HAL
open science

# Comparison between two multi objective optimization algorithms : PAES and MGDA. Testing MGDA on Kriging metamodels

Adrien Zerbinati, Jean-Antoine Désidéri, Régis Duvigneau

## To cite this version:

Adrien Zerbinati, Jean-Antoine Désidéri, Régis Duvigneau. Comparison between two multi objective optimization algorithms : PAES and MGDA. Testing MGDA on Kriging metamodels. Repin, S. and Tiihonen, T. and Tuovinen, T. Numerical Methods for Differential Equations, Optimization, and Technological Problems, 27, Springer Dordrecht, pp.237-252, 2013, Computational Methods in Applied Sciences, 978-94-007-5288-7. 10.1007/978-94-007-5288-7 . hal-00765314

**HAL Id: hal-00765314**

**https://inria.hal.science/hal-00765314v1**

Submitted on 14 Dec 2012

**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.

# Comparison between two multi objective optimization algorithms : PAES and MGDA. Testing MGDA on Kriging metamodels

Adrien Zerbinati, Jean-Antoine Désidéri and Régis Duvigneau

**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 [3], 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 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 presently. 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 the steepest-descent method and compared with PAES by numerical experiments. Using MGDA on a multi objective optimization problem requires the evaluation of a large number of points with regard to criteria, and their gradients. In the particular case of CFD problems, each point evaluation is very costly. Thus here we also propose to construct metamodels and to calculate approximate gradients by local finite differences.

## 1 Introduction

The numerical treatment of a multi-objective minimization is usually aimed to identify the Pareto set or a convenient subset of it. In the literature, several authors have proposed to achieve this goal by various algorithms, each one adapting a particular

Adrien Zerbinati
INRIA, 2004 route des lucioles 06902 Sophia Antipolis, e-mail: adrien.zerbinati@inria.fr

Jean-Antoine Désidéri
INRIA, 2004 route des lucioles 06902 Sophia Antipolis e-mail: jean-antoine.desideri@inria.fr

Régis Duvigneau
INRIA, 2004 route des lucioles 06902 Sophia Antipolis e-mail: regis.duvigneau@inria.fr

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, the most commonly used methods 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 the objective functions, a direction exists along which all criteria diminish [1]. 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 [1]. 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 [1]. 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 [1], 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 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:

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

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 $(u_i)_{i\in I}, 1 \leq i \leq n$. The following lemma holds :

**Lemma 1 (Existence and uniqueness of the minimal-norm element).** *Assume :*

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

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

*and* $\overline{\mathscr{U}}$ *its closure, or convex hull.*

*Then,*

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

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

In the case of two criteria, three configurations of the two gradients can be considered, as illustrated below 1:

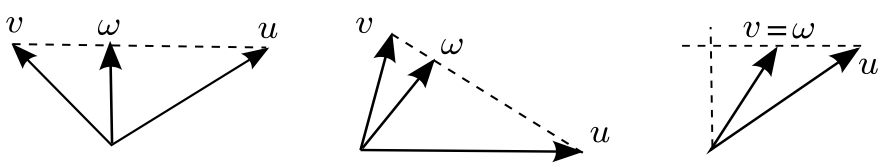



**Fig. 1** Various possible configurations of the two gradient-vectors $u = u_1$ and $v = u_2$ and the minimal-norm element $\omega$.

This result applies in particular to $u_i$ for all $i$. But, $(u_i, \omega)$ is the Frechet-derivative of $J_i$ in the direction $\omega$. 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 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 :*

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

*Let* $\omega$ *be the minimal-norm element of the convex hull* $\overline{\mathscr{U}}$*, closure of* $\mathscr{U}$*. 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 $\omega \in \mathscr{U}$, the inner product $(\bar{u}, \omega)$ is equal to $\|\omega\|^2$ for all $\bar{u} \in \overline{\mathscr{U}}$.*

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 $\varepsilon_{TOL}$ on $\|\omega\|$ below which the linesearch is not performed.

