C. R. Acad. Sci. Paris, Ser. I 350 (2012) 313–318

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Numerical Analysis/Calculus of Variations

Multiple-gradient descent algorithm (MGDA) for multiobjective optimization

Algorithme de descente à gradients multiples pour l'optimisation multiobjectif

Jean-Antoine Désidéri

INRIA, Centre de Sophia Antipolis Méditerranée, 2004, route des Lucioles, BP 93, 06902 Sophia Antipolis cedex, France

ARTICLE INFO

Article history:
Received 6 March 2012
Accepted 20 March 2012
Available online 29 March 2012

Presented by Olivier Pironneau

ABSTRACT

One considers the context of the concurrent optimization of several criteria $J_i(Y)$ ($i = 1, \dots, n$), supposed to be smooth functions of the design vector $Y \in \mathbb{R}^N$ ($n \leqslant N$). An original constructive solution is given to the problem of identifying a descent direction common to all criteria when the current design-point Y^0 is not Pareto-optimal. This leads us to generalize the classical steepest-descent method to the multiobjective context by utilizing this direction for the descent. The algorithm is then proved to converge to a Pareto-stationary design-point.

RÉSUMÉ

On se place dans le contexte de l'optimisation concourante de plusieurs critères $J_i(Y)$ ($i = 1, \dots, n$), fonctions régulières du vecteur de conception $Y \in \mathbb{R}^N$ ($n \leqslant N$). On donne une solution constructive originale au problème de l'identification d'une direction de descente commune à tous les critères en un point Y^0 non optimal au sens de Pareto. On est conduit à généraliser la méthode classique du gradient au contexte multiobjectif en utilisant cette direction pour la descente. On prouve que l'algorithme converge alors vers un point de conception Pareto-stationnaire.

Version française abrégée

On se focalise sur l'optimisation concourante de n critères $\{J_i(Y)\}$ supposés fonctions régulières du vecteur de conception $Y \in \Omega \subset \mathbb{R}^N$ (Ω : domaine admissible ; $J_i \in C^1(\Omega)$; $i = 1, \dots, n \leqslant N$).

On considère un point Y^0 au centre d'une boule ouverte de Ω. Ayant d'abord posé $u_i = \nabla J_i(Y^0)$ (gradients de critères), on propose de définir la notion de « Pareto-stationnarité » par (1). On établit alors que cette propriété constitue une condition nécessaire d'optimalité au sens de Pareto [4]. (La démonstration est conduite selon le rang r de la famille de vecteurs $\{u_i\}$; on traite rapidement les cas simples $r = 0$, et 1 ; puis le cas $0 \leqslant r \leqslant n-1 \leqslant N-1$, pour lequel on pose la relation de dépendance linéaire (2) ; on démontre alors que l'hypothèse $\mu_k < 0$ conduirait à une contradiction avec celle d'optimalité de Pareto ; la conclusion s'en déduit ; enfin, on montre que $r = n$ est impossible car chaque critère est optimal sous la contrainte des autres, (4), ce qui implique (5), contradictoire avec l'hypothèse sur le rang.)

E-mail address: Jean-Antoine.Desideri@inria.fr.

1631-073X/$ – see front matter
http://dx.doi.org/10.1016/j.crma.2012.03.014

On se place ensuite dans le cas inverse où Y^0 n'est pas Pareto-stationnaire, donc pas non plus Pareto-optimal. On introduit l'enveloppe convexe des gradients $\bar{\mathrm{U}}$ défini en (6). Cet ensemble est un fermé convexe, fermeture de l'intérieur U de $\bar{\mathrm{U}}$ constitué des combinaisons convexes strictes ($\alpha_i > 0$ $(\forall i)$). Il admet donc un unique élément de plus petite norme, noté ω. On établit le Lemme 2.1 selon lequel le produit scalaire (u, ω) d'un élément quelconque $u \in \bar{\mathrm{U}}$ avec ω est au moins égal à la constante $\|\omega\|^2$. On en déduit le résultat principal suivant :

Théorème 0.1. *Soit Y^0 le centre d'une boule ouverte de l'espace de conception Ω dans lequel les critères admettent des gradients continus. On note ω l'élément de plus petite norme de l'enveloppe convexe $\bar{\mathrm{U}}$ définie en* (6)*. Deux cas sont possibles* :

