



HAL
open science

Heuristic global optimization and identification of general pareto fronts

Bijan Mohammadi

► **To cite this version:**

Bijan Mohammadi. Heuristic global optimization and identification of general pareto fronts. 2017.
hal-01474671

HAL Id: hal-01474671

<https://hal.science/hal-01474671>

Preprint submitted on 23 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

HEURISTIC GLOBAL OPTIMIZATION AND IDENTIFICATION OF GENERAL PARETO FRONTS

BIJAN MOHAMMADI

MONTPELLIER UNIVERSITY

MATHEMATICS (IMAG), CC51, 34095 MONTPELLIER, FRANCE

ABSTRACT. We recall some of the multi-criteria and multi-disciplinary optimization formulations with emphasis on algorithmic and implementation issues. We use a partial controllability result which states that the vicinity of any point in the optimization space can be reached in finite time from any other point in the space following a trajectory solution of a particular second order dynamical system. We use this continuous point of view to propose a class of heuristic optimization algorithms also addressing global search issues. For multi-criteria problems this means that any point on a Pareto front can be reached even if the front is non-convex as far as the optimization algorithm has inertial features. The paper also recalls the conditions under which a Pareto front can be accessed by descent methods applied to multicriteria problems using the weighted sum approach. The question of the density of the distribution of the points over the front is also addressed. The approach is illustrated on academic examples as well as on a multi-criteria distributed inverse problem for a heterogeneous data fusion problem.

1. INTRODUCTION

The paper is dedicated to two problems of interest in optimization. First, we address the question of how descent methods can be improved to perform global search ? Once this question is discussed, one sees why global search methods are necessary to reach points on general non-convex Pareto fronts where descent methods fail. Addressing these questions leads to the development of a class of heuristic global optimization methods as discretization of special 2^{nd} order continuous dynamical systems.

Descent algorithms can be interpreted as discrete forms of first order systems [20]: a descent method solves a Cauchy problem for such a system. Now, suppose we know the infimum of the functional and the question of global optimization is then where it is reached ? Global optimization becomes then a boundary value problem with an initial condition in the admissible space and a final condition on the functional. Therefore, solving a global optimization problem with such methods leads to an over-determined problem. In other words, if the final condition is given, the initial condition should be prescribed and should be a variable of the problem as well. Intuitively, this suggests that a descent method is willing to succeed if it starts somewhere in the optimal attraction basin (i.e. where the infimum is).

The paper discusses how a controllability result can be used in the development of heuristic optimization algorithms using recursive shootings in order to solve a

Key words and phrases. Heuristic optimization, Multidisciplinary, multicriteria, Optimization, data fusion, Pareto front.

boundary value problem. We recover the philosophy behind the Bock's multiple shooting algorithm [5] and the Lion's parareal method [15].

The paper then considers the application of these ideas to multi-criteria problems. We will see that weighted sums used to transform a multi-criteria problem into a mono-criteria one can still be used to recover general Pareto fronts and not only convex ones. This is interesting because this method is widely used and appreciated by engineers.

Finally, we discuss the question of the distribution and localization of the points on the Pareto front. Indeed, we will see that despite general fronts can be reached, the points are often oddly distributed with concentrations in some regions of the front while we might want some kind of uniform distribution of the points on the front. To address this issue, we propose to couple the different trials introducing a particular repulsion force in the corresponding second order dynamical systems. This point is however not central in the paper and the concepts of the paper can be used together with other proposals such as the normal boundary intersection or the normal constraint methods [8, 1]. The proposed method has however the advantage to remain compatible with the weighted sum formulation and easy to explain in engineering environments.

2. MULTI-CRITERIA OPTIMIZATION

The literature on multi-criteria optimization is vast [7, 8, 9, 17, 22, 24]. Let us briefly recall some notions and take this opportunity to introduce the notations which will be used throughout the paper. We consider n real functions $\{J_i(x), i = 1, \dots, n\}$. One way to look at multidisciplinary optimization is minimizing a vector [16]:

$$(1) \quad \min_{x \in \mathbf{O}_{ad} \subset \mathbb{R}^N} \vec{\mathbf{J}}, \quad \vec{\mathbf{J}} = (J_1(x), \dots, J_n(x))^t,$$

which brings the question of order relation in a vector space which is in turn related to the norm one chooses for the optimization: shall we consider an L^1 , L^2 or ... L^∞ norm ? or shall we consider several or all of them ? There is no single answer to this question. Instead, one should look for trade-off between conflicting objectives and Pareto equilibrium [22].

2.1. Multicriteria optimization algorithms. Let us recall a few strategies to solve problem (1) which has a wide literature dedicated to [11, 16, 17, 21]. This paper is not on introducing a new algorithm for this purpose. The algorithms can be cast into two classes considering either the functionals separately or through a unique representation, after weighted summation for instance.

A natural representative of the first class is the min-max approach where one looks for minimizing sequences for the following problem starting from an initial value $x_0 \in \mathbf{O}_{ad}$:

$$(2) \quad \min_{x \in \mathbf{O}_{ad}} \max_{j=1, \dots, n} \left\{ \frac{J_j(x)}{J_j(x_0)} \right\}.$$

To make sure the sequence x^p is a descent sequence for each of the vector functional components, one should require a new proposal x^{p+1} not to degrade any of the current achievements. This means that x^{p+1} should be solution of the following

constrained minimization problem:

$$\min_{x \in \mathbf{O}_{ad}} \max_{j=1, \dots, n} \left\{ \frac{J_j(x)}{J_j(x_0)} \right\} \quad \text{such that} \quad C_i = J_i(x) - J_i(x^p) \leq 0, \quad i = 1, \dots, n, \quad i \neq j.$$

This can be achieved solving the KKT conditions for the Lagrangian $L_j(x, \Lambda) = J_j + \Lambda^t C$ with $C \in \mathbb{R}^{n-1}$ the vector of constraints and $\Lambda \in \mathbb{R}_+^{n-1}$ is the vector of positive Lagrange multipliers. This Lagrangian might change at each iteration. Looking to achieve the KKT conditions means building x^{p+1} aiming at vanishing $\nabla_x L_j = \nabla_x J_j + \Lambda^t \nabla_x C$ with the complementary condition $\Lambda^t C = 0$ making a distinction between an inactive constraint where $C_i < 0$ and $\lambda_i = 0$ and an active one where $C_i = 0$ and $\lambda_i > 0$. In this former situation we have $\lambda_i = -\partial J_j / \partial C_i$ meaning that the Lagrange multiplier gives the sensitivity of J_j with respect to a perturbation of this particular active constraint. Once the optimum is reached, this information permits a better construction of the Pareto front.