## 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 [1]. One 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 $\omega$

In the general case ($n > 2$), $\omega$ 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, $\omega$ can be expressed explicitly. Recall Figure 1, for which $u = u_1 = \nabla J_1$ and $v = u_2 = \nabla J_2$. 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 $\omega$; otherwise, $\omega$ 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 optimization, it is not easy to compute a satisfactory step with respect to all criteria producing a significant evolution. An adaptative method to compute a satisfactory 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 $\omega \in \mathscr{U}$). For each criterion, a surrogate quadratic model is constructed after computing three function values, and a related optimum stepsize $\rho_i$ is calculated corresponding to the location of the ith surrogate model's minimum.

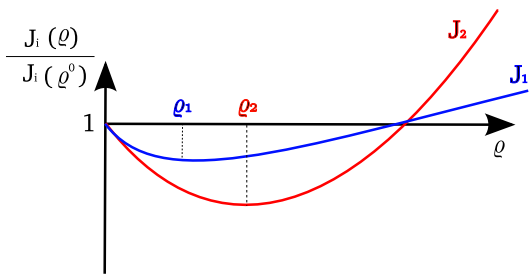



**Fig. 2** Variation of *normalized* the cost functions with the stepsize $\rho$ in $-\omega$ direction.

Now, we choose the global step $\rho$ as the smallest $\rho_i$ :

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

The vector $\omega$ is such that, $\forall i$, $\rho_i \geq 0$ and $\rho \geq 0$. Whenever $\rho = 0$, MGDA is interrupted.

## 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].

### 3.1 Fonseca test case

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

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

The design variable is $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. This testcase is known to yield a continuous but concave Pareto set in function space. 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 by MGDA, we have applied the method starting from a set of some 50 initial design points located on a sphere in the design-space (Figure 3).

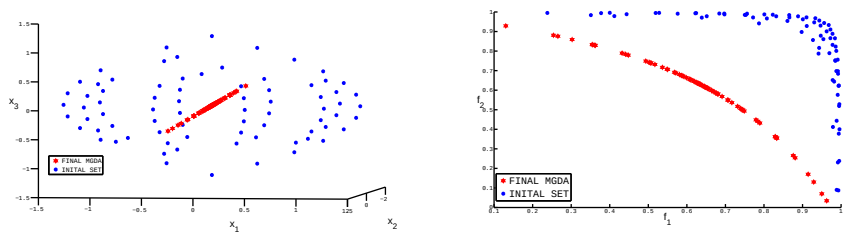



**Fig. 3** Convergence of MGDA to the Pareto front, for several initial design points., in design space $(x, y, z)$ (left) and in function space $(f_1, f_2)$ (right)

In all cases, MGDA converges and provides an accurately defined point on the Pareto set (Figure 4).

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 5 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 here in a fewer number.

However, at identical computational cost, generally, PAES introduces more diversity 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.

