Multiple-Gradient Descent Algorithm (MGDA)

Jean-Antoine Désidéri

Multiple-Gradient Descent Algorithm (MGDA)

Jean-Antoine Désidéri

RESEARCH REPORT

N° 6953

June 5, 2009

Project-Team Opale

ISSN 0249-6399 ISRN INRIA/RR--6953--FR+ENG

Multiple-Gradient Descent Algorithm (MGDA)

Jean-Antoine Désidéri*

Project-Team Opale

Research Report n° 6953 — version 3† — initial version June 5, 2009 — revised version November 5, 2012 — 16 pages

Abstract: In a previous report [4], a methodology for the numerical treatment of a two-objective optimization problem, possibly subject to equality constraints, was proposed. The method was devised to be adapted to cases where an initial design-point is known and such that one of the two disciplines, considered to be preponderant, or fragile, and said to be the *primary discipline*, achieves a local or global optimum at this point. Then, a particular split of the design variables was proposed to accomplish a *competitive-optimization* phase by a Nash game, whose equilibrium point realizes an improvement of a *secondary discipline*, while causing the least possible degradation of the primary discipline from the initial optimum. In this new report, the initial design point and the number of disciplines are arbitrary. Certain theoretical results are established and they lead us to define a preliminary *cooperative-optimization* phase throughout which all the criteria improve, by a so-called *Multiple-Gradient Descent Algorithm (MGDA)*, which generalizes to n disciplines ($n \geq 2$) the classical steepest-descent method. This phase is conducted until a design-point on the Pareto set is reached; then, the optimization is interrupted or continued in a subsequent competitive phase by a generalization of the former approach by territory splitting and Nash game.

Key-words: Optimum–shape design, concurrent engineering, multi-criterion optimization, split of territory, Nash and Stackelberg game strategies, Pareto optimality, descent direction, steepest-descent direction

* INRIA Research Director, Opale Project-Team Head

† In this revised version, the proof of Lemma 1 has been generalized to permit the number n of objective-functions and the dimension N of the working design-space to compare arbitrarily in the definition of Pareto-stationarity.

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Algorithme de descente à gradients multiples (MGDA)

Résumé : Dans un précédent rapport [4], on a proposé une méthodologie pour le traitement numérique d'un problème d'optimisation bicritère, éventuellement soumis à des contraintes d'égalité. La méthode a été conçue pour s'adapter aux cas où l'on connait un point de conception initial où l'une des disciplines, dite *discipline principale* car prépondérante, ou fragile, atteint un optimum local ou global. Alors, on a proposé une méthode de partage des variables afin de réaliser une phase d'*optimisation compétitive* par un jeu de Nash, dont le point d'équilibre produit une amélioration de la *discipline secondaire*, tout en causant la moindre dégradation possible de la discipline principale par rapport à l'optimum initial. Dans ce nouveau rapport, le point de conception initial et le nombre de disciplines sont arbitraires. On établit certains résultats théoriques qui nous conduisent à définir une phase préliminaire d'*optimisation coopérative* au cours de laquelle tous les critères s'améliorent, par un *Algorithme de Descente à Gradients Multiples (MGDA)* qui généralise à n disciplines l'algorithme classique du gradient. Cette phase est prolongée jusqu'à atteindre un point de conception du front de Pareto; alors, on peut éventuellement continuer l'optimisation par une phase suivante, compétitive, en généralisant l'approche précédente par partage de territoire et jeu de Nash.

Mots-clés : Conception optimale de forme, ingénierie concourante, optimisation multi-critère, partage de territoire, stratégies de jeux de Nash ou de Stackelberg, Pareto-optimalité, direction de descente, direction de plus grande pente

1 Introduction, problem setting and notations

In a previous report [4], a methodology for the numerical treatment of a two-objective minimization problem, possibly subject to equality constraints was proposed. Such a problem is a basic step in multidisciplinary optimization, sometimes referred to as *concurrent engineering*. There, we considered the case where one criterion to be minimized, J_A, is preponderant over the second, J_B, that is, either more critical or more fragile. The problem was formulated as a parametric optimization in which the two criteria are smooth functions of a common design vector $Y \in \mathbb{R}^N$. The numerical procedure was defined in two steps.

In the first step, the criterion associated with the preponderant criterion $J_A(Y)$, or primary discipline, is minimized first, alone, to full convergence by hypothesis, yielding the design vector $Y_A^\star$. As a result, the gradient vector, $\nabla J_A^\star$, the Hessian matrix, $H_A^\star$, and the K constraint gradients, $\nabla g_k^\star$ are assumed to be known at $Y = Y_A^\star$. In practice, such information may be difficult to calculate exactly when the finite-dimensional parametric formulation is the result of discretizing functionals of the distributed solution of a complex set of partial differential equations, as it is the case in the prototype example of aerodynamic optimum-shape design; then, possibly, but not necessarily, the exact derivatives can be replaced by approximations through meta-modeling of the functionals.

