Comparison between MGDA and PAES for Multi-Objective Optimization

Adrien Zerbinati, Jean-Antoine Desideri, Régis Duvigneau

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Comparison between MGDA and PAES for Multi-Objective Optimization

Adrien Zerbinati — Jean-Antoine Désidéri — Régis Duvigneau

N° 7667

Juin 2011

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7667--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

centre de recherche
SOPHIA ANTIPOLIS - MÉDITERRANÉE

Comparison between MGDA and PAES for Multi-Objective Optimization

Adrien Zerbinati , Jean-Antoine Désidéri, Régis Duvigneau

Theme :
Équipes-Projets Opale

Rapport de recherche n° 7667 — Juin 2011 — 12 pages

Abstract: In multi-objective optimization, the knowledge of the Pareto set provides valuable information on the reachable optimal performance. A number of evolutionary strategies (PAES [4], NSGA-II [1], etc), have been proposed in the literature and proved to be successful to identify the Pareto set. However, these derivative-free algorithms are very demanding in terms of computational time. Today, in many areas of computational sciences, codes are developed that include the calculation of the gradient, cautiously validated and calibrated. Thus, an alternate method applicable when the gradients are known is introduced here. Using a clever combination of the gradients, a descent direction common to all criteria is identified. As a natural outcome, the Multiple Gradient Descent Algorithm (MGDA) is defined as a generalization of steepest-descent method and compared with PAES by numerical experiments.

Key-words: Optimization, gradient descent, Pareto optimality, Pareto front, performances

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Comparaison des algorithmes MGDA et PAES en optimisation multiobjectif

Résumé : Dans le cadre d'une étude d'optimisation multiobjectif, la connaissance du front de Pareto permet de cerner efficacement le champ de recherche des paramètres optimaux. Pour ce faire, des algorithmes basés sur des méthodes évolutionnaires ont été développés (PAES [4], NSGA-II [1], etc). Nous proposons ici un algorithme alternatif, basé sur l'utilisation des gradients de critères permettant d'obtenir un échantillon du front de Pareto. Nous commençons par montrer qu'une combinaison judicieuse de ces gradients est une direction de descente commune à tous les critères.

Mots-clés : Optimisation, gradient de descente, Pareto optimalitée, front de Pareto

Contents

1 Introduction

The numerical treatment of a multi-objective minimization is usually aimed to identify the Pareto set. In the literature, several authors have proposed to achieve this goal by various algorithms, each one adapting a particular Evolution Strategy (ES). Such approaches are compared in the book of Deb [3]. Using a sufficiently diverse initial sample, these methods produce a discrete set of 2 by 2 non-dominated points. However, they are very demanding in terms of computational time, as ES do in general.

In the particular case in which the gradients of the objective functions are at reach, at the current design point, faster algorithms can be developed. In the convex hull of the gradients of objective functions, a direction exists along which all criteria diminish. The MGDA results in utilizing this direction as search direction and optimizing the stepsize appropriately. In this way, the classical steepest-descent method is generalized to multi-objective optimization. Applying MGDA thus corresponds to a phase of *cooperative optimization*.

In section 2, theoretical aspects leading to MGDA are briefly recalled. A complete presentation is available in [2]. In section 3, results of a numerical experimentation on a classical test case are presented and commented.

2 Theoretical aspects

2.1 Cooperative-optimization phase : Multiple-Gradient Descent Algorithm (MGDA)

Here, to be complete, we review briefly the notions developed in [2]. The general context is the simultaneous minimization of n ($n \in \mathbb{N}$) smooth criteria (or disciplines) $J_i(Y)$ (Y : design vector, $Y \in \mathbb{R}^N$). Starting from an initial design point that is not Pareto optimal, a cooperative optimization phase is defined that is beneficial to all criteria.

2.1.1 Pareto concepts

Following [2], we introduce the notion of *Pareto stationarity*: a design point Y^0 is said to be Pareto stationary if there exists a convex combination of the gradients of the smooth criteria J_i that is equal to 0 at this point. Thus :

Definition 2.1. *The smooth criteria $J_i(Y)$ $(1 \leq n \leq N)$ are said to be Pareto stationary at the design point Y^0 if:*

- $\forall i = 1, .., n, \quad u_i^0 = \nabla J_i\left(Y^0\right)$ *;*
- $\exists \left(\alpha_i\right)_{i=1,..,n}, \alpha_i \geq 0, \sum_{i=0}^{n} \alpha_i = 1, \quad \sum_{i=0}^{n} \alpha_i u_i^0 = 0.$

Inversely, if the smooth criteria $J_i(Y)$ $(1 \leq i \leq n)$ are not Pareto-stationary at the given design point Y^0, a descent direction *common to all criteria* exists.