2.2. Geometric interpretation. This construction can be understood geometrically. To be admissible, a search direction d needs to be orthogonal to the following convex cone:

$$(3) \quad S_p = \left\{ x \mid x = \sum_{i=1}^p \beta_i \nabla_x C_i, \beta_i > 0 \mid C_i = 0 \right\} \subseteq \mathbb{R}^p \subset \mathbb{R}^{n-1},$$

where we have p active constraints. At the solution, $\nabla_x J$ is orthogonal to this cone. Before working on the cone, let us start defining a local orthonormal basis $\{\tilde{\mathbf{q}}_{i=1, \dots, p}\}$ for \tilde{S}_p defined by (3) but with $\beta_i \in \mathbb{R}$. \tilde{S}_p is a vector space and an orthonormal basis can be defined for \tilde{S}_p orthonormalizing the columns of $\nabla_x C$ by the Gram-Schmidt procedure. Now, we consider the vectors $\mathbf{q}_i = \pm \tilde{\mathbf{q}}_i$ and the sign chosen such that $\langle \mathbf{q}_{i=1, \dots, p}, \nabla_x C_k \rangle \geq 0$, if $C_k = 0$ for $k = 1, \dots, p$ (i.e. pointing inside the cone for an active constraint). $\{\mathbf{q}_{i=1, \dots, p}\}$ are the generators of the cone S_p deduced from a basis of \tilde{S}_p . If the generators cannot be defined, the problem is found having no solution as at least two of the constraints are incompatible with their gradients parallel and pointing in opposite directions. These generators permit to define an admissible search direction d at iteration n following:

$$(4) \quad d = \nabla_x J_j - \sum_{i=1}^p \chi_i \langle \mathbf{q}_i, \nabla_x J_j \rangle \mathbf{q}_i,$$

with $\chi_i = 0$ if $\langle \mathbf{q}_i, \nabla_x J_j \rangle \geq 0$ and $\chi_i = 1$ if $\langle \mathbf{q}_i, \nabla_x J_j \rangle < 0$. At the optimum, the Lagrange multipliers for active constraints correspond to $0 < \lambda_i = -\langle \mathbf{q}_i, \nabla_x J_j \rangle$. This construction can also be obtained applying the Uzawa algorithm for the solution of:

$$\min_{x \in \mathbf{O}_{ad}} \max_{\Lambda \in \mathbb{R}_+^{n-1}} L_j(x, \Lambda),$$

at each iteration of problem (2). The solution of (2) belongs to the Pareto front as there is no way to improve one of the criteria without degrading the others.

This approach requires the gradients of all the functionals and, for large dimensional optimization problems, this means the solution of n adjoint problems [20] at each iteration which can be very costly, especially with sophisticated state equations solvers. With black-box solvers and if the size of the problem is small finite

differences are often used instead of adjoint because of their natural parallelism and simplicity together with a weighted sum formulation to only handle one functional.

2.3. Weighted sum formulation. The classical method for multi-objective optimization remains the weighted sum method because of its simplicity [7, 8, 24]:

$$(5) \quad J(x) = \sum \alpha_i J_i(x), \quad \sum \alpha_i = 1, \quad \alpha_i \in [0, 1].$$

A sampling in α_i enables to reach various points in the Pareto front using a descent method for $J(x)$. Indeed any point $\bar{x} \in \mathcal{O}_{ad}$ minimizing J is a Pareto equilibrium. Actually any point $x \in \mathcal{O}_{ad}$ dominating \bar{x} would yield an improved weighted sum $J(x) < J(\bar{x})$ in contradiction with the minimizing property of \bar{x} .

The converse is true in a convex setting. That is, any Pareto equilibrium \bar{x} is a minimum point of some weighted sum J given by (5), provided that the functionals J_i and the set \mathcal{O}_{ad} are convex. Indeed set

$$C = \{ \xi \in \mathbb{R}^n, \exists x \in \mathcal{O}_{ad}, \xi_i \geq J_i(x), \forall i \in \{1, \dots, n\} \},$$

$$K = \{ \xi \in \mathbb{R}^n, \xi < J_i(\bar{x}), \forall i \in \{1, \dots, n\} \}.$$

These are convex nonvoid subsets of \mathbb{R}^n , and K is open. Moreover $C \cap K = \emptyset$, since any point $\xi \in C \cap K$ would verify $J_i(\xi) < J_i(\bar{x})$ in contradiction with \bar{x} being a Pareto equilibrium. Hence, there exists a hyperplane separating sets C and K (see figure 1); that is, there exist a vector $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n \setminus \{0\}$ and a scalar b such that

$$(6) \quad \xi \in K \Rightarrow \alpha \xi^t \leq b,$$

$$(7) \quad \xi \in C \Rightarrow \alpha \xi^t \geq b.$$

Let ξ be some fixed point in K . For all $\lambda \in \mathbb{R}^n$ with nonnegative components, $\xi - \lambda$ lies in K too, and with (6) we have $\alpha \xi^t - \alpha \lambda^t \leq b$. Hence necessarily $\alpha_i \geq 0$ for all $i \in \{1, \dots, n\}$, and there is no harm in supposing $\sum_{i=1}^n \alpha_i = 1$.