In preparation of the second step, the entire parametric space is then split into two supplementary subspaces on the basis of the analysis of the second variation of the primary functional, subject to the K active constraints. The construction is such that infinitesimal perturbations in the design vector Y about $Y_A^\star$ lying in the second subspace, whose dimension p is adjustable ($p \leq N - K$), cause potentially the least degradation to the primary functional value. In other words, the second subspace is the subspace of dimension p of least sensitivity of the preponderant criterion J_A.

In the second step of the optimization, a Nash equilibrium [6] is sought between the two disciplines by introducing two virtual players, each one in charge of minimizing its own criterion, J_A or J_{AB} (a convex combination of $J_A/J_A^\star$ and $\theta J_B/J_B^\star$), w.r.t. either subset of parameters that generates one of the two supplementary subspaces. In this way, the secondary criterion is potentially reduced, while not increasing unduly the primary criterion from its initial minimum.

The Nash game is given a particular form in which a continuation parameter ε ($0 \leq \varepsilon \leq 1$) is introduced. The optimum solution $Y_A^\star$ achieved at completion of the first step of the optimization is proved to be a Nash equilibrium solution of our formulation for $\varepsilon = 0$. Thus, as ε increases from 0 to 1, the formulation provides a continuum of Nash equilibrium solutions, corresponding to a smooth introduction of the trade-off between the two disciplines. Along the continuum, the initial derivative of the primary functional w.r.t. ε is also proved to be equal to zero, which can be viewed as a type of robust-design result. In practice this offers the designer the possibility to elect a design point along the continuum, not necessarily at the endpoint $\varepsilon = 1$.

Lastly we observed that the hierarchy introduced above between the criteria was applied to the split of territory in preparation of a Nash game, which is by essence symmetrical. The bias is therefore different in nature from the unsymmetrical treatment of the variables introduced in a Stackelberg-type game [2].

In [4], our formulation was first demonstrated in the simple case of the minimization of two quadratic forms in $\mathbb{R}^4$ subject to a linear, or a nonlinear equality constraint. The methodology was further illustrated by the treatment a difficult exercise of a generic aircraft wing shape optimization w.r.t. two criteria, one representative of the aerodynamic performance (drag subject to a lift constraint) and the other of the structural design (average stress subject to geometrical constraints) taken from B. Abou El Majd's doctoral thesis [1].

In this new report, we demonstrate how the general problem of the unconstrained simultaneous minimization of n smooth criteria (or disciplines) $J_i(Y)$ (Y : design vector; $Y \in \mathcal{H}$; $\mathcal{H}$: working space, a Hilbert space usually equal to $\mathbb{R}^N$, but possibly a subspace of L^2 also) can formally be accomplished in two phases :

1. A *cooperative phase*, beneficial to all criteria, in the sense that all criteria are reduced at each step of this phase.
2. A *competitive phase* realized through a Nash game with an appropriate definition of the split of territory.

In the usual case of a finite-dimensional working space $\mathcal{H} = \mathbb{R}^N$, no assumption is made on how the integers n and N compare.

The cooperative phase is conducted by an algorithm that generalizes the classical steepest-descent method to n concurrent disciplines. At completion of this phase, the design vector provides a point on the Pareto set that is not necessarily Pareto-optimal [5], if not all the criteria are locally convex. In that case only, it makes sense to define a subsequent minimization phase, competitive in nature. For this, we propose to generalize the split of territory defined in the former report [4] in preparation of a dynamic Nash game, implemented numerically, preferably on a parallel architecture.

2 Pareto concepts

We refer to the excellent textbook by K. Miettinen [5] for a detailed review of fundamentals in nonlinear multiobjective optimization. and much more. Here, we simply formulate a number of theoretical results that are essential to our subsequent algorithmic construction.

Thus consider n smooth criteria $J_i(Y)$ (Y : design vector; $Y \in \mathcal{H}$; $\mathcal{H}$: working space, a Hilbert space usually equal to $\mathbb{R}^N$, but possibly a subspace of L^2 also). In practice, these functions or functionals are assumed to be of class C^1 and convex in some working open ball $\mathcal{B}$ of the design space $\mathcal{H}$. Throughout this report, unless specified otherwise, the symbol N denotes the dimension of the design-space $\mathcal{H}$ when it is finite and the symbol ∞ otherwise.

In this revised version of the report, the following two lemmas are introduced in the alternate order (and renumbered). Lemma 2 (formerly Lemma 1) is now given a more rigorous and general proof based on Lemma 1. In particular, the case $n > N$ is now encompassed.

Lemma 1
Let $\mathcal{H}$ be a Hilbert space of finite or infinite dimension N, and $\{u_i\}$ $(1 \le i \le n)$ a family of n vectors in $\mathcal{H}$. Let $\mathcal{U}$ be the set of strict convex combinations of these vectors :

$$\mathcal{U} = \left\{ w \in \mathcal{H} \ / \ w = \sum_{i=1}^{n} \alpha_i \, u_i \, ; \ \alpha_i > 0 \ (\forall i) \, ; \ \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{1}$$

and $\overline{\mathcal{U}}$ its closure (the convex hull of the family). Then, there exists a unique element $\omega \in \overline{\mathcal{U}}$ of minimum norm, and :

$$\forall \bar{u} \in \overline{\mathcal{U}} \ : \ (\bar{u}, \omega) \ge (\omega, \omega) = \|\omega\|^2 := C_\omega \tag{2}$$

Proof : the convex hull $\overline{\mathcal{U}}$ is a closed and convex set, and this implies existence and uniqueness of the element ω of minimum norm in $\overline{\mathcal{U}}$.