2.1.2 Existence and uniqueness of the minimal-norm element

Consider a family of vectors, denoted $(u_i)_{i \in I}, I \subset \mathbb{N}$. The following lemma holds :

Lemma 2.1 (Existence and uniqueness of the minimal-norm element)**.** *Assume :*

- $\{u_i\}$ *$(1 \leq i \leq n)$ a family of n vectors in $\mathbb{R}^N$;*
- *$\mathcal{U}$ be the set of strict convex combinations of these vectors :*

$$\mathcal{U} = \left\{ w \in \mathbb{R}^n / w = \sum_{i=0}^{n} \alpha_i u_i^0 \, ; \, \alpha_i > 0, \forall i \, ; \, \sum_{i=0}^{n} \alpha_i = 1 \right\}.$$

Then,

$$\exists! \omega \in \overline{\mathcal{U}}, \quad \forall \bar{u} \in \overline{\mathcal{U}} \quad : \quad (\bar{u}, \omega) \geq (\omega, \omega) = \|\omega\|^2.$$

(The element ω exists since $\overline{\mathcal{U}}$ is closed, and it is unique since $\overline{\mathcal{U}}$ is convex; as a result, $\forall \bar{u} \in \overline{\mathcal{U}}$, and $\forall \epsilon \in [0,1]$, $\omega + \epsilon(\overline{u} - \omega) \in \overline{\mathcal{U}}$, and $\|\omega + \epsilon(\overline{u} - \omega)\| \geq \|\omega\|$, and this yields the conclusion [2]).

In the case of two criteria, three configurations of the two gradients can be considered, as illustrated below:

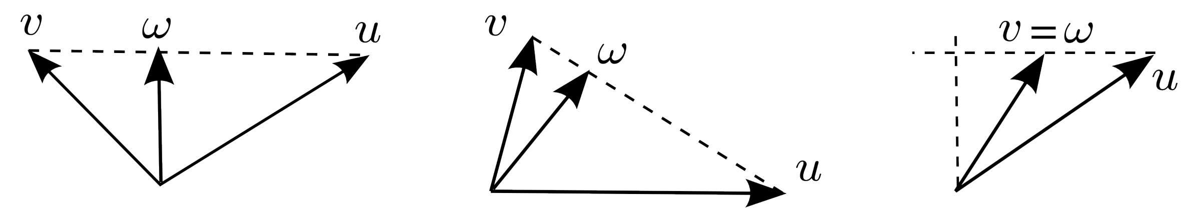

Figure 1: Various possible configurations of the two gradient-vectors u and v and the minimal-norm element ω.

This result applies in particular to u_i for all i. But, (u_i, ω) is the Frechet-derivative of J_i in the direction ω. Hence, if $\omega \neq 0$, the Frechet-derivatives of all the criteria are bounded from below by the strictly positive number $\|\omega\|^2$. The direction $-\omega$ is therefore a descent direction common to all criteria. These considerations yield the following:

Theorem 2.1. *Let $J_i(Y)$ ($1 \leq i \leq n \leq N$, $N \in \mathbb{N}$) be n smooth functions of the vector $Y \in \mathbb{R}^N$. Assume Y^0 is an admissible design-point. We denote $u_i^0 = \nabla J_i(Y^0)$ and :*

$$\mathcal{U} = \left\{ w \in \mathbb{R}^N, \quad w = \sum_{i=1}^{n} \alpha_i u_i^0; \forall i, \alpha_i > 0; \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{1}$$

Let ω be the minimal-norm element of the convex hull $\overline{\mathcal{U}}$, closure of $\mathcal{U}$. Then :

1. *Either $\omega = 0$, and the criteria $J_i(Y)$ ($1 \leq i \leq n$) are Pareto-stationary ;*
2. *Or $\omega \neq 0$ and $-\omega$ is a descent direction common to all the criteria; additionally, if $\omega \in \mathcal{U}$, the inner product $(\bar{u}, \omega)$ is equal to $\|\omega\|^2$ for all $\bar{u} \in \overline{\mathcal{U}}$.*

Based on these results, when the gradients of all the criteria can be computed, the following algorithm (MGDA) proceeds by successive steps that are beneficial to all criteria. In the practical implementation, one specifies a tolerance ϵ_{TOL} on $\|\omega\|$ below which the linesearch is not performed.

Algorithm 1 MGDA

Initialisation: $Y := Y^0$
Loop (WHILE $\|\omega\| \geq \varepsilon$)

- Evaluate $J_i(Y)$, $(1 \leq i \leq n)$
- Compute $\nabla J_i(Y)$ $(1 \leq i \leq n)$;
- Identify ω, as the mininal-norm element in the convex hull of $\{\nabla J_i(Y)\}$ $(1 \leq i \leq n)$
- Linesearch : determine optimal l ;
- Update design vector $Y := Y - l\omega$.