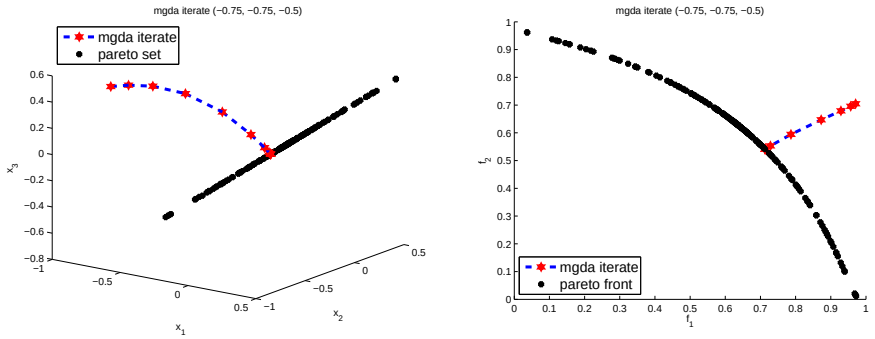



**Fig. 4** (Fonseca test case) Convergence of MGDA to the Pareto front, for several initial design points.

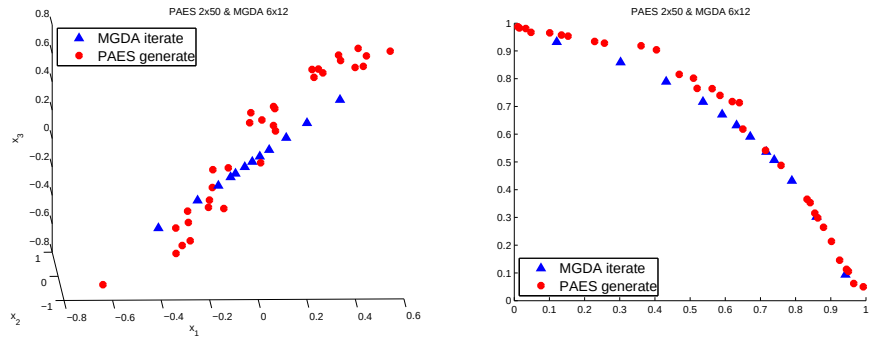



**Fig. 5** (Fonseca test case) Pareto set approximated discretely by PAES and MGDA.

In each case about 3 to 4 iterations are sufficient to converge and produce the accurate result indicated on Figure 6.

## 4 Applying MGDA on a Kriging metamodel

In this section, we conduct numerical experiments to demonstrate the convergence of MGDA to Pareto optimal solutions in conjunction with Kriging metamodels. A first Kriging metamodel is constructed with an initial database. From each initial point, MGDA yields to a better point used subsequently to update the metamodel.

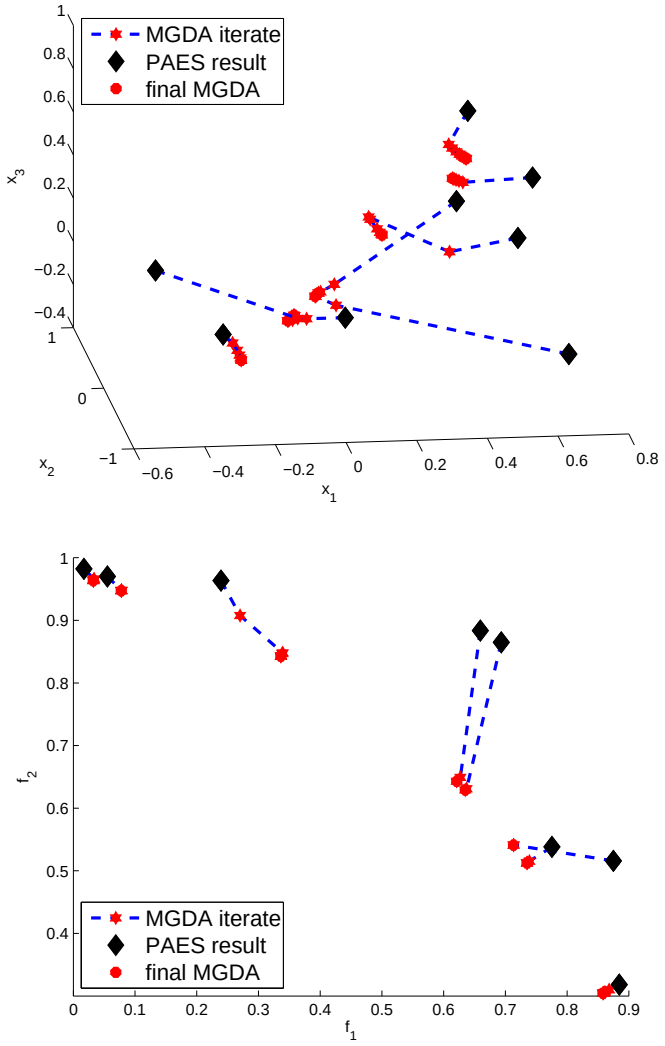



**Fig. 6** (Fonseca test case) First step with a large PAES followed by MGDA iterates on each non dominated point found. Design space on the left, functional space on the right.