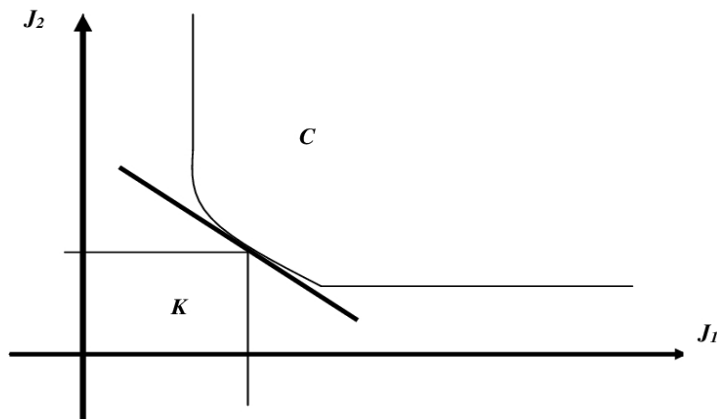


FIGURE 1. Hyperplane separating sets C and K .

Let \bar{K} denote the closure of K in \mathbb{R}^n . Since $(J_1(\bar{x}), \dots, J_n(\bar{x})) \in \bar{K} \cap C$, we have $\sum_{i=1}^n \alpha_i J_i(\bar{x}) \leq b \leq \sum_{i=1}^n \alpha_i J_i(\bar{x})$, hence $b = \sum_{i=1}^n \alpha_i J_i(\bar{x})$. Finally with (7) we obtain in particular

$$x \in \mathcal{O}_{ad} \Rightarrow \sum_{i=1}^n \alpha_i J_i(x) \geq \sum_{i=1}^n \alpha_i J_i(\bar{x}),$$

which means that \bar{x} is a minimum point of J .

A shorter proof may be given under reinforced assumptions. If \bar{x} is a Pareto equilibrium, it is a solution to the constrained optimization problem:

$$\inf \{J_1(x), x \text{ subject to } J_i(x) \leq J_i(\bar{x}), i \in \{2, \dots, n\}\}.$$

Indeed $J_i(x) \leq J_i(\bar{x})$, for $i \in \{1, \dots, n\}$, entails $J_i(x) = J_i(\bar{x})$ for $i \in \{1, \dots, n\}$. Under the extra assumption (see [13] e.g.)

$$\bigcap_{i \in \{2, \dots, n\}} \{x \in \mathcal{O}_{ad}, J_i(x) < J_i(\bar{x})\} \neq \emptyset,$$

there exist nonnegative multipliers α_i for $i \in \{2, \dots, n\}$ such that \bar{x} minimizes the functional $J_1 + \sum_{i \in \{2, \dots, n\}} \alpha_i (J_i - J_i(\bar{x}))$, and the functional $J_1 + \sum_{i \in \{2, \dots, n\}} \alpha_i J_i$ as well. Now divide the latter by $1 + \sum_{i \in \{2, \dots, n\}} \alpha_i$ to recover a weighted sum J such as given by (5).

It is reported that this approach does not find Pareto optimal solutions in non-convex regions of the Pareto front [2]. This is linked to the fact that when the tangent to the Pareto front is parallel to one of the axes the corresponding functional has a plateau which cannot be crossed by a descent method not having inertial features.

Now consider the situation of two functionals and a Pareto front $P(J_1, J_2) = 0$. In cases, where the front is non-convex but the tangent never parallel to one of the axes the implicit function theorem can be applied and one can find $J_1 = j_1(J_2)$, $J_2 = j_2(J_1)$. In these cases, as shown in [2], one can still use a descent method for increasing power m in the functional:

$$\tilde{J}(x) = \alpha J_1^m(x) + (1 - \alpha) J_2^m(x), \quad \alpha \in [0, 1].$$

The aim is to find the lowest m making the Pareto front $\tilde{P}(J_1^m, J_2^m)$ convex. But, this approach has the disadvantage of degrading the conditioning of the optimization problem.

In the sequel we show that the failure of the weighted sum approach is due to a lack in global minimization feature in descent methods and propose an alternative solution which is not submitted to the mentioned limitation on the tangent to the Pareto front.

3. CONTROLLABILITY AND GLOBAL OPTIMIZATION

Consider the minimization of a functional $J(x) \in \mathbb{R}$, $x \in \mathcal{O}_{ad}$. We suppose the problem admissible (i.e. there exist at least one solution x_m to the problem: $J(x_m) = J_m$). Most minimization algorithms can be seen as discrete expressions for the following Cauchy problem [20]:

$$(8) \quad M(x(t))x_t = -d(x(t)), \quad x(t=0) = x_0.$$

M is aimed to be positive definite and $M^{-1}d$ is built to be an admissible direction. Global solution of (8) means, for instance, finding $x(T)$ for finite T such that

$$(9) \quad M(x(t))x_t = -d(x(t)), \quad x(0) = x_0, \quad J(x(T)) = 0.$$

This is an over-determined boundary value problem and it tells us why one should not solve global optimization problems with methods which are discrete form of (8). This limitation can be somehow removed if one can provide an initial condition in the attraction basin of the global optimum. To remove this overdetermination one shall consider, for instance, the following second order system:

$$(10) \quad \eta x_{tt} + M(x(t))x_t = -d(x(t)), \quad x(t=0) = x_0, \quad J(x(T)) = J_m.$$

This can be solved using solution techniques for boundary value problems with free surface. An analogy can be given with the problem of finding the interface between water and ice which is only implicitly known through the iso-value of zero temperature [3].

3.1. Practical impacts. In practice the final condition is not aimed at being exactly realized. Rather, one would like, for a given precision δ in the functional, to build at least one trajectory $(x(t), 0 \leq t \leq T_\delta)$ passing for finite T_δ in the ball $B_\delta(x_m)$. This can be summarized as:

$$(11) \quad \forall \delta > 0, \exists (v, T_\delta) \in \mathcal{O}_{ad} \times [0, +\infty[\quad \text{such that} \quad J(x_v(T_\delta)) - J_m \leq \delta.$$

If J_m is unknown T_δ defines the maximum calculation complexity wanted. This is how optimization algorithms are actually used: one chooses the maximum calculation/memory effort one can afford and one looks for the best solution for this complexity. In these cases, setting $J_m = -\infty$, one retains the best solution obtained over $[0, T_\delta]$. In other words, one solves:

$$(12) \quad \forall (\delta, T_\delta) \in \mathbb{R}^+ \times [0, +\infty[, \exists (v, \tau) \in \mathcal{O}_{ad} \times [0, T_\delta], \\ \text{such that} \quad J(x_v(\tau)) - J_m \leq \delta.$$

The following theorem formalizes this controllability question with $M = Id$ and $d = \nabla J$.