(i) *ou bien $\omega = 0$, et le point de conception Y^0 est Pareto-stationnaire* ;
(ii) *ou bien $\omega \neq 0$, et en $Y = Y^0$, le vecteur $-\omega$ définit une direction de descente commune à tous les critères* ; *si de plus ω appartient à l'intérieur U de $\bar{\mathrm{U}}$, les dérivées de Fréchet (u_i, ω) $(i = 1, \dots, n)$, et plus généralement le produit scalaire (u, ω) $(u \in \bar{\mathrm{U}})$, sont égaux à la constante $\|\omega\|^2$.*

S'ensuit très naturellement, la définition de l'*Algorithme de Descente à Gradients Multiples* (*MGDA*) qui généralise à l'optimisation concourante la méthode classique du gradient en utilisant la direction de descente $-\omega$. On suppose que le pas est optimisé de manière à ce que tous les critères diminuent à chaque itération. En supposant les critères positifs et infinis à l'infini, on démontre [2] alors le

Théorème 0.2. *Lorsque l'itération de MGDA est infinie, elle génère une sous-suite convergente vers un point de conception Pareto-stationnaire.*

On termine par plusieurs remarques à caractère computationnel, et une illustration numérique.

Remarque 1. Les résultats formels s'appliquent directement à l'optimisation sous contraintes à condition de considérer les gradients projetés sur le sous-espace tangent aux surfaces de contraintes.

Remarque 2. Afin de déterminer l'elément de plus petite norme ω, ou de manière équivalente, le vecteur de coefficients $\alpha = (\alpha_1, \dots, \alpha_n)$, on fait plusieurs changements de variables. D'abord on pose : $\alpha_i = \sigma_i^2$ $(\forall i = 1, \dots, n)$ afin de satisfaire la condition de positivité. En conséquence $\sigma = (\sigma_1, \dots, \sigma_n)$ appartient à la sphère unité de $\mathbb{R}^n$, que l'on paramétrise par le biais de $n-1$ coordonnées sphériques : $\sigma_i = \sin\phi_{i-1} . \prod_{j=i}^{n-1} \cos\phi_j$. Enfin, comme seul compte σ_i^2, α_i s'exprime par (8) où : $c_0 = 0$, $c_j = \cos^2\phi_j \in [0, 1]$. Ainsi, l'identification de ω est ramenée à la recherche des $n-1$ coefficients $\{c_j\}$ dans $[0, 1]^{n-1}$. Par exemple, pour $n = 4$ critères, 3 paramètres c_1, c_2, c_3 tous dans $[0, 1]$ sont ajustés après avoir posé (9).

Remarque 3. Sans normalisation préalable des gradients, on s'attend à ce la direction de l'élément de norme minimale ω soit principalement influencée par les vecteurs de petites normes de la famille, comme le suggère le cas $n = 2$ de la Fig. 1. Or, au cours de l'optimisation, ces vecteurs sont souvent ceux associés aux critères qui ont déjà atteint un très bon degré de convergence. Cette direction peut correspondre à celle d'un chemin très direct vers le front de Pareto, mais on peut doûter qu'elle corresponde au choix le plus judicieux pour une optimisation multiobjectif bien équilibrée. Nos recherches actuelles portent sur l'analyse de différentes procédures de normalisation conçues pour s'opposer à cette tendance. On étudie notamment les définitions alternatives données par (10) (k : index d'itération ; $\delta > 0$, petit). La première formule correspond à une normalisation standard dont le mérite principal réside dans sa stabilité ; la deuxième est conçue pour que les premières variations logarithmiques des critères soient égales lorsque ω appartient à l'intérieur de l'enveloppe convexe ($\omega \in \mathrm{U}$) ; les deux dernières sont inspirées de la méthode de Newton (dans l'hypothèse où $\lim J_i = 0$ pour la première). Cette question reste ouverte.

1. Introduction: Pareto-stationarity

Our focus is on the concurrent optimization of n criteria $\{J_i(Y)\}$ supposed to be smooth functions of the design vector $Y \in \Omega \subset \mathbb{R}^N$ (Ω: admissible domain; $J_i \in C^1(\Omega)$, $i = 1, \dots, n \leqslant N$).