2.2 Convergence of the MGDA

Provided that the criteria are formulated to be smooth, positive and infinite at infinity, the sequence of iterates produced by the MGDA has been proved to admit a subsequence converging to a Pareto-optimal point [2]. The main purpose of this report is to illustrate this convergence by numerical experiments using testcases of variable complexity.

2.3 Practical determination of the vector ω

In the general case ($n > 2$), ω can be calculated by numerical minimization of the quadratic form that expresses $\|\omega\|^2$ in terms of the coefficients $\{\alpha_i\}$ of the convex combination, subject to the inequality constraints $\alpha_i \geq 0$ ($\forall i$), and the linear equality constraint $\sum_i \alpha_i = 1$. Many routines are effective to perform this optimization, for instance certain evolution strategies. However, the problem may become ill-conditioned for large dimensions.

However, in the particular case of two objectives, ω can be expressed explicitly. Recall Figure 1, for which $u = \nabla J_1$ and $v = \nabla J_2$. In this figure, the gradient vectors, elements of $\mathbb{R}^N$ are represented as vectors of $\mathbb{R}^2$ with same origin O. This results in no loss of generality since only the norms of the two vectors, and the angle between them do matter. Eliminating the trivial case in which $u = v$ (for which $\omega = u = v$), the convex hull is then represented by the segment uv connecting the extremities of these representative vectors. Let $\omega^\perp$ be the vector whose origin is O, and extremity is the orthogonal projection of O onto the line that supports the segment uv (convex-hull). If the vector $\omega^\perp$ is in the convex hull, that is, if its representative points on the segment uv, it is ω; otherwise, ω is the vector of smallest norm between u and v. Thus let:

$$\omega = (1-\alpha)u + \alpha v \tag{2}$$

and compute $\alpha^\perp$ for which the above convex combination is orthogonal to $u - v$, that is :

$$\alpha^\perp = \frac{(u, u-v)}{(u-v, u-v)}$$

If $\alpha^\perp \in [0, 1]$, $\alpha = \alpha^\perp$; otherwise, $\alpha = 0$ or 1, that is, $\omega = u$ or v, depending on whether $\alpha^\perp < 0$ or > 1.

2.4 Line-search

This part deals with the determination of the step length (line-search). In multi criterion otimization, it is not easy to compute a step giving satisfaction to all criteria and a significant evolution. An adaptative method to compute a satisfying step for each multiobjective problem would be convenient.

At the current design point, the Frechet-derivatives of all the criteria are strictly negative (and equal if $\omega \in \mathcal{U}$). For each criterion, a surrogate quadratic model is constructed after computing three function values, and a related optimum stepsize ρ_i is calculated corresponding to the location of the surrogate model's minimum (see Figure 2).
Now, we choose the global step ρ as the smallest ρ_i :

$$\rho = \min_{i, 1 \leq i \leq n} \rho_i$$

Thanks to the definition of ω, $\rho_i \geq 0$ and $\rho \geq 0$.

3 Numerical experimentation

In this section, we conduct numerical experiments to demonstrate the convergence of MGDA to Pareto optimal solutions, and to compare this algorithm with PAES [4] in two analytical testcases of two-objective optimization corresponding to a convex and a concave Pareto fronts.

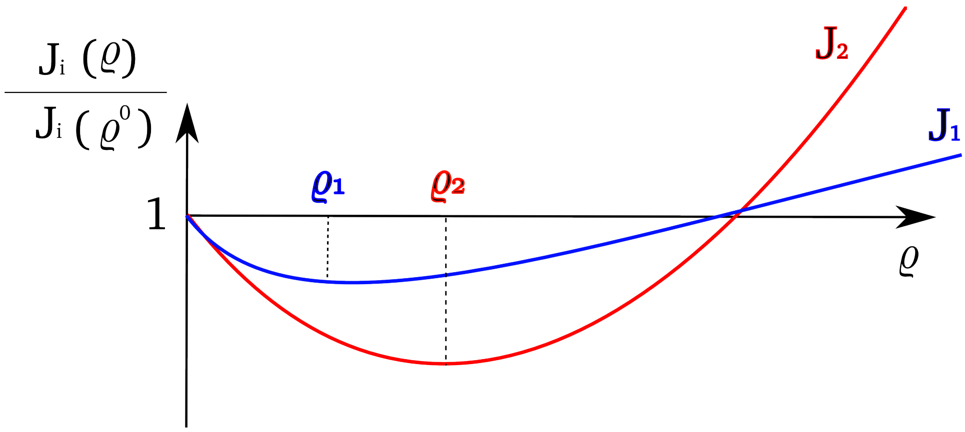

Figure 2: Variation of *normalized* the cost functions with the stepsize ρ in $-\omega$ direction.

3.1 Analytical test case

In this testcase, two functions from $\mathbb{R}^2 \to \mathbb{R}$, denoted $f(x,y)$ and $g(x,y)$, are defined analytically by :

- $f(x,y) = 4x^2 + y^2 + xy$;
- $g(x,y) = (x-1)^2 + 3(y-1)^2$.