## *4.1 Kur test case*

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

$$g_1(x) = -\sum_{i=1}^{2} -10\exp\left(-0.2\sqrt{x_i^2 + x_{i+1}^2}\right), \quad g_2(x) = \sum_{i=1}^{3}\left(|x_i|^{0.8} + 0.5\sin(x_i^3)\right)$$

The design variable is $x = (x_1, x_2, x_3) \in \mathbb{R}^3$. This testcase is known to yield a nonconvex discontinuous Pareto set in function space. Two generations of non dominated points applying PAES from different initial configurations gives a good discrete approximation of the Pareto front obtained by Deb [3]. Figure 7 shows that the Pareto set is here discontinuous, especially in the design space, where three distinct groups of points are evident.

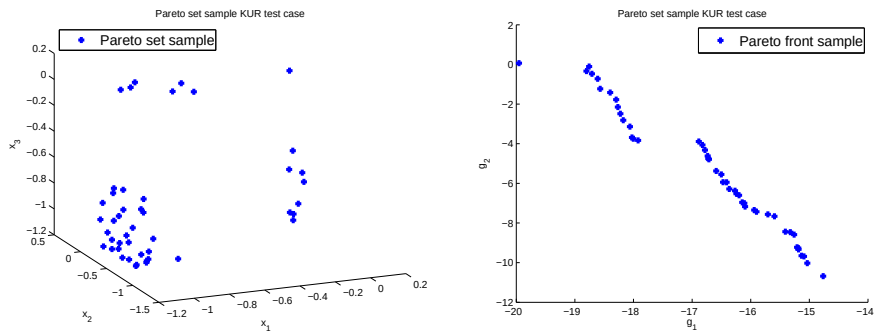



**Fig. 7** (Kur test case) Discrete Pareto front produced by 2 generations of PAES optimization.

In the next experiment, we have first applied PAES once from one initial design point to generate 100 new points. PAES sorts out 11 non dominated points from these 100. For each point obtained, MGDA produces a new one closer to the Pareto front, as illustrated by Figure 8.

Because of the sine in the second function, this test case is a multi modal problem. Thus optimization algorithms based on Gradient descent methods have experience difficulties. To asses the MGDA, a clever strategy must be adopted to generate a sufficiently diverse set of initial points. Presently we use an initial small and diverse set of design points forming a sample of a latin hypercube. This set gives a Kriging metamodel on which MGDA drives each initial point to a better one in terms of function values. If the MGDA points are sufficently widespread, a new metamodel is constructed with the initial set augmented. Whenever a new point is found too close to another one in the database, it is not considered to update the metamodel. In a few iterations of this method, best points obtained are close to the Pareto front. The following experiment (Figure 9) is based on a set of 10 initial design points from $[-5,5]^3$ evaluated with respect to $g_1$ and $g_2$, after 10 iterations of the process described above. The computational cost corresponds to 43 point evaluations, including the database but not the metamodel construction.