Consider an admissible point $Y^0 \in \Omega$. If this design-point is Pareto-optimal (see [4] for classical concepts), it is not possible to reduce the local value of any criterion without increasing at least one of the other criteria. In the inverse case, assuming locally smooth criteria, descent directions common to all criteria do exist. We propose to identify one such direction appropriately.

Before developing our construction, we propose the following notion of Pareto-stationarity:

Definition 1.1 *(Pareto-stationarity).* Let Y^0 be a design-point at the center of an open ball in the admissible domain Ω in which the n criteria $J_i(Y)$ $(1 \leqslant i \leqslant n \leqslant N)$ are smooth. Let $u_i = \nabla J_i(Y^0)$ be the local gradients. The design-point Y^0 is said to be Pareto-stationary iff there exists a convex combination of the gradient-vectors, $\{u_i\}$, that is equal to zero:

$$\sum_{i=1}^{n} \alpha_i u_i = 0, \qquad \alpha_i \geqslant 0 \quad (\forall i), \qquad \sum_{i=1}^{n} \alpha_i = 1. \tag{1}$$

The hypotheses of this definition being made, we have:

Lemma 1.2. *If the design-point Y^0 is Pareto-optimal, it is Pareto-stationary.*

Proof. Let r be the rank of the family of gradient-vectors, $r = \mathrm{rank}(\{u_i\}_{1\leqslant i\leqslant n}) = \mathrm{dimSp}(\{u_i\}_{1\leqslant i\leqslant n})$, and let us examine the different possible cases depending on the value of r.

If $r = 0$, all the gradient-vectors are equal to zero and the result is trivial.

If $r = 1$, the gradient-vectors are colinear, say $u_i = \beta_i \mathrm{u}$, where u denotes some unit vector and the coefficients β_i's are not all equal to zero. Then, a perturbation $\delta Y = -\varepsilon \mathrm{u}$ about Y^0 would cause for sufficiently small $\varepsilon > 0$, the perturbation $\delta J_i = -\varepsilon\beta_i + O(\varepsilon^2)$. Thus, if all the coefficients β_i's were of the same sign, say positive, at least one criterion would diminish while the other ones would remain unchanged to $O(\varepsilon^2)$, and this would be a contradiction with the hypothesis that Y^0 is Pareto-optimal. Hence, necessarily, both strictly positive and negative coefficients are present, and perhaps some are equal to zero. Assume, for example, $\beta_1\beta_2 < 0$ and let: $\alpha_1 = -\beta_2/(\beta_1 - \beta_2)$, $\alpha_2 = \beta_1/(\beta_1 - \beta_2)$ so that: $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_1 + \alpha_2 = 1$, $\alpha_1 u_1 + \alpha_2 u_2 = 0$; this establishes the result for $r = 1$.

Now consider the more general case where $2 \leqslant r \leqslant n-1$. Possibly after a permutation of indices, a linear combination of the first $r+1$ $(r+1 \leqslant n)$ gradient-vectors is equal to zero:

$$u_1 + \sum_{k=2}^{r+1} \mu_k u_k = 0. \tag{2}$$

Then we claim that $\mu_k \geqslant 0$ $(\forall k \geqslant 2)$. To establish this, assume instead that $\mu_2 < 0$, and define: $\mathrm{V} = \mathrm{Sp}(\{u_i\}_{3\leqslant i\leqslant r+1})$. Then $\dim \mathrm{V} \leqslant r-1 \leqslant n-2 \leqslant N-2$, and $\dim \mathrm{V}^\perp \geqslant 2$. Let $\omega \in \mathrm{V}^\perp$, arbitrary. Taking the scalar product of (2) with ω yields the following Fréchet derivatives:

$$\forall \omega \in \mathrm{V}^\perp, \quad J'_{2,\omega} = \gamma J'_{1,\omega} \quad (\gamma = -1/\mu_2 > 0). \tag{3}$$