Theorem 1: Let $J : \mathbb{R}^N \rightarrow \mathbb{R}$ be a C^2 -function such that $\min_{\mathbb{R}^N} J$ exists and is reached at $x_m \in \mathbb{R}^N$. Then for every $(x_0, \delta) \in \mathbb{R}^N \times \mathbb{R}^+$, there exists $(\sigma, t') \in \mathbb{R}^N \times \mathbb{R}^+$ such that the solution of the following system:

$$(13) \quad \eta x_{tt}(t) + x_t(t) = -\nabla J(x(t)), \quad t \geq 0, \quad x(0) = x_0, \quad x_t(0) = \sigma,$$

with $\eta \in \mathbb{R}$, passes at time $t' = t$ into the ball $B_\delta(x_m)$.

Proof :

We assume $x_0 \neq x_m$ ($x_0 = x_m$ is a trivial case). Let $\varepsilon > 0$, we consider first the initial value problem:

$$(14) \quad \eta y_{tt}(t) + \varepsilon y_t(t) = -\varepsilon^2 \nabla J(y(t)), \quad t \geq 0, \quad y(0) = x_0, \quad y_t(0) = \varrho(x_m - x_0),$$

with $\varrho \in \mathbb{R}^+ \setminus \{0\}$.

Let us show that $y(\cdot)$ passes at some time into the ball $B_\delta(x_m)$:

- If $\varepsilon = 0$, we obtain the following system :

$$(15) \quad \eta y_{tt,0}(t) = 0, \quad t \geq 0, \quad y_0(0) = x_0, \quad y_{t,0}(0) = \varrho(x_m - x_0).$$

System (15) describes a straight line of origin x_0 and passing at some time θ_ϱ by the point x_m , i.e. $y_0(\theta_\varrho) = x_m$.

- If $\varepsilon \neq 0$, system (14) can be rewritten as $(w_2)_t = f(\eta, w, \varepsilon)$, with $w = (w_1, w_2)^t = (y(t), \eta y_t(t))^t$ and $f = -\varepsilon/\eta w_2 - \varepsilon \nabla J(w_1(t))$ satisfying the Cauchy-Lipschitz conditions (J is a C^2 function). Then $\lim_{\varepsilon \rightarrow 0} |y_\varepsilon(\theta_\varrho) - y_0(\theta_\varrho)| = 0$. Thus, for every $\delta \in \mathbb{R}^+ \setminus \{0\}$, there exists ε_δ such that for every $\varepsilon < \varepsilon_\delta$:

$$(16) \quad |y_\varepsilon(\theta_\varrho) - x_m| < \delta.$$

Finally, let us consider the change of variables $t' = \varepsilon_\delta t$ and $x(t') = y_{\varepsilon_\delta}(\frac{t'}{\varepsilon_\delta})$. Then system (14) becomes:

$$(17) \quad \eta x_{t't'}(t') + x_{t'}(t') = -\nabla J(x(t')), \quad t' \geq 0, \quad x(0) = x_0, \quad \dot{x}(0) = \frac{\varrho}{\varepsilon_\delta}(x_m - x_0).$$

Let $\vartheta = \varepsilon_\delta \theta_\varrho$. Under this assumption, $x(\vartheta) = y_{\varepsilon_\delta}(\theta_\varrho)$. Thus, due to (16): $|x(\vartheta) - x_m| < \delta$. We have found $\sigma = \frac{\varrho}{\varepsilon_\delta}(x_m - x_0) \in \mathbb{R}^N$ and $t_b = \vartheta \in \mathbb{R}^+$ such that the solution of system (13) passes at time t_b into the ball $B_\delta(x_m)$.

□

3.2. Algorithmic considerations. This theorem is insightful and permits to derive a global minimization algorithm which requires $\delta > 0$, the infimum J_m and $0 \leq T_\delta < \infty$ [4]. If J_m is unknown, one sets $J_m = -\infty$. Again, T_δ shall be seen as a maximal admissible complexity for the calculation. We then minimize the functional $h_{\delta, T_\delta, J_m} : \mathcal{O}_{ad} \rightarrow \mathbb{R}^+$:

$$(18) \quad h_{\delta, T_\delta, J_m}(v) = \min_{\mathcal{A}} (J(x_v(\tau)) - J_m),$$

where $\mathcal{A} = \{x_v(\tau) \in \mathcal{O}_{ad}, \tau \in [0, T_\delta]\}$. Hence, global minimization becomes a nested minimization problem where one looks to improve the initial impulse v for a 'core' minimization sub-problem to generate the trajectory $(x_v(\tau), 0 \leq \tau \leq T_\delta)$ by:

$$(19) \quad \eta x_{tt} + M(x(t))x_t = -d(x(t)),$$

$$x(t=0) = x_0, \quad x_t(t=0) = v, \quad t \in [0, T_\delta].$$

In cases one would like to use a core minimization algorithm which is a discrete form of a first order dynamical system (i.e. $\eta = 0$), (19) becomes:

$$(20) \quad M(x(t))x_t = -d(x(t)), \quad x(t=0) = v, \quad t \in [0, T_\delta].$$

In other words, the external minimization problem aims at improving the initial condition for the core optimization sub-problem. Both the initial condition and initial impulse can therefore be considered as optimization variables by the external optimization problem. To solve the external minimization problem, our approach is based on constructing a multi-level reduced-order parametric model by 'data assimilation' or 'learning' [4, 3, 14, 20]. This algorithm has shown great efficiency for cases where the functional is coercive and comes from perturbation of a convex functional [20]. However, we keep its discussion to a minimum as it is not central to our topic here.

4. APPLICATION TO THE IDENTIFICATION OF PARETO FRONTS

We saw how to link any couple of points in the admissible space in finite time with some given accuracy by a trajectory solution of a second order dynamical system. If the problem is admissible a point in a Pareto front corresponds to a point in the admissible control space. Theorem 1 states that the vicinity of this point can be reached in finite time from any other point in the control space following a trajectory solution of the second order system (13). The algorithm above can therefore be used to recover any point in a Pareto front with:

$$(21) \quad h_{\delta, T, \vec{J}_m}(v) = \min_{\mathcal{A}} \|\vec{J}(x_v(\tau)) - \vec{J}_m\|,$$

regardless of whether the front is convex or not. The unknown components of \vec{J}_m are set to $-\infty$ as in mono-objective minimization. One looks for the best solution at given complexity. In all cases, this means that one shall use a minimization method which has global search features to cross local minima in the case of non-convex Pareto fronts. Let us illustrate this construction on a simple example.