Figure 3 illustrates the pattern of their isovalue contours. Each function has a visible distinctive minimum. The curve connecting the locations of these minima is a representation of the Pareto-optimal solutions of the two-objective unconstrained minimization problem of these functions, here easily calculated analytically.

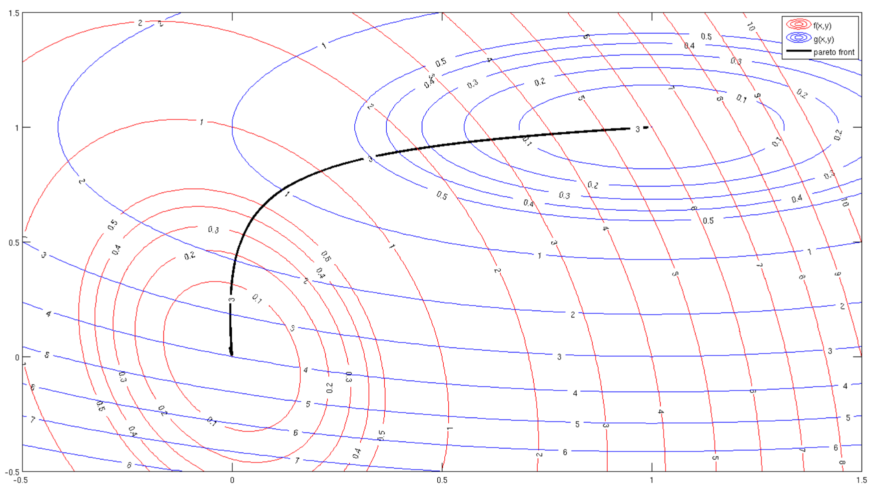

Figure 3: Isolines of f and g with Pareto set of the optimization problem, in design space (x, y).

Figure 4 illustrates the step-by-step convergence of the MGDA to a Pareto-optimal point, in three different cases corresponding to three different initial points. On the left of the figure, the convergence path is indicated in the $\mathbb{R}^2$ search space, and on the right in the $\mathbb{R}^2$ function space. Each segment of the dashed jagged line corresponds to one MGDA iteration. Depending on the initial point, the convergence can be to the minimum of f, the minimum of g, or to an intermediate point of the Pareto-set. Evidently, the Pareto set (in function space) is here continuous and convex, which corresponds to a favorable situation.