Let $\bar{u}$ be an arbitrary element of $\overline{\mathcal{U}}$; set $r = \bar{u} - \omega$ so that $\bar{u} = \omega + r$. Since the convex hull $\overline{\mathcal{U}}$ is convex,

$$\forall \varepsilon \in [0,1] \, , \ \omega + \varepsilon r \in \overline{\mathcal{U}} \tag{3}$$

Since ω is the element of $\overline{\mathcal{U}}$ of minimum norm, $\|\omega + \varepsilon r\| \ge \|\omega\|$, which writes :

$$\|\omega + \varepsilon r\|^2 - \|\omega\|^2 = (\omega + \varepsilon r, \omega + \varepsilon r) - (\omega, \omega) = 2\varepsilon \, (r, \omega) + \varepsilon^2 \, (r, r) \ge 0 \tag{4}$$

and since ε can be arbitrarily small, this requires that :

$$(r,\omega) = (\bar{u} - \omega, \omega) \geq 0 \tag{5}$$

from which the result follows directly. □

Let us introduce and use throughout the following very natural definition, although the terminology does not seem to be standard :

Definition 1 (Pareto-stationarity)
Let $J_i(Y)$ $(1 \leq i \leq n,\ Y \in \mathcal{B} \subseteq \mathbb{R}^N)$ be smooth and convex objective-functions over the open ball $\mathcal{B}$ centered at the design-point Y^0. These objective-functions are said to be Pareto-stationary at Y^0 iff there exists a convex combination of the gradient-vectors, $u_i^0 = \nabla J_i(Y^0)$, that is equal to zero:

$$\exists \alpha = \{\alpha_i\} \text{ such that } \alpha_i \geq 0 \ \ (\forall i)\ ,\ \sum_{i=1}^{n} \alpha_i = 1\ ,\ \text{and } \sum_{i=1}^{n} \alpha_i\, u_i^0 = 0\,. \tag{6}$$

Lemma 2
Let Y^0 be a Pareto-optimal point, center of an open ball $\mathcal{B}$ in which the objective-functions $J_i(Y)$ $(1 \leq i \leq n)$ are smooth and convex, and define the gradient-vectors $u_i^0 = \nabla J_i\left(Y^0\right)$ in which ∇ is the symbol for the gradient; then, the objective-functions are Pareto-stationary at Y^0.

Proof : Without loss of generality, assume $J_i(Y^0) = 0$ $(\forall i)$. Then:

$$Y^0 = \mathsf{Arg}\min_Y J_n(Y) \quad \text{subject to: } J_i(Y) \leq 0\ (\forall i \leq n-1)\,. \tag{7}$$

Let $\overline{U}_{n-1}$ be the convex hull of the gradients $\{u_1^0, u_2^0, \ldots, u_{n-1}^0\}$ and $\omega_{n-1} = \mathsf{Arg}\min_{u \in \overline{U}_{n-1}} \|u\|$. By virtue of Lemma 1, the vector ω_{n-1} exists, is unique, and such that:

$$\left(u_i^0, \omega_{n-1}\right) \geq \|\omega_{n-1}\|^2 \ \ (\forall i \leq n-1)\,. \tag{8}$$

<u>Two situations are then possible:</u>

1. $\omega_{n-1} = 0$, and the objective-functions $\{J_1, J_2, \ldots, J_{n-1}\}$ satisfy the Pareto stationarity condition at $Y = Y^0$. *A fortiori,* the condition is also satisfied by the whole set of objective-functions.

2. Otherwise $\omega_{n-1} \neq 0$. Then let

$$j_i(\epsilon) = J_i(Y^0 - \epsilon\omega_{n-1})\ (i = 1, \ldots, n-1) \tag{9}$$

so that

$$j_i(0) = 0 \text{ and } j_i'(0) = -\left(u_i, \omega_{n-1}\right) \leq -\|\omega_{n-1}\|^2 < 0\,, \tag{10}$$

and for sufficiently small strictly-positive ϵ:

$$j_i(\epsilon) = J_i(Y^0 - \epsilon\omega_{n-1}) < 0 \qquad (\forall i \leq n-1) \tag{11}$$

which establishes that Slater's qualification condition [3] is satisfied for the optimization problem (7) subject to inequality constraints. Thus, the Lagrangian

$$\mathcal{L} = J_n(Y) + \sum_{i=1}^{n-1} \lambda_i J_i(Y) \tag{12}$$

is stationary w.r.t. Y at $Y = Y^0$, and this gives:

$$u_n^0 + \sum_{i=1}^{n-1} \lambda_i u_i^0 = 0 \tag{13}$$

in which $\lambda_i > 0$ ($\forall i \le n-1$) since equality constraints hold $J_i(Y^0) = 0$ (KKT condition). Normalizing this equation by dividing by the number $1 + \sum_{i=1}^{n-1} \lambda_i$ which is greater then 1 (thus, strictly-positive), results in the Pareto-stationarity condition. $\square$

