3**(3), 337–358 (1993). DOI 10.1007/BF01096775. URL http://dx.doi.org/10.1007/BF01096775