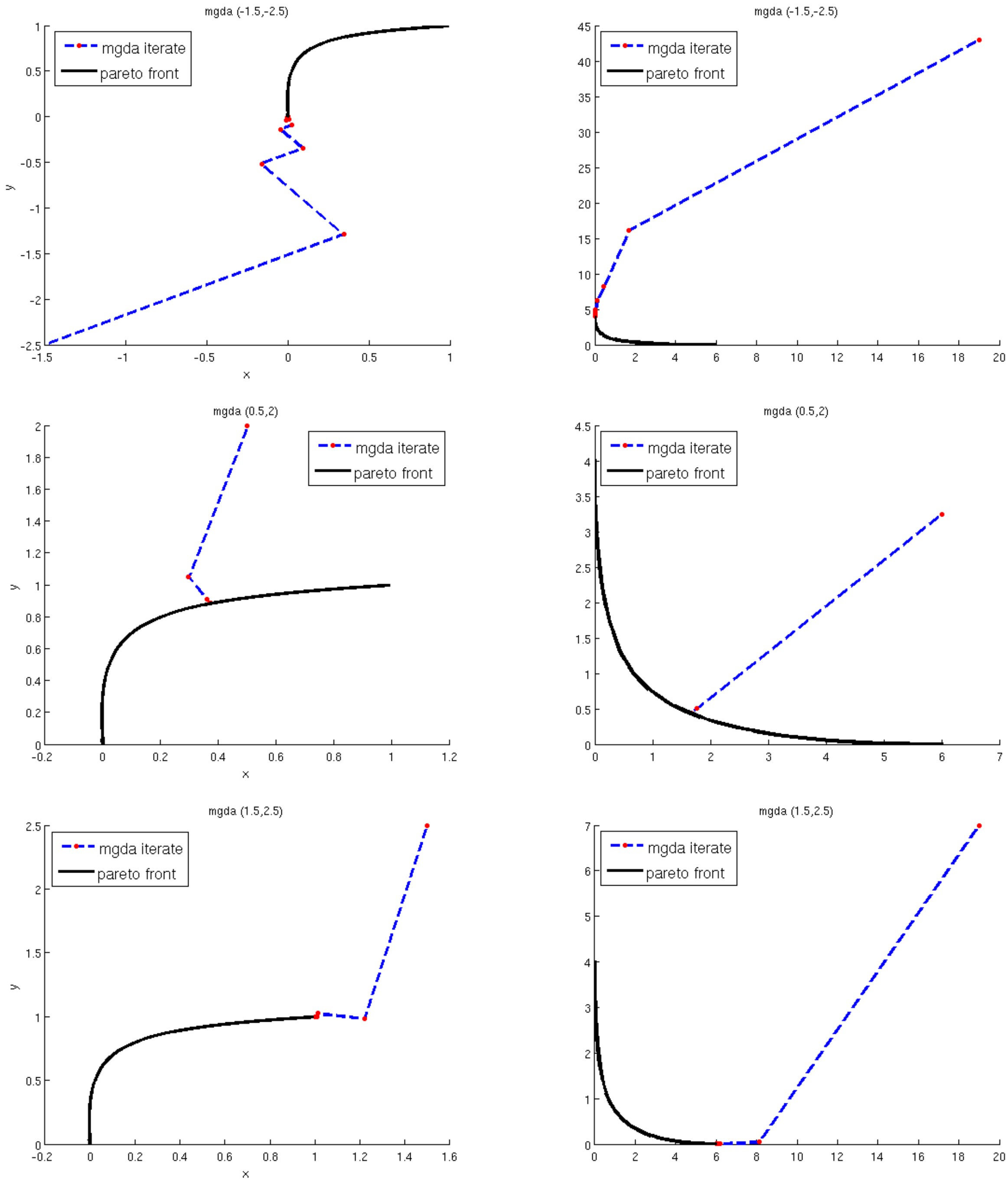

Figure 4: Convergence to the Pareto set and front, using MGDA for several initial design points. In design space (x, y) on the left, in functional space (f, g) on the right.

Convergence has been verified by browsing the design space thoroughly by considering a sample of initial design points located on a circle whose interior includes the Pareto set in design space (see Figure 5.

As a result, the achieved discrete representation of the Pareto set is very accurate. It is compared on Figure 6 to the analytical set.

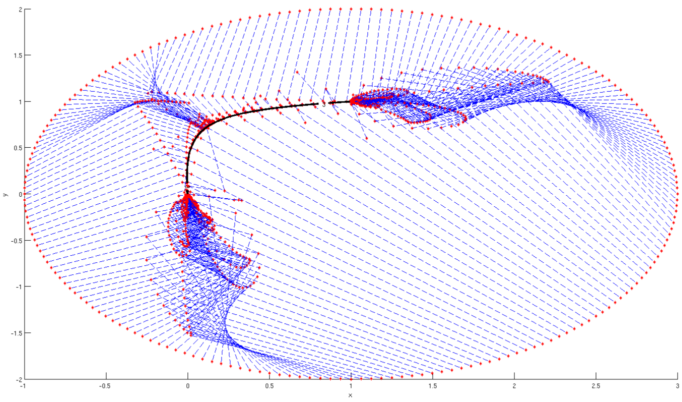

Figure 5: Convergenge to the Pareto set using MGDA from a set of initial design sample points.

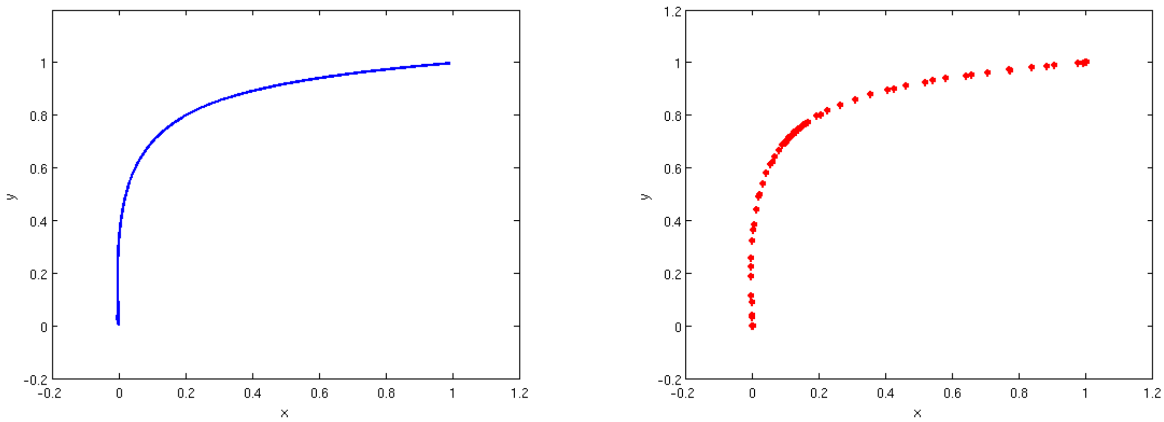

Figure 6: Pareto set in design space : analytical determination (left), and discrete results by convergence of MGDA (right)

A similar experiment has been conducted using PAES [4] instead of MGDA. In this experiment the population size is 100, and 100 generations have been computed. The resulting approximate determination of the Pareto set is indicated on Figure 7. Evidently, in this simple testcase of continuous and convex Pareto set, in which the gradients are available and smooth, the gradient-based method is far superior in both computing effort, and accuracy.