Thus, in general, for smooth and convex unconstrained criteria, Pareto-stationarity is a necessary condition for Pareto-optimality. Inversely, if the smooth criteria $J_i(Y)$ ($1 \le i \le n$) are not Pareto-stationary at a given design-point Y^0, descent directions common to all criteria exist. We now examine how such a direction can be identified. We have the following :

Combining Lemma 1 and Lemma 2 with Definition 1 yields the following :

Theorem 1
Let $\mathcal{H}$ be a Hilbert space of finite or infinite dimension N. Let $J_i(Y)$ ($1 \le i \le n$) be n smooth and convex objective-functions of the vector $Y \in \mathcal{B} \subseteq \mathcal{H}$, and Y^0 a particular admissible design-point, at which the gradient-vectors are denoted $u_i^0 = \nabla J_i(Y^0)$, and

$$\mathcal{U} = \left\{ w \in \mathcal{H} \; / \; w = \sum_{i=1}^{n} \alpha_i \, u_i^0 \, ; \; \alpha_i > 0 \; (\forall i) \, ; \; \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{14}$$

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

Proof : all the elements of this theorem are reformulations of previous results, except for the statement concerning the scalar product $(\bar{u}, \omega)$ in the second case when additionally $\omega \in \mathcal{U}$ (and not simply $\overline{\mathcal{U}}$). To establish this last point, observe that under these assumptions, the element ω is the solution to the following minimization problem :

$$\omega = u = \sum_{i=1}^{n} \alpha_i u_i^0 \, , \;\; \alpha = \mathsf{Argmin}\, \mathsf{j}(u) \, , \;\; \mathsf{j}(u) = (u, u) \, , \;\; \sum_{i=1}^{n} \alpha_i = 1 \tag{15}$$

since by hypothesis, none of the inequality constraints, $\alpha_i > 0$, is saturated. Consequently, using the vector $\alpha \in \mathbb{R}^n$ as the finite-dimensional variable, the Lagrangian writes :

$$\mathsf{L}(\alpha, \lambda) = \mathsf{j} + \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right) \tag{16}$$

and the optimality conditions satisfied by the vector α are the following :

$$\frac{\partial \mathsf{L}}{\partial \alpha_i} = 0 \; (\forall i) \, , \;\; \frac{\partial \mathsf{L}}{\partial \lambda} = 0 \tag{17}$$

These equations imply that for all indices i :

$$\frac{\partial \mathsf{j}}{\partial \alpha_i} + \lambda = 0 \tag{18}$$

But, $\mathsf{j}(u) = (u, u)$ and for $u = \omega = \sum_{i=1}^{n} \alpha_i u_i$, one has :

$$\frac{\partial \mathsf{j}}{\partial \alpha_i} = 2(\frac{\partial u}{\partial \alpha_i}, u) = 2(u_i^0, \omega) = -\lambda \Longrightarrow (u_i^0, \omega) = -\lambda/2 \tag{19}$$

independently of i. Now consider an arbitrary element $\bar{u} \in \overline{\mathcal{U}}$:

$$\bar{u} = \sum_{i=1}^{n} \mu_i \, u_i^0 \tag{20}$$

where $\mu_i \geq 0$ $(\forall i)$ and $\sum_{i=1}^{n} \mu_i = 1$. Then :

$$(\bar{u}, \omega) = \sum_{i=1}^{n} \mu_i \, (u_i^0, \omega) = -\lambda/2 \text{ (a constant)} = \|\omega\|^2 \tag{21}$$

where the constant has been evaluated by letting $\bar{u} = \omega$. $\square$

In summary, one is led to identify the vector

$$\omega = \sum_{i=1}^{n} \alpha_i \, u_i^0 \tag{22}$$

by solving the following quadratic-form constrained minimization problem in $\mathbb{R}^n$:

$$\min_{\alpha \in \mathbb{R}^n} \left\| \sum_{i=1}^{n} \alpha_i \, u_i^0 \right\|^2 \tag{23}$$

subject to :

$$\alpha_i \geq 0 \; (\forall i) \, , \; \sum_{i=1}^{n} \alpha_i = 1 \tag{24}$$

Note that in a finite-dimensional setting, and in a functional-space setting as well, the above problem can be solved in $\mathbb{R}^n$, so long as the gradients $\{u_i^0\}$ $(1 \leq i \leq n)$ and their scalar products $\{u_{ij}^0 := (u_i^0, u_j^0)\}$ are known. Then, a call to a library procedure should be sufficient.