If the above relation were a trivial equality $(0 = \gamma \times 0)$ for all $\omega \in \mathrm{V}^\perp$, the vectors u_1 and u_2 would be in V; consequently, so would be the whole family of gradient-vectors $\{u_i\}$ $(1 \leqslant i \leqslant n)$, which is of rank r by assumption; this is not possible since $\dim \mathrm{V} \leqslant r-1 < r$. Hence, for some $\omega \in \mathrm{V}$, (3) holds and $J'_{1,\omega} \neq 0$. Suppose for example that $J'_{1,\omega} > 0$. Then an infinitesimal perturbation of the design vector Y in the direction of $-\omega$ causes a reduction of both criteria J_1 and J_2 and leaves the other criteria unchanged to second order. This is in contradiction with the assumption that Y^0 belongs to the Pareto set. Hence, we have instead $\mu_2 \geqslant 0$, and for the same reason: $\forall k \geqslant 2$: $\mu_k \geqslant 0$. Then we define $\mu_1 = 1$ and for all i $(1 \leqslant i \leqslant n)$: $\alpha_i = \mu_i / \sum_{k=1}^n \mu_k$. This establishes the result for $2 \leqslant r \leqslant n-1$.

Finally consider the case $r = n$, and let $C_k = J_k(Y^0)$ $(1 \leqslant k \leqslant n)$. Then for some index i (in fact all), Y^0 is the solution of the following minimization problem subject to inequality constraints:

$$\min_Y J_i(Y) \quad \text{subject to:} \quad g_k(Y) := J_k(Y) - C_k \leqslant 0 \quad (\forall k \neq i). \tag{4}$$

Assume $i = 1$ to fix the ideas. Hence, the Lagrangian $\mathrm{L} = J_1(Y) + \sum_{k=2}^n \lambda_k g_k(Y)$ in which the λ_k's $(2 \leqslant k \leqslant n)$ are the Lagrange multipliers, is stationary at $Y = Y^0$:

$$u_1 + \sum_{k=2}^{n} \lambda_k u_k = 0. \tag{5}$$

With our sign conventions, the Lagrange multiplier λ_k is nonnegative, and when $g_k < 0$, it is equal to zero (see e.g. [3]); but this is not essential here. Instead note that (5) is in contradiction with the hypothesis $r = n$, which we now reject. Therefore, $r \leqslant n-1$, and all possible cases have been examined. □

Thus, in general, for smooth unconstrained criteria, Pareto-stationarity is a necessary condition for Pareto-optimality. Inversely, if the smooth criteria $J_i(Y)$ $(1 \leqslant i \leqslant n)$ are not Pareto-stationary at a given design-point Y^0, descent directions common to all criteria exist. We now examine how can such a direction be identified. Denoting $(.,.)$ the usual scalar product in $\mathbb{R}^N$, the problem is equivalent to finding a vector $\omega \in \mathbb{R}^N$ such that: $(\nabla J_i(Y^0), \omega) \geqslant 0$ $(\forall i = 1, \dots, n)$. Then, $-\omega$ is one such descent direction. Noting that an arbitrary normalization of the gradients leaves the problem unchanged, the condition is restated as follows: $(u_i, \omega) \geqslant 0$ $(\forall i = 1, \dots, n)$ where now $u_i = \nabla J_i(Y^0)/S_i$, and the definition of the strictly-positive scaling factors $\{S_i\}$ $(i = 1, \dots, n)$ will be examined ultimately.

This problem admits a general solution in the convex hull of the family of vectors $\{u_i\}$ $(i = 1, \ldots, n)$:

$$\bar{\mathsf{U}} = \left\{ u \in \mathbb{R}^N \;\middle|\; u = \sum_{i=1}^{n} \alpha_i u_i;\ \alpha_i \geqslant 0\ (\forall i = 1, \ldots, n);\ \sum_{i=1}^{n} \alpha_i = 1 \right\}. \tag{6}$$

This result is established in the next section and then exploited to define a *Multiple-Gradient Descent Algorithm* (*MGDA*) that generalizes the *steepest-descent method* [3] to multiobjective optimization. The algorithm is then proved to converge to a Pareto-stationary design-point.

2. Main results

The convex hull $\bar{\mathsf{U}}$ is a closed set, closure of its interior U, made of the elements of $\bar{\mathsf{U}}$ associated with strictly-positive coefficients $\{\alpha_i\}$ $(i = 1, \ldots, n)$. Hence the norm admits at least one realization of a minimum in $\bar{\mathsf{U}}$. Secondly, $\bar{\mathsf{U}}$ is evidently convex, and the minimum is unique.