3.2 Fonseca test case

This testcase corresponds to the two-objective unconstrained minimization of the functions

$$f_1(x) = 1 - \exp\left(-\sum_{i=1}^{3}\left(x_i - \frac{1}{\sqrt{3}}\right)^2\right) \quad f_2(x) = 1 - \exp\left(-\sum_{i=1}^{3}\left(x_i + \frac{1}{\sqrt{3}}\right)^2\right)$$

of the design variable $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. This testcase is known to yield a concave Pareto set in function space, a usually more straining situation for numerical optimizers than previously.

As before, we first illustrate a few iterations of MGDA in two cases differing by the intial design point. The algorithm yields non-dominated design points. Only a few iterations are sufficient (see Figure 8).

Here, the Pareto set is not known analytically, but has been well identified by Deb using the well-known genetic algorithm NSGA-II [3]. To obtain an accurate discrete representation of the Pareto set

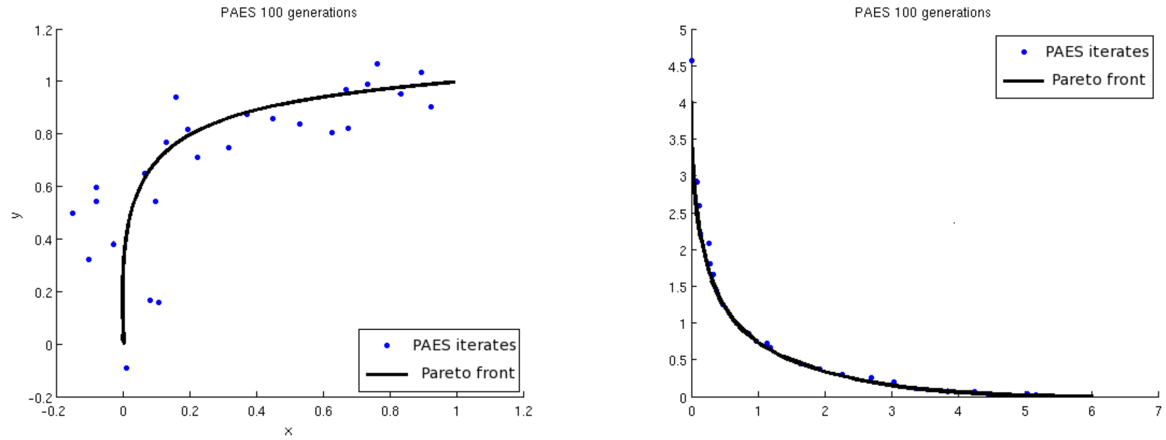

Figure 7: Pareto set in design space (left) and in function space (right); the solid lines correspond to the analytical determination, and the discrete points are given by application of PAES