4.1. A first illustration. We consider the functionals $J_1 = x_1 + x_2$ and $J_2 = (1/x_1 + \|x\|^2) + \beta(\exp(-100(x_1 - 0.3)^2) + \exp(-100(x_1 - 0.6)^2))$ for $x \in [0.1, 1]^2$. Functional J_1 is bilinear and J_2 has local minima for non zero β . We consider the cases of $\beta = 0$ and $\beta = 3$. One expects a local minimization algorithm to be captured by local minima for the latter situation showing that the corresponding Pareto front generation requires a global minimization algorithm.

Figure 2 shows examples of convex and non-convex and non-connected Pareto fronts for these two functionals when the parameter space has been uniformly sampled. The figure also shows the front identified by a steepest descent method with optimal step size and when enriching the method with second order dynamics features as presented in the paper. This improves the global search capabilities of the method and permits to reach the non-convex region in the front. We have solved 20 minimization problems based on uniform sampling of the weighting parameter $\alpha \in [0, 1]$ in the weighted sum with the steepest descent method which corresponds to a discrete expression of (20). We have then solved the same minimization problems with a discrete expression of (19) with $M = Id, d = \alpha \nabla J_1 + (1 - \alpha) \nabla J_2$ and $\eta = 1$.

4.2. Improving the sampling of the Pareto front coupling dynamical systems . We saw that the vicinity of any admissible point in the Pareto front can be reached in finite time from any point in the admissible domain. As we are interested in a discretization of the Pareto front, this means we proceed with several minimization in parallel as described in previous sections. However, these minimization by our dynamical systems can interact in order to increase their global search features. This coupling can also be used to achieve a more uniform distribution of the points over the Pareto front to avoid the kind of situation one sees in figure 2 even for a convex Pareto front where points accumulate in some regions regardless of whether the dynamics is first or second order. A natural way to introduce the uniformity of the distribution is to ask for a point to be pushed away the positions already occupied by others. This can be done introducing a repulsive force in our dynamics. This couples the different paths using cross-information and involving a global gradient.

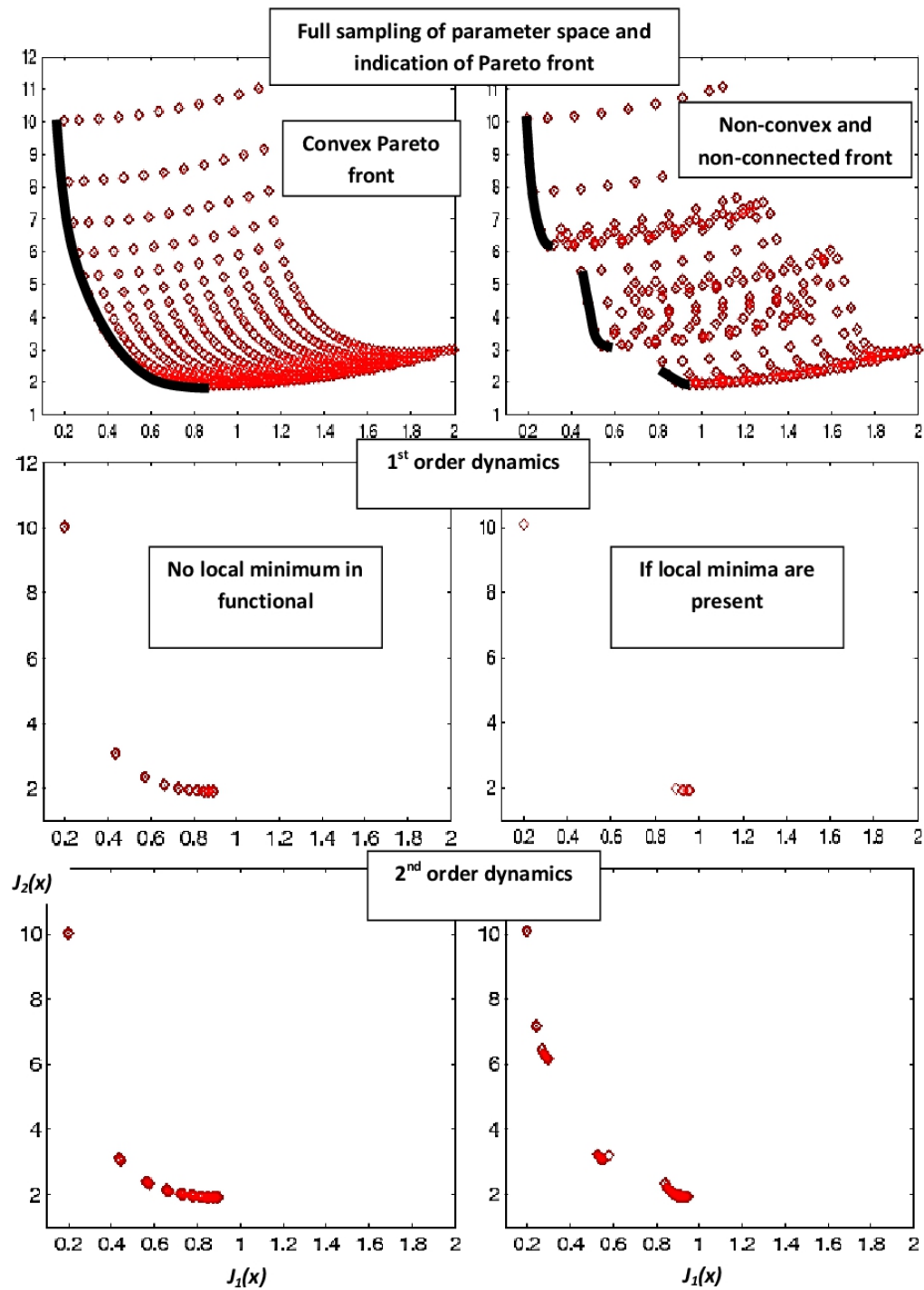


FIGURE 2. Top: Pareto front for (J_1, J_2) given in section 4.1 for $\beta = 0$ and $\beta = 3$ after uniform sampling of the parameter space. Solutions by optimization algorithms issued from discretization of a first (middle) and second-order (bottom) dynamical system for a uniform 20 point sampling of $\alpha \in [0, 1]$.