Proof. Suppose there were two realizations of the minimum norm in $\bar{\mathsf{U}}$: $\|\omega_1\| = \|\omega_2\|$. Since $\bar{\mathsf{U}}$ is convex, $\forall \epsilon \in [0, 1]$, $u = (1 - \epsilon)\omega_1 + \epsilon\omega_2 \in \bar{\mathsf{U}}$; therefore: $\|u\|^2 \geqslant \|\omega_1\|^2$, that is: $(\omega_1 + \epsilon\omega_{12}, \omega_1 + \epsilon\omega_{12}) \geqslant (\omega_1, \omega_1)$ where $\omega_{12} = \omega_2 - \omega_1$. Hence $2\epsilon(\omega_1, \omega_{12}) + \epsilon^2(\omega_{12}, \omega_{12}) \geqslant 0$. Then $(\omega_1, \omega_{12}) \geqslant 0$, since otherwise a contradiction would appear for sufficiently small ϵ. Consequently, for $\epsilon = 1$, the inequality is strict unless $\omega_{12} = 0$; but then $u = \omega_2$ and the equality should hold; thus the only possibility is that $\omega_{12} = 0$. □

Let us denote ω the minimum-norm element in $\bar{\mathsf{U}}$: $\omega = \mathrm{Argmin}_{u \in \bar{\mathsf{U}}} \|u\|$. Then the following holds:

Lemma 2.1. *For all* $u \in \bar{\mathsf{U}}$: $(u, \omega) \geqslant \|\omega\|^2$.

Proof. Let $v = u - \omega$. Since $\bar{\mathsf{U}}$ is convex, $\forall \epsilon \in [0, 1]$, $\omega + \epsilon v \in \bar{\mathsf{U}}$ and $(\omega + \epsilon v, \omega + \epsilon v) \geqslant (\omega, \omega)$. Hence $2\epsilon(\omega, v) + \epsilon^2(v, v) \geqslant 0$. Consequently, $(\omega, v) \geqslant 0$ since otherwise a contradiction would appear for sufficiently small ϵ. Replacing v by $u - \omega$ then gives the result. □

Theorem 2.2. *With the same setting as in Lemmas* 1.2 *and* 2.1*, two cases are possible*:

(i) *either* $\omega = 0$*, and the design-point* Y^0 *is Pareto-stationary*;
(ii) *or* $\omega \neq 0$*, and at* $Y = Y^0$*, the vector* $-\omega$ *defines a descent direction common to all the criteria*; *if additionally* ω *belongs to the interior* U *of* $\bar{\mathsf{U}}$*, the Fréchet derivatives* (u_i, ω) $(i = 1, \ldots, n)$*, and more generally the scalar product* (u, ω) $(u \in \bar{\mathsf{U}})$*, are equal to the constant* $\|\omega\|^2$.

Proof. The conclusions of this theorem are reformulations of previous lemmas, except for the statement concerning the scalar product (u, ω) in the second case, when additionally $\omega \in \mathsf{U}$ (and not simply $\bar{\mathsf{U}}$). Under these assumptions, the element ω is the solution to the following minimization problem:

$$\omega = u = \sum_{i=1}^{n} \alpha_i u_i, \quad \alpha = \mathrm{Argmin}\,\mathrm{j}(u), \quad \mathrm{j}(u) = (u, u), \quad \sum_{i=1}^{n} \alpha_i = 1 \tag{7}$$