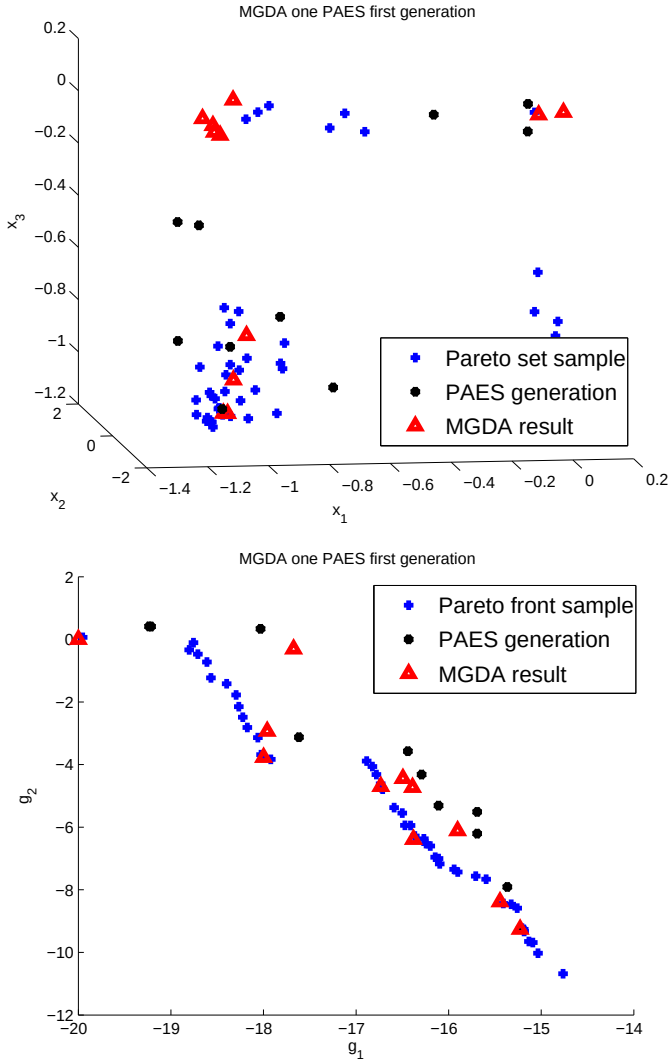



**Fig. 8** (Kur test case) Applying MGDA to each non dominated point from one PAES generation of 100 points (sort 11 non dominated).

## *4.2 CFD test case*

The last experiment is an optimum-shape design in compressible aerodynamics. The transonic flow ($M_\infty = 0.83$, $\alpha = AoA = 2^o$) about a generic aircraft wing is simulated by the solution of the 3D Euler equations by an upwind finite volume method over an unstructured mesh of some 200,000 points generated by the software GMSH [7].