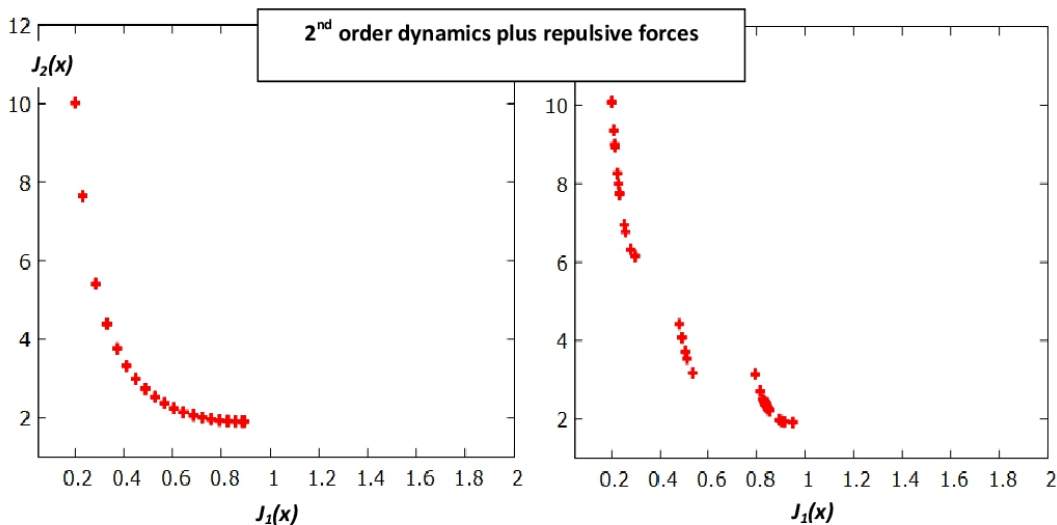


FIGURE 3. Pareto front by second-order dynamics accounting for repulsive forces (22) to be compared to lower images in figure 2.

More precisely, we consider q individuals $(x^j, j = 1, \dots, q)$ with their motions prescribed by q systems (19):

$$(22) \quad \eta x_{tt}^j + M x_t^j = d^j + G^j, \quad x^j(t=0) = x_0, \quad x_t^j(t=0) = v^j.$$

d^j is based on the gradient of the functional J^j given by (5) for a particular choice of the n-uplet $\alpha^j \in [0, 1]^n$ such that $\sum_{i=1}^n \alpha_i^j = 1$. G^j is a global gradient representing the interaction between the individuals:

$$G^j = - \sum_{k=1, k \neq j}^q \frac{J^j - J^k}{\|x^j - x^k\|^2} (x^j - x^k), \quad \text{for } j = 1, \dots, q.$$

This repulsive force aims at forcing the search to take place in regions not already visited. The right-hand side in (22) is therefore a balance between local and global forces.

This construction aims at achieving similar behavior than methods such as the normal boundary intersection [8] or the normal constraint [1]. It has the great advantage to be easy to implement and to be naturally parallel. Figure 3 shows the impact of this correction on the Pareto fronts given in figure 2. In the sequel, we will apply this combination to a more sophisticated multi-criteria and multi-physics data fusion problem.

5. APPLICATION TO MULTI-CRITERIA AND MULTI-PHYSICS DATA FUSION

As previously, our multi-criteria problem involves a vector of functionals:

$$\vec{J} = (J_1(u_1(x)), \dots, J_n(u_n(x)))^t \in \mathbb{R}^n,$$

for a same set of optimization parameters $x \in \mathbf{O}_{ad} \subset \mathbb{R}^N$. For simplicity we assume all the parameters are public and do not split the optimization parameters between public or global and private or local or disciplinary sets. One natural way to achieve this splitting in our context is to consider territory splitting strategies and to affect or not a given parameter $x_{i=1, \dots, N}$ to a given discipline j represented

by $J_{j=1,\dots,n}$ following its sensitivity dJ_j/dx_i . The difficulty is that these sensitivities are a priori unknown. This splitting is an important step in the formulation of a multidisciplinary optimization problem but it is not central in our discussion [10, 12, 18].

5.1. Multicriteria and multiphysics data fusion. Our interest is on data assimilation/fusion and m represents the different data set to be assimilated. In this case, J_i reads for instance as:

$$(23) \quad J_i(u_i(x), u_i^*) = \|u_i(x) - u_i^*\|_{\Sigma_i^{-1}}^2 + T_i(x), \quad \text{with } T_i(x) \geq 0.$$

$T_i(x)$ represents regularity control requirements or additional constraint. The norm $\|\cdot\|_{\Sigma_i^{-1}}$ uses the inverse of the covariance matrix Σ_i of the different target observations u_i^* .

The problem is multi-physics as getting the functional J_i necessitates the solution of a different state equations $F_i(u_i(x)) = 0$ for the description of the different state variable $u_i(x) \in \mathbb{R}^M$. One expects M to be very large. For instance, in the case of a distributed optimization problem involving the solution of PDE models, M typically represents the size of the mesh.

5.2. An example of vector and scalar fields assimilation. Consider the general problem of mixing a passive scalar (here the temperature) in a fluid. The question is to find the flow field which permits to realize a given distribution for a transported quantity with constraint on what the admissible field is. Inversely, one wants to answer the question of source identification from a few flow field and advected quantity observations. The state equations we consider are the incompressible Navier-Stokes equations together and an advection and diffusion equation for the temperature. Details on these equations and the flow solver can be found in [19]. The flow field can be controlled by various means as, for instance, injection devices or variations of a magnetic field. The former is taken into account through a modification of the boundary conditions for the velocity and the later by the introduction of a source term in the flow equations. Here we consider a vector x of 12 control parameters with lower and upper bounds on each component. The bounds express the admissibility domain for flow directions and maximum possible velocity amplitude. Figure 4 shows an example of flow velocity intensity and temperature distributions for a centrifugal device in which control of such quantities might of interest for suitable functioning.