since by hypothesis, none of the inequality constraints is saturated $(\alpha_i > 0\ (\forall i))$. Consequently, using the vector $\alpha \in \mathbb{R}^n$ as the finite-dimensional variable, the Lagrangian writes: $\mathrm{L}(\alpha, \lambda) = \mathrm{j} + \lambda(\sum_{i=1}^{n} \alpha_i - 1)$ and the optimality conditions satisfied by the vector α are the following: $\partial \mathrm{L}/\partial \alpha_i = 0$ $(\forall i)$, $\partial \mathrm{L}/\partial \lambda = 0$. These equations imply that for all indices i: $\partial \mathrm{j}/\partial \alpha_i + \lambda = 0$. But, $\mathrm{j}(u) = (u, u)$ and for $u = \omega = \sum_{i=1}^{n} \alpha_i u_i$, one has: $\partial \mathrm{j}/\partial \alpha_i = 2(\partial u/\partial \alpha_i, u) = 2(u_i, \omega) = -\lambda$. Hence, the Fréchet derivatives $(u_i, \omega) = -\lambda/2$ $(\forall i)$ are equal. Finally, for any $u \in \bar{\mathsf{U}}$, say $u = \sum_{i=1}^{n} \mu_i u_i$ where $\mu_i \geqslant 0$ $(\forall i)$ and $\sum_{i=1}^{n} \mu_i = 1$, one has: $(u, \omega) = \sum_{i=1}^{n} \mu_i (u_i, \omega) = -\lambda/2$, a constant, necessarily equal to $\|\omega\|^2$. □

This theorem leads us to generalize the classical *steepest-descent method* to multiobjective optimization by defining the *Multiple-Gradient Descent Algorithm* (*MGDA*) as the iteration whose generic step is:

(i) Compute the normalized gradient-vectors $u_i = \nabla J_i(Y^0)/S_i$, and determine the minimum-norm element ω in the convex hull $\bar{\mathsf{U}}$. If $\omega = 0$ (or sufficiently small), stop.
(ii) Otherwise, determine the step-size h as the largest strictly-positive real number for which all the functions $j_i(t) = J_i(Y^0 - t\omega)$ $(1 \leqslant i \leqslant n)$ are monotone-decreasing over the interval $[0, h]$.
(iii) Reset Y^0 to $Y^0 - h\omega$.

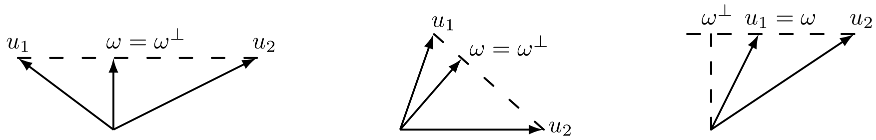

Fig. 1. Case $n = 2$: possible positions of vector ω with respect to the two gradients u_1 and u_2.

Since at each iteration of the *MGDA*, all the criteria diminish, we refer to this process as a *cooperative-optimization* method. The above *MGDA* can stop after a finite number of iterations if a Pareto-stationary design-point has been reached. Otherwise, we have the following:

Theorem 2.3. *If the sequence of iterates $\{Y^k\}$ generated by the MGDA is infinite, it admits a subsequence that converges to a Pareto-stationary design-point.*

Proof. Without loss of generality, it can be assumed that the multiobjective problem has been formulated by means of strictly-positive criteria that are infinite when $\|Y\|$ is (see [2]). Then:

– Since the sequence of values of any considered criterion, say $\{J_1(Y^k)\}$, is positive and monotone-decreasing, it is bounded.
– Since $J_1(Y)$ is infinite whenever $\|Y\|$ is infinite, the sequence of iterates $\{Y^k\}$ is also bounded, and this implies the convergence of a subsequence to some $Y^\star$.

The limiting design-point $Y^\star$ is necessarily Pareto-stationary since otherwise, if $\omega(Y^\star) \neq 0$, a new iteration would potentially diminish each criterion of a finite amount, and a better iterate be found. □

Remark 1. The formal results can be applied straightforwardly to the case of constrained optimization provided the gradients are projected onto the subspace tangent to the constraint surfaces.

Remark 2. For the determination of the minimum-norm element ω, or, equivalently, the vector of coefficients $\alpha = (\alpha_1, \ldots, \alpha_n)$, appropriate changes of variables are made. First, one lets $\alpha_i = \sigma_i^2$ ($\forall i = 1, \ldots, n$) to satisfy the positivity condition. As a result $\sigma = (\sigma_1, \ldots, \sigma_n)$ is to be found on the unit sphere of $\mathbb{R}^n$, which is then parameterized using $n-1$ spherical coordinates: $\sigma_i = \sin\phi_{i-1} \cdot \prod_{j=i}^{n-1} \cos\phi_j$. But since only σ_i^2 matters, α_i is expressed as follows:

$$\alpha_i = (1 - c_{i-1}) \cdot \prod_{j=i}^{n-1} c_j \quad (i = 1, \ldots, n;\ j = 1, \ldots, n-1) \tag{8}$$

where: $c_0 = 0$, $c_j = \cos^2\phi_j \in [0, 1]$. Thus, the identification of ω is realized by a search in $[0, 1]^{n-1}$. For example, with $n = 4$ criteria, 3 parameters $c_1,\ c_2,\ c_3$ all in $[0, 1]$ are adjusted optimally after letting:

$$\alpha_1 = c_1c_2c_3, \qquad \alpha_2 = (1 - c_1)c_2c_3, \qquad \alpha_3 = (1 - c_2)c_3, \qquad \alpha_4 = (1 - c_3). \tag{9}$$

Remark 3. If the gradients are not normalized ($S_i = 1,\ \forall i$), the direction of the minimum-norm element ω is expected to be mostly influenced by the gradients of small norms in the family, as the case $n = 2$ illustrated in Fig. 1 suggests. In the course of the iterative optimization, these vectors are often associated with the criteria that have already achieved a fair degree of convergence. If this direction may yield a very direct path to the Pareto front, one may question whether it is adequate for a well-balanced multiobjective iteration. On-going research is focused on analyzing various normalization procedures to circumvent this undesirable trend. These correspond to the following possible alternative definitions:

$$u_i = \frac{\nabla J_i(Y^0)}{\|\nabla J_i(Y^0)\|}, \frac{\nabla J_i(Y^0)}{J_i(Y^0)}, \frac{J_i(Y^0)}{\|\nabla J_i(Y^0)\|^2}\nabla J_i\big(Y^0\big), \text{ or } \frac{\max(J_i^{(k-1)}(Y^0) - J_i^{(k)}(Y^0), \delta)}{\|\nabla J_i(Y^0)\|^2}\nabla J_i\big(Y^0\big) \tag{10}$$

(k: iteration number; $\delta > 0$, small). The first formula is a standard normalization: it has the merit of providing a stable definition; the second realizes equal logarithmic first variations of the criteria whenever ω belongs to the interior U of the convex hull since then, the Fréchet derivatives (u_i, ω) are equal; the last two are inspired from Newton's method (assuming $\lim J_i = 0$ for the first). This question is still open.

Remark 4. If the original gradients $\{\nabla J_i(Y^0)\}$ $(i = 1, \ldots, n)$ are linearly independent, applying the Gram–Schmidt orthogonalization process to them produces the family $u_i' = [\nabla J_i(Y^0) - \sum_{k<i} \mu_{i,k} u_k']/S_i$, where: $S_1 = \|\nabla J_1(Y^0)\|$, and