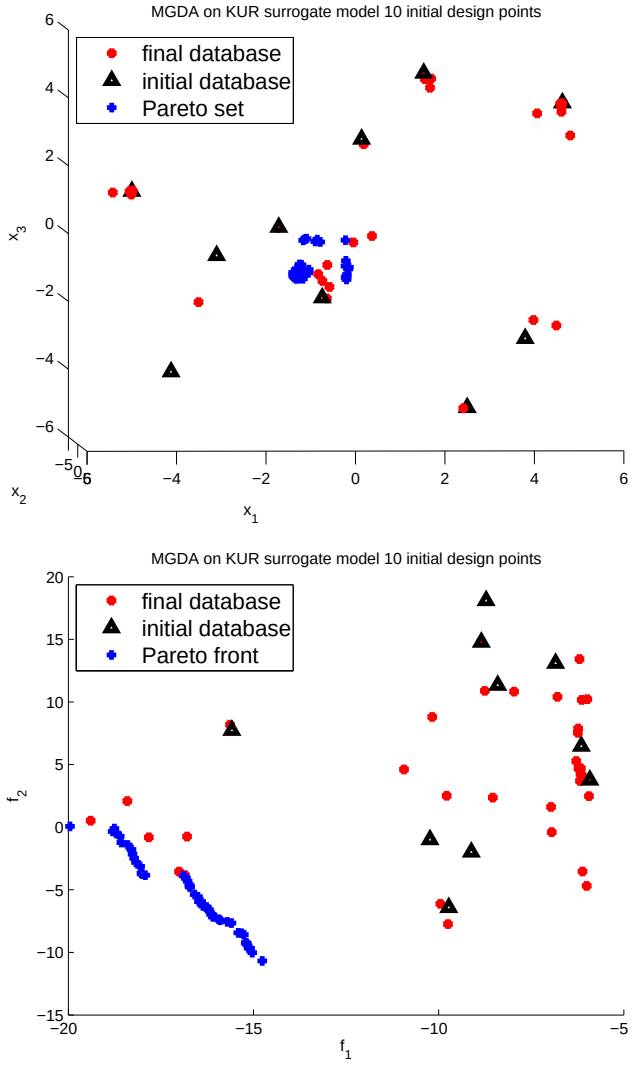



**Fig. 9** (Kur test case) Evolution of points gived by MGDA on an evoluting Kriging metamodel. 10 initial design points lead to 43 points

The cross sections of the wing are made homothetic with a linear variation in the spanwise direction. Thus only the shape of these sections, an airfoil, is optimized. This airfoil is represented by 7 B-spline functions for the upper surface, and 7 other ones for the lower surface. The leading and trailing edges are fixed, and this permits us to introduce a total of 10 geometrical design variables. Intially, these variables are set to define a cross section close to the classical NACA0012 airfoil.

MGDA is used here to solve the two-criterion optimization problem consisting of maximizing the lift coefficient and minimizing the drag coefficient simultaneously, starting from the specified initial geometry.

An initial set of 40 design points forming a sample of a latin hypercube in $\mathbb{R}^{10}$ haa been considered. This first set of datapoints is employed for two purposes. First,

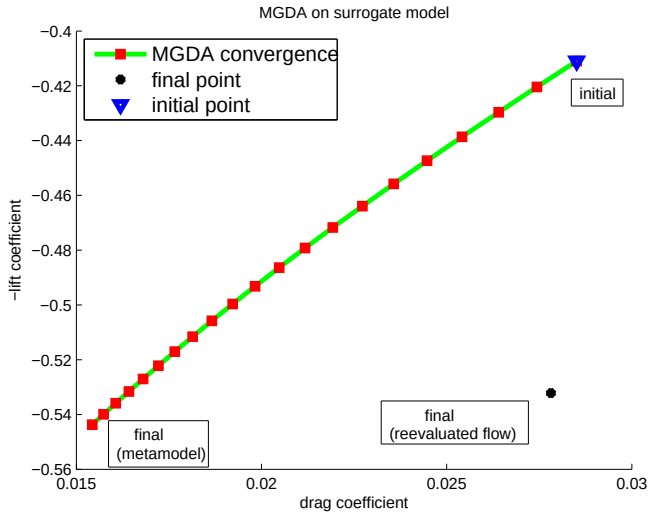



**Fig. 10** (Eulerian flow test case) Example of the convergence of MGDA from an initial database point on the metamodel and the corresponding simulation result point.

it is used to construct initial Kriging metamodels of both functions (lift and drag). Second, it is used throughout the following cycle to provide starting points to initiate the MGDA iteration in different conditions. This iteration is conducted until convergence using at every iteration, gradients that are calculated by local finite-differences of the metamodels. Each converged point belongs to the Pareto set associated with the two-criterion problem related to the metamodels. It is then reealuated by a flow computation, and added to the database unless it is found too close to an existing point. At completion of this database enrichment process, the metamodels are updated, and this completes the cycle. In practice, in what follows, only two cycles were performed.

Figure 10 represents the convergence of MGDA from a particular initial database point. The figure indicates the converged point and the point obtained by the same design reevaluated by an Euler flow computation (actual lift and drag).

Figure 11 represents the initial database of 40 points and the ultimate database. With only 95 calls to the CFD solver, a significant improvement of both criteria is achieved and visibly, an approximate Pareto front begins to form.