Let us now examine the instructive particular case of two criteria $(n = 2)$. Let $u = u_1^0$, $v = u_2^0$, $\alpha = \alpha_1$ and $\alpha_2 = 1 - \alpha$, to simplify the notation, and consider the following quadratic form :

$$j(\alpha) = \|\alpha u + (1-\alpha) v\|^2 = \Big(\alpha u + (1-\alpha)v, \alpha u + (1-\alpha) v\Big) \tag{25}$$

so that :

$$j'(\alpha) = 2\Big(u - v, \alpha(u - v) + v\Big) \tag{26}$$

Hence, putting aside the trivial case where $u = v$, the minimum is achieved for

$$\alpha = \frac{v.(v-u)}{\|u-v\|^2} = \frac{\|v\|^2 - v.u}{\|u\|^2 + \|v\|^2 - 2u.v} \tag{27}$$

In fact, the orthogonal projection of the null vector 0 onto the the convex hull given by $\alpha u + (1-\alpha)v$ (without limitation on α), is given by the condition :

$$\Big(\alpha u + (1-\alpha)v\Big).(u - v) = 0 \tag{28}$$

which gives the same expression for α.

This expression shows that if the vectors u and v are first normalized so that $\|u\| = \|v\| = 1$, one gets $\alpha = \frac{1}{2}$. Inversely, if the vectors u and v are not normalized first, it is not certain that the above α be in [0,1]; indeed, utilizing here the usual dot-product notation for the scalar product, one has :

$$\begin{aligned} 0 < \alpha < 1 &\Longleftrightarrow 0 < \|v\|^2 - v.u < \|u\|^2 + \|v\|^2 - 2u.v \\ &\Longleftrightarrow u.v < \min\Big(\|u\|^2, \|v\|^2\Big) = \min\Big(\|u\|, \|v\|\Big)^2 \\ &\Longleftrightarrow \cos\widehat{(u,v)} < \frac{\min\Big(\|u\|, \|v\|\Big)}{\max\Big(\|u\|, \|v\|\Big)} \\ &\Longleftrightarrow \widehat{(u,v)} > \cos^{-1} \frac{\min\Big(\|u\|, \|v\|\Big)}{\max\Big(\|u\|, \|v\|\Big)} \end{aligned} \tag{29}$$

Thus the condition is that the angle $\widehat{(u,v)}$ between the two gradient-vectors u and v be at least equal to a certain limit-angle which is a function of their norms whose values are in $[0,\pi/2]$. A sufficient condition is therefore that this angle be obtuse ($u.v < 0$). Whenever the norms of the gradient-vectors u and v are very different, the limit-angle is close to $\pi/2$. Inversely, if the norms are close to one another, the limit-angle is small, and the condition is satisfied except if the directions of the two gradient-vectors are too close (see Figure 1).

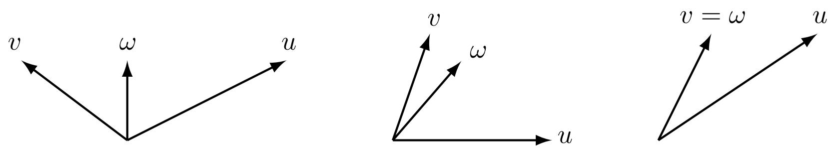

Figure 1: Various possible configurations of the two gradient-vectors u and v and the minimal-norm element ω depending on whether the angle $\widehat{(u,v)}$ is obtuse (left), acute but superior to the limit-angle (center), or acute and inferior to the limit-angle (right).

Also note that in the first two cases on the left of Figure 1, the vector ω points strictly in between the vectors representing u and v, and this corresponds to situations in which $\omega \in \mathcal{U}$, and not simply $\overline{\mathcal{U}}$. Then, ω is orthogonal to the dashed line that connects the extremities of these representatives, and evidently, as a result, the scalar product $(\bar{u},\omega)$ is constant and equal to $\|\omega\|^2$ for all $\bar{u} \in \overline{\mathcal{U}}$, as stated in the theorem.

Thus, the solution is the following :

$$\alpha = \begin{cases} \dfrac{v.(v-u)}{\|u-v\|^2} & \text{if } u.v < \min\Big(\|u\|\,,\|v\|\Big)^2 \\ 0 \text{ or } 1 & \text{otherwise, depending on whether } \min\Big(\|u\|\,,\|v\|\Big) = v \text{ or } u\,. \end{cases} \tag{30}$$

Lastly, when $\alpha \in [0,1]$, let :

$$\theta = \widehat{(u,v)} \qquad \gamma = u.v = \|u\|\ \|v\|\ \cos\theta \tag{31}$$

One has :

$$\omega = \frac{\|v\|^2 - u.v}{\|u\|^2 + \|v\|^2 - 2u.v}\, u + \frac{\|u\|^2 - u.v}{\|u\|^2 + \|v\|^2 - 2u.v}\, v \tag{32}$$

Then :

$$u.\omega = \frac{\Big(\|v\|^2 - \gamma\Big)\ \|u\|^2 + \Big(\|u\|^2 - \gamma\Big)\,\gamma}{\|u-v\|^2} = \frac{\|u\|^2\ \|v\|^2 - \gamma^2}{\|u-v\|^2} = \frac{\|u\|^2\ \|v\|^2}{\|u-v\|^2}\sin^2\widehat{(u,v)} \geq 0 \tag{33}$$

which is symmetric w.r.t. u and v; hence :

$$u.\omega = v.\omega \tag{34}$$

which, in part, constitutes a direct verification of Theorem 1.