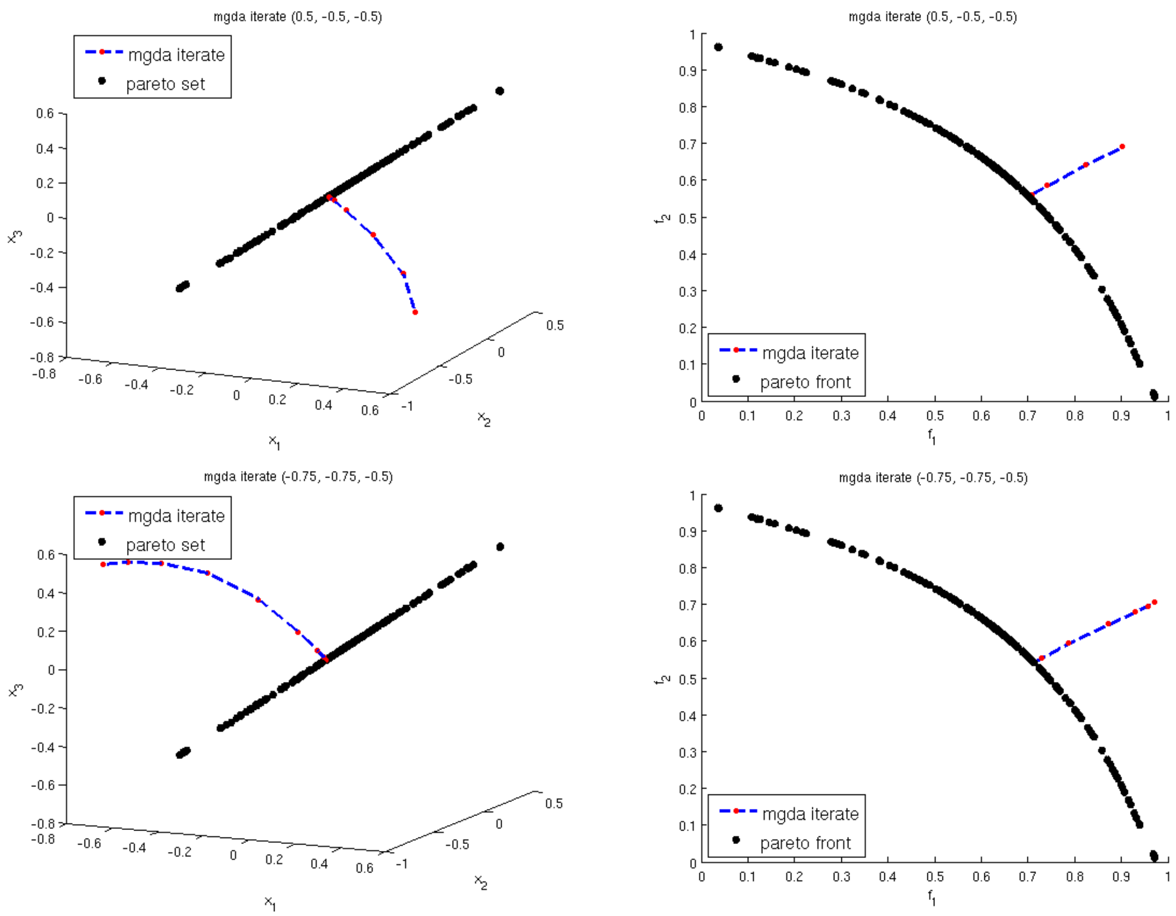

Figure 8: Convergence of MGDA to the Pareto front, for several initial design points., in design space (x, y, z) (left) and in function space (f_1, f_2) (right)

by MGDA, we have applied the method starting from a large set of initial design points located on a sphere in the design-space (Figure 9).

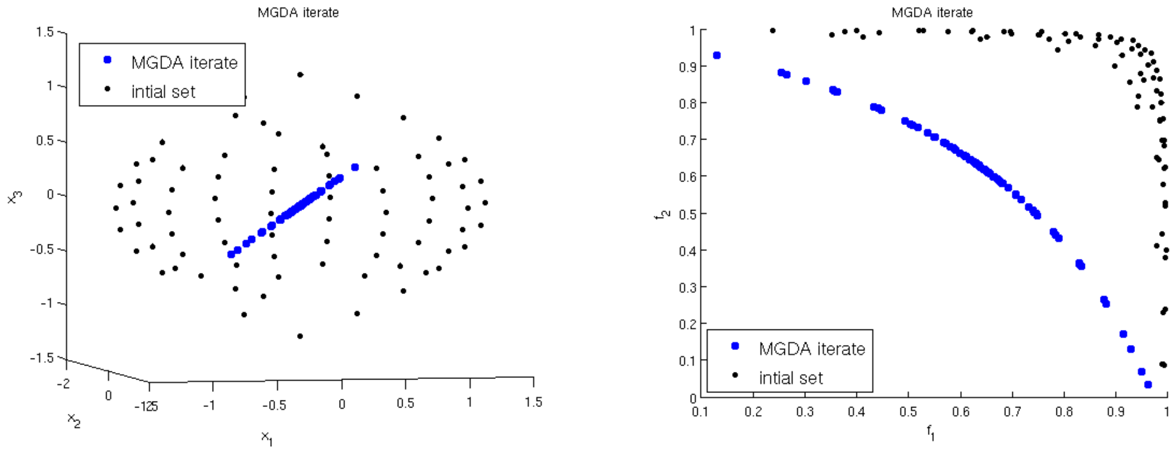

Figure 9: Case of a nonconvex Pareto set: convergence of the MGDA from a set of initial design points: in the space of criteria (left), and in the design space (right).

In the next experiment, we have first applied PAES twice, each time starting from a different design point and generating 50 others. Then the remaining dominated design points have been discarded. Thus less than one hundred design points have been archived. This set is compared on Figure 10 with the result of applying MGDA starting from 12 well-distributed initial design points, so that the number of function evaluations is the same in the two cases. MGDA again produces design points closer to the Pareto set (improved accuracy), but in fewer number.