Figure 12 represents pressure fields on the wing and the symmetry plane, corresponding to 3 particular non dominated points of the ultimate database (points 1 and

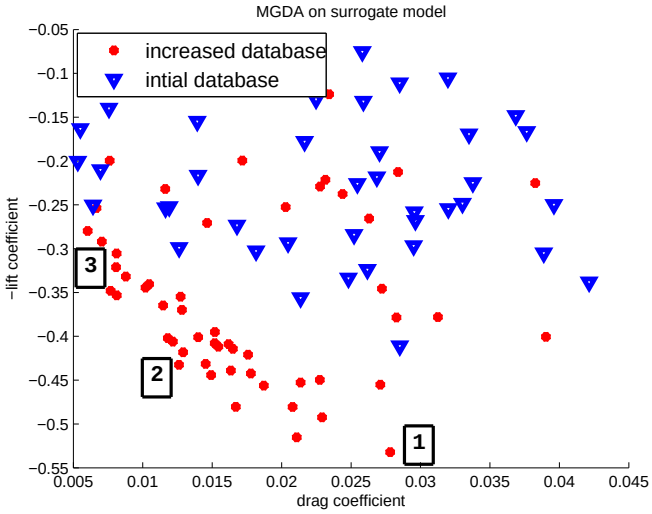
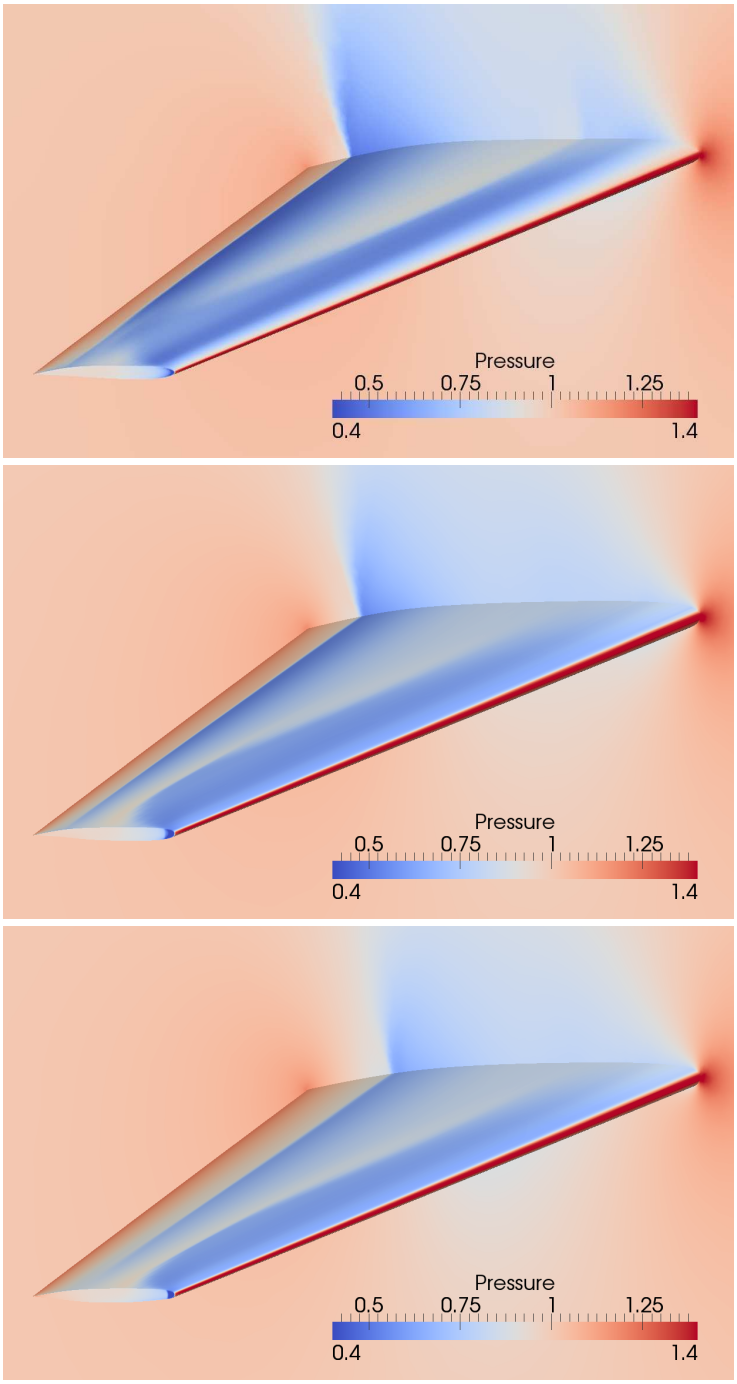