3 Numerical strategy for multiobjective optimization

Given an initial admissible design-point Y^0 at which the smooth criteria $J_i(Y)$ $(1 \leq i \leq n)$ are not Pareto-stationary, we propose to develop the optimization process in several stages described in the following subsections.

3.1 Optional preliminary reformulation of criteria

In numerical experiments, it is preferable that the various criteria all be positive, and scaled in a somewhat unified way. To achieve this, we propose to modify the definitions of the criteria without altering the sense of the associated minimization problems.

For this purpose, let :

$$\mathcal{B}_R = \mathcal{B}\left(Y^0, R\right) \tag{35}$$

be a working ball in the design space about the initial design-point Y^0.

In a first step, we propose to replace each criterion $J_i(Y)$ by the following :

$$\tilde{J}_i(Y) = \exp\left(\alpha_i \frac{\left\|H_i^0\right\|}{\left\|\nabla J_i^0\right\|^2}\,\left(J_i(Y) - J_i^0\right)\right) \tag{36}$$

where :

- the superscript ^{0} indicates an evaluation at $Y = Y^0$;
- ∇J_i^0 and H_i^0 denote the gradient-vector and the Hessian matrix, and $\left\|H_i^0\right\|$ can be computed economically as : $\sqrt{\text{trace}\,\left[(H_i^0)^2\right]}$;
- α_i is a dimensionless constant.

In this way, the new criteria are dimensionless, they vary in the same sense as the original ones, and :

$$\forall i\,,\;\; \tilde{J}_i(Y^0) = 1\,,\;\; \nabla J_i(Y^0) = \frac{\gamma}{R} \tag{37}$$

provided the constants α_i's are chosen to satisfy :

$$\alpha_i \frac{\left\|H_i^0\right\|}{\left\|\nabla J_i^0\right\|} = \frac{\gamma}{R} \sim 1 \tag{38}$$

In the above, the dimensionless constant γ is given a value equal or close to the possibly-dimensional measure of R in the utilized system of units.

For a reason that will appear later, without altering the regularity of the criteria, we would like them to be infinite when $\|Y\|$ is infinite. For this, define the following function :

$$\phi(x) = \begin{cases} 0 & \text{if } x \le 0 \\ x \exp\left(-\dfrac{1}{x^2}\right) & \text{if } x > 0 \end{cases} \tag{39}$$

This function is C^∞ including at 0, and $\phi(x) \sim x$ as $x \to +\infty$. The new criterion

$$\tilde{\tilde{J}}_i(Y) = \tilde{J}_i(Y) + \varepsilon_0\,\phi\left(\frac{\left\|Y - Y^0\right\|^2}{R^2} - 1\right) \tag{40}$$

in which ε_0 is some strictly-positive constant, is identical to the former one, $\tilde{J}$, inside the working ball $\mathcal{B}_R$, and grows at least like $\|Y\|^2$ outside. The match of $\tilde{\tilde{J}}_i(Y)$ with $\tilde{J}_i(Y)$ and $J_i(Y)$ at the boundary of the working ball is infinitely smooth. Additionally :

$$\lim_{\|Y\| \to \infty} \tilde{\tilde{J}}_i(Y) = \infty \tag{41}$$

In what follows, it is implicit that the original criteria have been replaced by $\{\tilde{\tilde{J}}_i(Y)\}$ $(1 \le i \le n)$ and the double superscript $\tilde{\tilde{}}$ is omitted.

3.2 Cooperative-optimization phase : the Multiple-Gradient Descent Algorithm (*MGDA*)

The *MGDA* relies on the results of Theorem 1. The *MGDA* consists in iterating the following sequence :

1. Compute the gradient-vectors $u_i^0 = \nabla J_i(Y^0)$, and determine the minimum-norm element ω in the convex hull $\overline{\mathcal{U}}$. If $\omega = 0$, stop.

2. Otherwise, determine the step-size h which is, presumably optimally, the largest strictly-positive real number for which all the functions $j_i(t) = J_i(Y^0 - t\omega)$ $(1 \leq i \leq n)$ are monotone-decreasing over the interval $[0, h]$.

3. Reset Y^0 to $Y^0 - h\omega$.

In practice, the test $\omega = 0$ will be made with a tolerance ($\|\omega\| < tol$). In addition, note that the determination of the step-size h can be realized by the adaptation of nearly all standard one-dimensional search methods. This algorithm can be repeated a finite number of iterations if these iterations yield a design-point at which the Pareto-stationarity condition is satisfied, or indefinitely, if this never occurs.

Since at each iteration of the *MGDA*, all the criteria diminish, we refer to this process as a *cooperative-optimization* phase.

3.3 Convergence of the *MGDA*

The above *MGDA* can stop after a finite number of iterations if a Pareto-stationary design-point is reached. Otherwise, we have the following :

Theorem 2
If the sequence of iterates $\{Y^r\}$ of the MGDA is infinite, it admits a weakly convergent subsequence. (Here, the working Hilbert space $\mathcal{H}$ is assumed to be reflexive.)