The two cost functions we consider are non dimensionalized and are of the form:

$$J_1(x) = \frac{1}{|\Omega| \overline{T}_{des}^2} \int_{\Omega} (T(x) - T_{des})^2, \quad J_2(x) = \frac{1}{|\Omega| \|\overline{\vec{u}}_{des}\|^2} \int_{\Omega} \|\vec{u}(x) - \vec{u}_{des}\|^2,$$

where upper bars indicate mean values over the domain. To simplify the presentation we have assumed identity covariance matrices in (23). These aim at realizing target flow velocity and temperature distributions in the domain or some part of it.

A sampling of the control space to generate the domain of possible $(J_1(x), J_2(x)) \in \mathbb{R}^2$ is too costly even if the use of sparse grids [6, 23] or other intelligent sampling techniques can bring some relief. Indeed, sampling an admissible domain in dimension n would require 2^n evaluations of the functional to have just the bounds for the variations of the cost function due to each of the parameters. Here, with only four points over each of the variables range we need over than 16 millions functional

evaluations. In this problem we obviously have no idea of how the Pareto front is and we can only guess the front from the simulations. Therefore, what we would like to mainly see through this example is the exploration capability of optimization methods based on discrete forms of either first (20) or second order (19) dynamical systems together with the repulsion forces given in section 4.2.

Figures 5 and 6 show visited $(J_1(x), J_2(x))$ points by first and second order dynamics for a same set of optimization problems corresponding to a 20 points uniform sampling of $\alpha \in [0, 1]$ and the weighted functional $\alpha J_1 + (1 - \alpha) J_2$ as described in the previous academic example. However, this problem is much more computer intensive.

One can see that second order dynamics have higher exploration capabilities. The functional J_1 and J_2 are apparently non-convex as the first order dynamics do not necessarily reach points on the non-convex part of the front as the trajectories are probably captured by local minima, away from the front. We can also see that it is interesting to combine the information from the first and second order constructions. Indeed, some points on the Pareto front (highlighted by the circle in figure 5) have been identified by the first order dynamics while missed by the second order construction. We do not have an explanation for this behavior yet. Combining the two sets of information, we have drawn an estimation of the Pareto front which we think is non-convex and also non-connected.

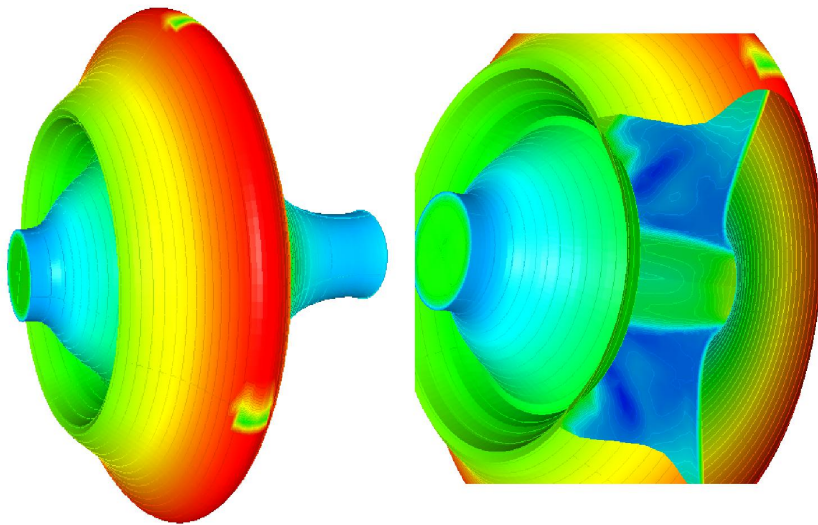


FIGURE 4. Typical flow field $\|\vec{u}\|$ distribution and temperature T iso-contours for a set of control parameters x . Two injection locations can be seen on the external wall.

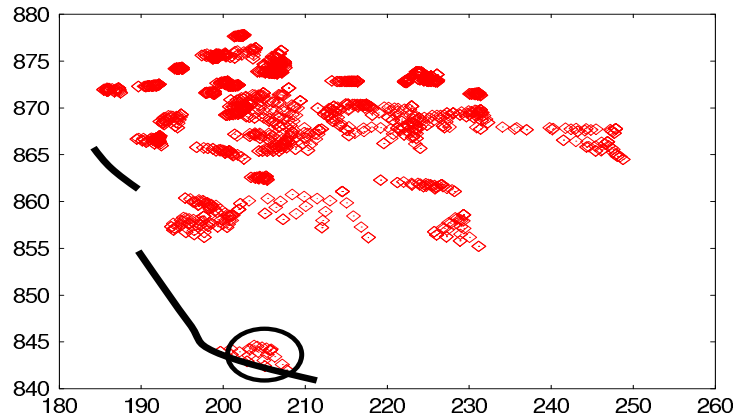


FIGURE 5. Visited $(J_1(x), J_2(x))$ points by first order dynamics together with repulsion forces and estimation (continuous line) of the Pareto front combining information from these simulations and those from the second order dynamics shown in figure 6. The first order dynamics has difficulty in reaching the Pareto front as most of the trials seem to be trapped by local minima.