**Fig. 11** (Eulerian flow test case) Evolution of datapoints by MGDA applied to Kriging metamodels of lift and drag; the dataset is made of 40 design points initially, and 95 ultimately. All points are Eulerian simulation result.

3 of Figure 11 on top and bottom respectively). Point 1 corresponds to the flow with the shock wave of strongest intensity of the three; it produces the largest values of both lift and drag. Inversely, point 3 is associated with the smallest values, and point 2 with intermediate.




(point 1)

(point 2)

(point 3)

**Fig. 12** Pressure field associated with design points 1, 2 and 3 of Figure 11

## 5 Conclusion

In this article, we have tested by numerical experiment a recently proposed gradient-based algorithm for multiobjective optimization, MGDA [1].

First, the convergence to Pareto-optimal solutions has been demonstrated in an analytical testcase corresponding to a continuous, but concave Pareto front (Fonseca test case). Additional information on this comparison can be found in [6].

Second, MGDA has been compared with the well-known PAES algorithm. Both the Fonseca and Kur testcases have been considered in this comparison. We found that the two algorithms have particular merits of their own. PAES is very effective to converge to a very diverse dataset, whereas MGDA achieves this only if the initial set of design points is itself diverse. However, the iterative convergence of MGDA which makes use of (approximate) gradients is much faster. Thus both algorithms are complementary.

Thirdly, a hybrid method has been proposed and tested over the above mathematical testcases, demonstrating promising potentials.

Lastly, in the context of a two objective aerodynamic wing shape optimization in which the 3D Euler equations have been solved, MGDA has been used to define a strategy to progressively enrich the database associated with metamodels of drag and lift. With less than 100 calls to the flow solver, both lift and drag have been improved significantly from an initial design of a wing whose cross section was close to the classical NACA0012 airfoil.

## References

1. Désidéri J.A. : Multiple-Gradient Descent Algorithm (MGDA), INRIA research report 6953, June 2009, http://hal.inria.fr/inria-00389811/en/
2. Désidéri J.A. : Split of Territories in Concurrent Optimization, INRIA research report 6108, November 2007, http://hal.inria.fr/inria-00127194/en/
3. Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, T. Meyarivan : A Fast and Elitist Multi-Objective Genetic Algorithm : NSGA-II, Transaction on evolutionary computation vol 6 NO. 2, , April 2002, p.182-197
4. Knowles J.D., Corne D.W. : Approximating the nondominated front using the Pareto Archived Evolution Strategy, Evolutionary Computation Volume 8 Issue 2, MIT press Cambridge, June 2000
5. Praveen C., Duvigneau R. : Radial Basis Functions and Kriging Metamodels for Aerodynamic Optimization, INRIA research report 6151, March 2007 . http://hal.inria.fr/docs/00/13/79/35/PDF/RR-6151.pdf
6. Zerbinati A., Désidéri J.A. Duvigneau R.: Comparison between MGDA and PAES for multi objective optimization, INRIA research report 7667, June 2011, http://hal.inria.fr/docs/00/60/54/23/PDF/RR-7667.pdf
7. Geuzaine C., Remacle J.F. : a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities, http://geuz.org/gmsh/