Proof : the following elements hold, in part by virtue of the reformulation of the criteria :

- Since the sequence of values of any considered criterion, say $\{J_1(Y^r)\}$, is positive and monotone-decreasing, it is bounded.

- Since $J_1(Y)$ is infinite whenever $\|Y\|$ is infinite, the sequence of design-vectors $\{Y^r\}$ is bounded, and this implies the statement. $\square$

Let $Y^\star$ be the limit. We conjecture that the design-point $Y^\star$ is Pareto-stationary. In what follows, Y^0 is then reset to $Y^\star$.

3.4 Competitive-optimization phase

From a Pareto-stationary design-point Y^0, we propose either to interrupt the whole optimization process if the performance of the design-point is already satisfactory, or to continue this process by a *competitive optimization* phase. The competitive-optimization phase can be accomplished by a Nash game based on an appropriate split of variables (see next section for the case of two disciplines).

4 Nash game from a Pareto-stationary design-point ($n = 2$)

In this section, we propose recommendations to construct a Nash game to carry out the competitive-optimization phase, after completion of the cooperative-optimization phase, in the case of two disciplines ($n = 2$).

In the report [4], a split of territory was defined from the knowledge of a stationary point of one discipline, the preponderant discipline. Since critical points of one discipline are particular Pareto-stationary points, this subsection is meant to generalize the results of the former report.

For simplicity, we consider the case of two disciplines only ($n = 2$). An initial Pareto-stationary design-point Y^0 is known, and should be used to define a split of variables based on local eigen-systems, and a Nash equilibrium-point determined subsequently.

Here, the two criteria are denoted J_A and J_B, and at $Y = Y^0$, the following holds :

$$\alpha_A \nabla J_A^0 + \alpha_B \nabla J_B^0 = 0 \qquad \alpha_A + \alpha_B = 1 \tag{42}$$

for some $\alpha_A \in [0, 1]$. Therefore, three cases are possible :

1. Pareto-stationarity of type I : $\nabla J_A^0 = \nabla J_B^0 = 0$;
2. Pareto-stationarity of type II : $\nabla J_A^0 = 0$ and $\nabla J_B^0 \neq 0$ (or vice versa);
3. Pareto-stationarity of type III : $\nabla J_A^0 + \lambda \nabla J_B^0 = 0$ for $\lambda = \frac{1-\alpha_A}{\alpha_A} > 0$ since $0 < \alpha_a < 1$.

The question is what to do next to reaching a design-point Y^0 of Pareto-stationarity of the criteria $(J_A\,,\ J_B)$? To better understand the question, let us examine first the above three cases assuming both criteria are locally convex.

Convex case :

1. Pareto-stationarity of type I : then, both criteria have simultaneously achieved at $Y = Y^0$ local minimums of their own. In general the optimization process is terminated.
2. Pareto-stationarity of type II : e.g. $\nabla J_A^0 = 0$ and $\nabla J_B^0 \neq 0$. Then, J_A has achieved a local minimum, whereas J_B is still reducible. The decision can be to interrupt the process if the achieved design is acceptable, or to continue it using the formulation of the former theory [4] : a Nash equilibrium is sought based on a hierarchical split of variables in the orthogonal basis made of the eigenvectors of matrix H_A^0.
3. Pareto-stationarity of type III : $\nabla J_A^0 + \lambda \nabla J_B^0 = 0$ ($\lambda > 0$). Here, Pareto-optimality has been achieved and in the absence of an additional criterion, the optimization process is terminated.

We now turn to the general case in which the criteria are not assumed to be locally convex at $Y = Y^0$.

Non convex case : In what follows, we discuss the different cases according to various assumptions that can be made on the eigenvalues of the Hessian matrices H_A^0 and H_B^0 of the two criteria at $Y = Y^0$.

1. Pareto-stationarity of type I :

   Since both gradients are equal to zero, the principal term in the expansion of the variations of the two criteria caused by a perturbation δY of the design vector Y about Y^0 are the quadratic terms associated with the respective Hessian matrices, one of which, at least, is not positive-definite by assumption, and perhaps both.

If H_A^0 is positive-definite and H_B^0 alone has some negative eigenvalues, J_A has achieved a minimum whereas J_B is still reducible. Then we propose to terminate the optimization process, or to continue it using the formulation of the former theory [4] : a Nash equilibrium is sought with a hierarchical split of variables based on the eigensystem of matrix H_A^0.