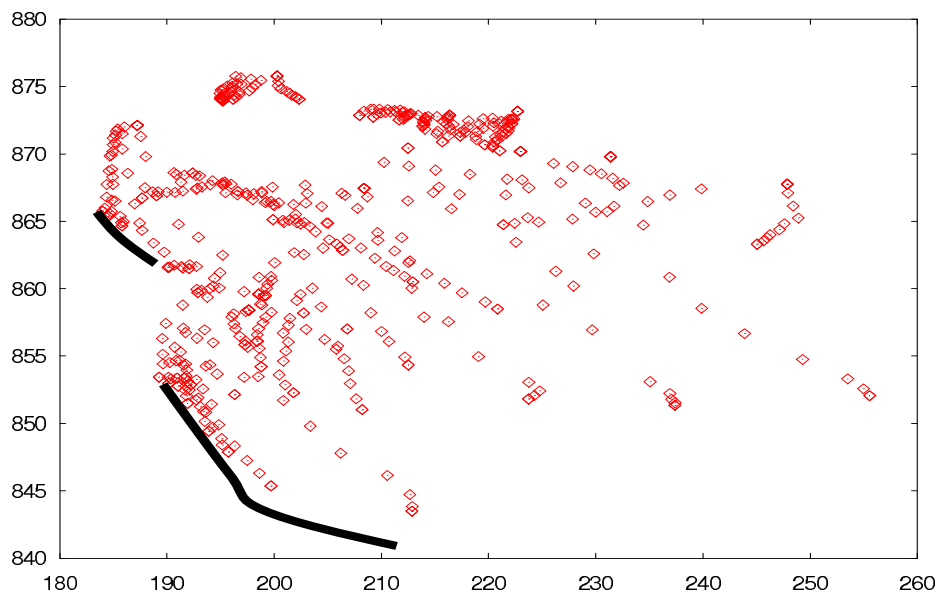


FIGURE 6. Visited $(J_1(x), J_2(x))$ points by second order dynamics together with repulsion forces and estimation (continuous line) of the Pareto front combining information from first and second orders dynamics.

6. CONCLUDING REMARKS

Different formulations of multicriteria optimization problems have been presented together with a class of metaheuristic optimization algorithms intuited from theoretical behavior of continuous dynamical systems. In that purpose, global optimization has been formulated as solution of boundary value problems for second order

systems. Two theoretical results have been given for the continuous systems on sub-optimal partial controllability in finite time and on the conditions under which a Pareto front can be accessed by descent methods applied to multicriteria problems using the weighted sum approach. It has been shown that global search features are suitable for the identification of points on general Pareto fronts. The behavior of minimization algorithms based on first and second order dynamical systems have been discussed on an academic example and a distributed data fusion problem in large dimension. For this latter case a sampling of the parameter space is clearly impossible. Both situations feature non-convex and non-connected fronts. It has been shown that minimization algorithms having inertial features permit to reach non-convex regions in Pareto fronts thanks to their higher global search capabilities. The question of the distribution and localization of the points on the Pareto front has also been discussed to see how to monitor the distribution of the points on the front. The different trials have been coupled introducing a particular repulsion force in the corresponding second order dynamical systems. However, to address this issue, the concepts of the paper can also be applied together with the normal boundary intersection [8] or the normal constraint [1] method.

REFERENCES

- [1] A. Ismail-Yahaya A. Messac and C.A. Mattson. The normalized normal constraint method for generating the pareto frontier. *Struct Multidisc Optim.*, 25:86–98, 2003.
- [2] R. Tappeta A. Messac, G. J. Sundararaj and J. E. Renaud. Ability of objective functions to generate points on non-convex pareto frontiers. *AIAA J.*, 38/6:915–937, 2000.
- [3] B. Mohammadi B. Ivorra and A. Ramos. semi-deterministic global optimization method and application to the control of burgers equation. *JOTA*, 135/1:415–427, 2007.
- [4] B. Mohammadi B. Ivorra and A. Manuel Ramos. A multi-layer line search method to improve the initialization of optimization algorithms. *EJOR European Journal of Operational Research*, 247:711–720, 2015.
- [5] H.-G. Bock and K.J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings of the 9th IFAC World Congress*, pages 22–31. Budapest Univ., 1984.
- [6] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.
- [7] I. Das and J. E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Struct Optim.*, 14:40–65, 1997.
- [8] I. Das and J. E. Dennis. Normal-boundary intersection: A new method for generating pareto optimal points in multicriteria optimization problems. *SIAM J Optim.*, 8:80–95, 1998.
- [9] K. Deb. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *IEEE J. Evolutionary Comp.*, 7:20–45, 1999.
- [10] J.-A. Desideri. Cooperation and competition in multidisciplinary optimization. *Computational Optimization and Applications*, 51/1:29–68, 2011.
- [11] M. Ehrgott. *Multicriteria Optimization*. ISBN 978-3-540-21398-7. Birkhuser, Berlin, 2012.
- [12] S. Garreau Ph. Guillaume and M. Masmoudi. The topological asymptotic for pde systems: The elasticity. *SIAM J. Control Optim.*, 39/6:1756–1778, 2001.
- [13] G. Michaille H. Attouch, G. Buttazzo. *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*. SIAM series on Optimization. NY, 2005.
- [14] B. Ivorra and B. Mohammadi. Optimization strategies in credit portfolio management. *Journal of Global Optimization*, 43/2:415–427, 2009.
- [15] Y. Maday J.-L. Lions and G. Turinici. A parareal in time discretization of pde’s. *C.R. Acad. Sci. Paris, Serie I.*, 332:1–8, 2001.
- [16] J. Jahn. *Vector Optimization: Theory, Applications and Extensions*. Springer, Berlin, 2004.
- [17] R.T. Marler and J.S. Arora. Survey of multi-objective optimization methods for engineering. *Struct Multidisc Optim.*, 26:369–395, 2004.

- [18] J. R. R. A. Martins and A. B. Lambe. Multidisciplinary design optimization: A survey of architectures. *AIAA J.*, 51(9):2049–2075, 2013.
- [19] G. Medic and B. Mohammadi. Nsike unstructured solver for laminar and turbulent incompressible flows simulation. In *INRIA Technical report 3644*, pages 1–31. INRIA, 1998.
- [20] B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids (2nd Edition)*. Oxford Univ. Press, Oxford, 2009.
- [21] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, NY, 2006.
- [22] V. Pareto. *Manuale di Economia Politica, Societa Editrice Libreria, Milano, Italy. Translated into English by Schwier, A.S. 1971: Manual of Political Economy*. Macmillan, NY, 1971.
- [23] SA. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR* **148**: 1042-1043. *Russian, Engl. Transl.: Soviet Math. Dokl.*, 4:240–243, 1963.
- [24] W. Stadler. A survey of multicriteria optimization, or the vector maximum problem. *JOTA*, 29:115–127, 1979.