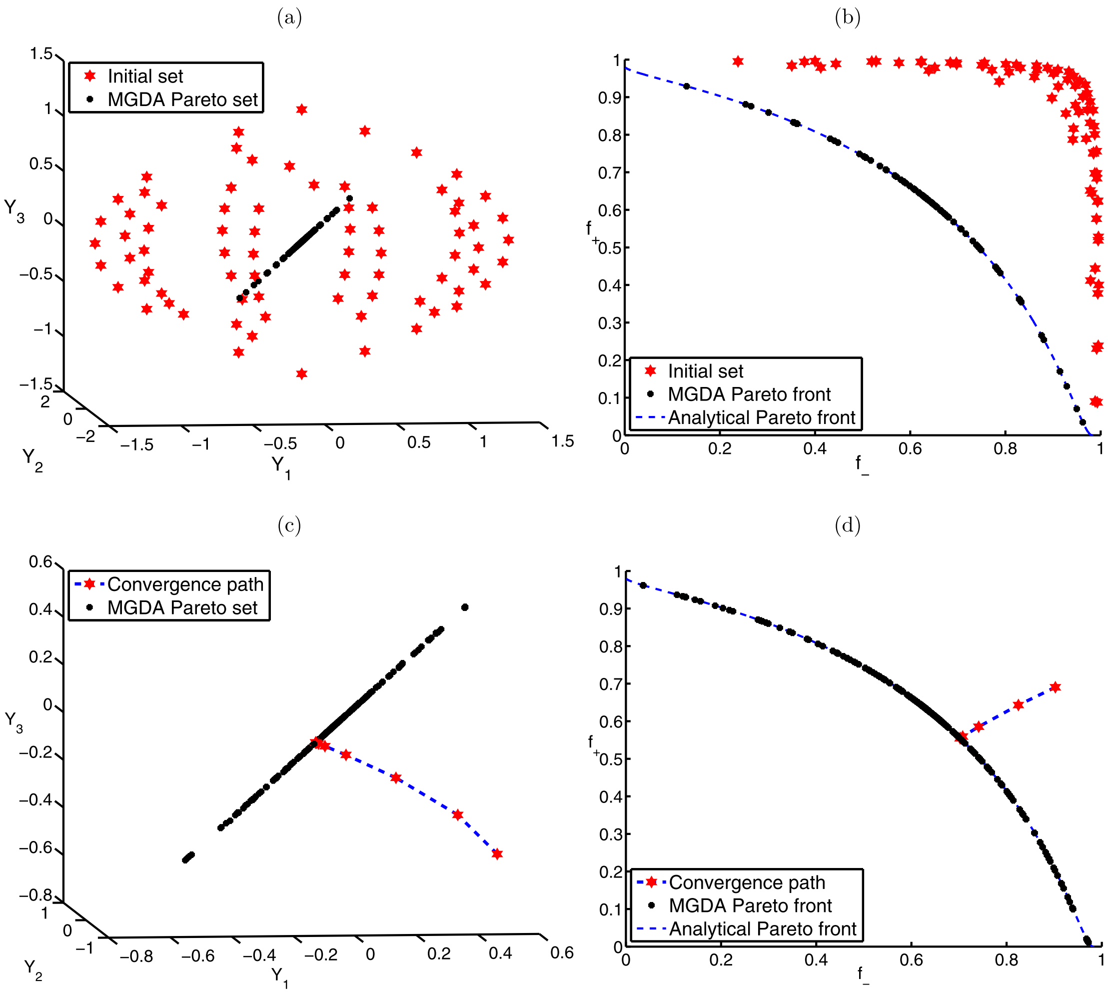

Fig. 2. Fonseca testcase – convergence of the MGDA in the concurrent optimization of the following two functions of the 3-component vector Y: $f_{\pm}(Y) = 1 - \exp[-\sum_{i=1}^{3} (Y_i \pm 1/\sqrt{3})^2]$. In (a) and (b): distribution of initial and Pareto-optimal points, in design and function space respectively; in (c) and (d): illustration of one particular path. The exact Pareto set is indicated by a dashed line. (Results from [5].)

$\forall i \geqslant 2$, $\forall k < i$: $\mu_{i,k} = (\nabla J_i(Y^0), u'_k)$, $S_i > 0$ and $(u'_i, u'_i) = 1$. The minimum-norm element of the convex hull associated with this new family is $\omega' = \sum_{i=1}^{n} u'_i/n$ and for all i: $(u'_i, \omega') = \|\omega'\|^2 = 1/n$. Consequently, $(\nabla J_i(Y^0), \omega') = S'_i/n$ where $S'_i = S_i + \sum_{k<i} \mu_{i,k}$. Then, if the condition $S'_i > 0$ ($\forall i$) is satisfied, $-\omega'$ is also an appropriate descent direction.

Lastly, on Fig. 2, we illustrate the capability of the MGDA to identify sharply a Pareto set in a testcase from the literature (Fonseca testcase in [1]) in which the front is nonconvex in function space.

References

[1] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2) (2002) 182–197.
[2] J.-A. Désidéri, Multiple-gradient descent algorithm (MGDA), INRIA Research Report No. 6953, June 2009, http://hal.inria.fr/inria-00389811.
[3] Ph.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, New York, London, 1986.
[4] K.M. Miettinen, Nonlinear Multiobjective Optimization, Kluwer Academic Publ., Boston, London, Dordrecht, 1999.
[5] A. Zerbinati, J.-A. Désidéri, R. Duvigneau, Comparison between MGDA and PAES for multi-objective optimization, INRIA Research Report No. 7667, June 2011, http://hal.inria.fr/inria-00605423.