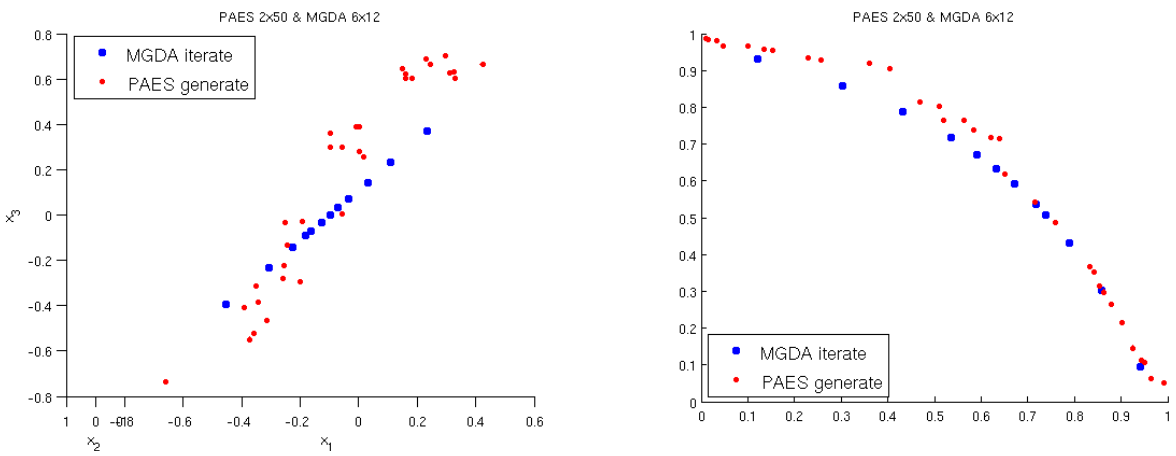

Figure 10: Pareto set approximated discretely by PAES and MGDA

As illustrated by Figure 10, when the set of initial points is chosen adequately, MGDA gives a better representation of the Pareto front than PAES. However, at identical computational cost, generally, PAES introduce more variety in the final result. Thus it appears interesting to combine the accuracy of MGDA with the robustness of PAES in a hybrid method. To check this, we have used the two methods sequentially: PAES first to generate 15 design points, retaining 8 nondominated design points then used as intial points for MGDA. In each case about 3 to 4 iterations are sufficient to converge and produce the accurate result indicated on Figure 11.

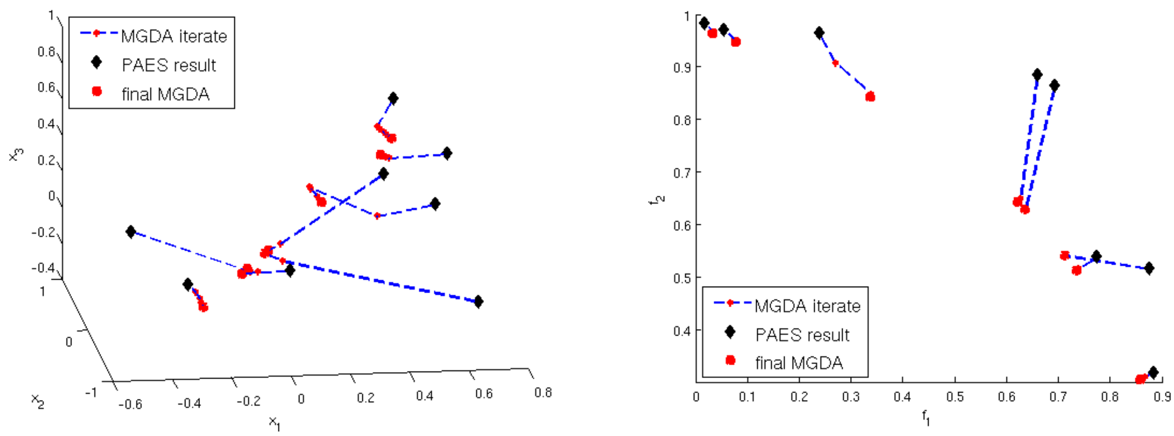

Figure 11: First step with a large PAES followed by MGDA iterates on each non dominated point found. Design space on the left, functional space on the right.

4 Conclusion

In this report, we have tested by numerical experiment a recently proposed gradient-based algorithm, MGDA [2], for mutiobjective optimization. The convergence to Pareto-optimal solutions has been demonstrated in two analytical testcases corresponding to a convex and a concave Pareto fronts. MGDA has been compared to PAES, and found to have complementary merits, and a hybrid method is promising.

References

[1] Pratap A. Agarwal S. Deb, K. and T Meyarivan. *A Fast and Elitist Multi-Objective Genetic Algorithm-NSGA-II.* KanGAL Report Number 2000001, 2000.

[2] Jean-Antoine Désidéri. Multiple-Gradient Descent Algorithm (MGDA). Research Report 6953, INRIA, 2009.

[3] Sameer Agarwal T. Meyrivan Kalyanmoy Deb, Amrit Pratap. *Transaction on evolutionary computation, vol 6, n 2.* IEEE, 2002.

[4] Corne D.W. Knowles, J.D. *Approximating the nondominated front using the Pareto Archived Evolution Strategy. Evolutionary Computation.* MIT press, 2000.

Centre de recherche INRIA Sophia Antipolis – Méditerranée
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399