If both Hessian matrices H_A^0 and H_B^0 have some negative eigenvalues, define the following families of linearly independent eigenvectors associated with these eigenvalues :

$$\mathcal{F}_A = \{ u_1 \, , \; u_2 \, , \; \dots \, , \; u_p \} \qquad \mathcal{F}_B = \{ v_1 \, , \; v_2 \, , \; \dots \, , \; v_q \} \tag{43}$$

Then :

- If the family $\mathcal{F}_A \cup \mathcal{F}_B$ is linearly dependent, say

$$\sum_{i=1}^{p} \alpha_i \, u_i - \sum_{j=1}^{q} \beta_j \, v_j = 0 \tag{44}$$

  in which $\{\alpha_i\}_{i=1,\dots,p} \cup \{\beta_j\}_{j=1,\dots,q} \neq \{0\}$, the vector

$$w^r = \sum_{i=1}^{p} \alpha_i \, u_i = \sum_{j=1}^{q} \beta_j \, v_j \tag{45}$$

  is not equal to zero (by linear independence of the families $\mathcal{F}_A$ and $\mathcal{F}_B$ separately), and it is a descent direction for both criteria. We then propose to make a step in that direction.
- Otherwise, $Sp\mathcal{F}_A \cap Sp\mathcal{F}_B = \{0\}$: then we propose to stop, or to determine the Nash equilibrium point using $\mathcal{F}_A$ (resp. $\mathcal{F}_B$) as the strategy of A (resp. B).

2. Pareto-stationarity of type II : say $\nabla J_A^0 = 0$ and $\nabla J_B^0 \neq 0$.

If the Hessian matrix H_A^0 is positive-definite, the criterion J_A has achieved a local minimum and this setting has been analyzed in [4] : a Nash equilibrium is sought with a hierarchical split based on the structure of the eigenvectors of H_A^0.

If instead the matrix H_A^0 has some negative eigenvalues, let

$$\mathcal{F}_A = \{ u_1 \, , \; u_2 \, , \; \dots \, , \; u_p \} \tag{46}$$

be a family of associated eigenvectors. Then :

- if ∇J_B^0 is not orthogonal to $Sp\mathcal{F}_A$: a descent direction common to J_A and J_B exists in $Sp\mathcal{F}_A$: use it to reduce both criteria.
- otherwise, $\nabla J_B^0 \perp Sp\mathcal{F}_A$: we propose to identify the Nash equilibrium using $\mathcal{F}_A$ as the strategy of player A and the remaining eigenvectors of H_A^0 as the strategy of player B.

3. Pareto-stationarity of type III : $\nabla J_A^0 + \lambda \nabla J_B^0 = 0$ ($\lambda > 0$).

Consider the direction defined by the vector :

$$\mathsf{u}_{AB} = \frac{\nabla J_A^0}{\|\nabla J_A^0\|} = -\frac{\nabla J_B^0}{\|\nabla J_B^0\|} \tag{47}$$

Along this direction, the two criteria vary in opposite ways and no rational decision can be made in the absence of other criteria. Thus consider instead possible move in the hyperplane orthogonal to u_{AB}. For this, consider reduced Hessian matrices :

$$H_A^{0\,\prime} = P_{AB} \, H_A^0 \, P_{AB} \qquad H_B^{0\,\prime} = P_{AB} \, H_B^0 \, P_{AB} \tag{48}$$

where :

$$P_{AB} = I - [\mathsf{u}_{AB}] \, [\mathsf{u}_{AB}]^t \; . \tag{49}$$

In this hyperplane, by orthogonality to the gradient-vectors, the analysis is that of Pareto-stationary point of type I in a subspace of dimension $N - 1$.

5 Conclusion

In this report, we have considered the problem of simultaneous unconstrained minimization of n smooth criteria. We have defined the notion of Pareto-stationarity, a weak form of the classical Pareto optimality. We have shown that when the current design point is not one of Pareto stationarity, the unique element ω of minimum norm in the convex hull of the gradient-vectors is nonzero, and $-\omega$ is a descent direction for all criteria simultaneously. This has led us to define the *Multiple-Gradient Descent Algorithm* (*MGDA*) which generalizes the classical steepest-descent algorithm to n disciplines, permitting to carry out a *cooperative-optimization* phase throughout which all criteria diminish at each iteration. This phase is possible until a Pareto-stationary design-point is reached associated with a representation point on the Pareto set. Then a *competitive-optimization* phase can be defined if the process of optimization should be continued. This new phase can be realized by a Nash game based on a split of variables guided by the analysis of the eigensystems of the Hessian matrices, in a way that generalizes the results of [4].

Acknowledgement

The author wishes to express his appreciation to his colleague Abderrahmane Habbal for extremely fruitful scientific discussions with him.

References

[1] B. Abou El Majd, *Algorithmes hiérarchiques et stratégies de jeux pour l'optimisation multidisciplinaire – Application à l'optimisation de la voilure d'un avion d'affaires*, Ph.D. thesis, Université de Nice-Sophia Antipolis, 2007.

[2] T. Basar and G. J. Olsder, *Dynamic noncooperative game theory*, Academic, Bodmin, Cornwall, Great Britain, 1995.

[3] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004, ISBN 978-0-521-83378-3. Retrieved October 3.

[4] J.-A. Désidéri, *Split of Territories in Concurrent Optimization*, Research Report 6108, INRIA, October 2007, URL: https://hal.inria.fr/inria-00127194 (version 6).

[5] K. M. Miettinen, *Nonlinear multiobjective optimization*, Kluwer Academic Publishers, Boston/London/Dordrecht, 1999.

[6] J. F. Nash, *Non-Cooperative Games*, Annals of Mathematics **54** (1951), no. 2, 286–295.

Contents

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399