MGDA II: A direct method for calculating a descent direction common to several criteria

Jean-Antoine Désidéri

RESEARCH REPORT
N° 7922
April 2012
Project-Team Opale

ISSN 0249-6399   ISRN INRIA/RR--7922--FR+ENG

MGDA II: A direct method for calculating a descent direction common to several criteria

Jean-Antoine Désidéri*

Project-Team Opale

Research Report n° 7922 — April 2012 — 8 pages

Abstract: This report is a sequel of the publications [1] [3] [2]. We consider the multiobjective optimization problem of the simultaneous minimization of n ($n \geq 2$) criteria, $\{J_i(Y)\}_{(i=1,\dots,n)}$, assumed to be smooth real-valued functions of the design vector $Y \in \Omega \subset \mathbb{R}^N$ ($n \leq N$) where Ω is the (open) admissible domain of $\mathbb{R}^N$ over which these functions admit gradients. Given a design point $Y^0 \in \Omega$ that is not Pareto-stationary, we introduce the gradients $\{J_i'\}_{(i=1,\dots,n)}$ at $Y = Y^0$, and assume them to be linearly independent. We also consider the possible "scaling factors", $\{S_i\}_{(i=1,\dots,n)}$ ($S_i > 0\,,\ \forall i$), as specified appropriate normalization constants for the gradients. Then we show that the Gram-Schmidt orthogonalization process, if conducted with a particular calibration of the normalization, yields a new set of orthogonal vectors $\{u_i\}_{(i=1,..,n)}$ spanning the same subspace as the original gradients; additionally, the minimum-norm element of the convex hull corresponding to this new family, ω, is calculated explicitly, and the Fréchet derivatives of the criteria in the direction of ω are all equal and positive. This direct process simplifies the implementation of the previously-defined *Multiple-Gradient Descent Algorithm (MGDA)*.

Key-words: multiobjective optimization, descent direction, convex hull, Gram-Schmidt orthogonalization process

* INRIA Research Director, Opale Project-Team Head

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

MGDA II: Une méthode directe de calcul de direction de descente de plusieurs critères

Résumé : Ce rapport est une suite des publications [1] [3] [2]. On considère le problème d'optimisation multiobjectif dans lequel on cherche à minimiser n ($n \geq 2$) critères, $\{J_i(Y)\}_{(i=1,\dots,n)}$, supposés fonctions régulières d'un vecteur de conception $Y \in \Omega \subset \mathbb{R}^N$ ($n \leq N$) où Ω est le domaine (ouvert) admissible, partie de $\mathbb{R}^N$ dans laquelle les critères admettent des gradients. Étant donné un point de conception $Y^0 \in \Omega$ qui n'est pas Pareto-stationnaire, on introduit les gradients $\{J'_i\}_{(i=1,\dots,n)}$ en $Y = Y^0$, et on les suppose linéairement indépendants. On considère également un ensemble de "facteurs d'échelles", $\{S_i\}_{(i=1,\dots,n)}$ ($S_i > 0$, $\forall i$), spécifiés par l'utilisateur, et considérés comme des constantes appropriées de normalisation des gradients. On montre alors que le processus d'orthogonalisation de Gram-Schmidt, lorsqu'on le conduit avec une calibration bien spécifique de la normalisation, produit un ensemble de vecteurs orthogonaux $\{u_i\}_{(i=1,..,n)}$ qui engendrent le même sous-espace que les gradients d'origine; de plus, l'élément de plus norme de l'enveloppe convexe de cette nouvelle famille, ω, se calcule explicitement, et les dérivées de Fréchet des critères dans la direction de ω sont égales et positives. Ce processus direct simplifie la mise en œuvre de l'*Algorithme de Descente à Gradients Multiples (MGDA)* défini précédemment.

Mots-clés : optimisation multiobjectif, direction de descente, enveloppe convexe, processus d'orthogonalisation de Gram-Schmidt

1 Introduction

We consider the context of the simultaneous minimization of n ($n \geq 2$) criteria, $\{J_i(Y)\}_{(i=1,\dots,n)}$, assumed to be smooth real-valued functions of the design vector $Y \in \Omega \subset \mathbb{R}^N$ ($n \leq N$) where Ω is the (open) admissible domain of $\mathbb{R}^N$ over which these functions admit gradients. Let $Y^0 \in \Omega$, and let:

$$J_i' = \nabla J_i(Y^0) \qquad (i = 1, \dots, n;\ J_i' \in \mathbb{R}^N) \tag{1}$$

be the gradients at the design-point Y^0.

In [1] and [2], we have introduced the local notion of Pareto-stationarity, defined as the existence of a convex combination of the gradients that is equal to zero. We established there that Pareto-stationarity was a necessary condition to Pareto-optimality. If inversely, the Pareto-stationarity condition is not satisfied, vectors having positive scalar products with all the gradients $\{J_i'\}_{(i=1,\dots,n)}$ exist. We focus on the question of identifying such vectors.

Consider a family $\{u_i\}_{(i=1,\dots,n)}$ of n vectors of $\mathbb{R}^N$, and recall the definition of their convex hull:

$$\overline{\mathsf{U}} = \left\{ u \in \mathbb{R}^N \ / \ u = \sum_{i=1}^{n} \alpha_i u_i\,;\ \alpha_i \geq 0\ (\forall i)\,;\ \sum_{i=1}^{n} \alpha_i = 1 \right\} \tag{2}$$

This set is closed and convex. Hence it admits a unique element ω of minimum norm. We established that:

$$\forall u \in \overline{\mathsf{U}} : \quad \left(u, \omega\right) \geq \|\omega\|^2 \tag{3}$$

By letting

$$u_i = J_i' \qquad (i = 1, \dots, n) \tag{4}$$

and identifying the corresponding vector ω, we were able to conclude that either $\omega = 0$ and the design-point Y^0 is Pareto-stationary, or $-\omega$ is a descent direction common to all criteria.

This observation has led us to propose the *Multiple-Gradient Descent Algorithm (MGDA)* that is an iteration generalizing the classical steepest-descent method to the context of multiobjective optimization. At a given iteration, the design point Y^0 is updated by a step in the direction opposite to ω:

$$\delta Y^0 = -\rho\omega \tag{5}$$

Assuming the stepsize ρ is optimized, *MGDA* converges to a Pareto-stationary design point [1] [2].

Thus, *MGDA* provides a technique to identify Pareto sets when gradients are available, as demonstrated in [3].

We have also shown that the convex hull was isomorphic to the positive part of the sphere of $\mathbb{R}^{n-1}$ (independently of N) and it can be parameterized by $n-1$ spherical coordinates. Hence, in the first version of our method, when $n > 2$, we proposed to identify the vector ω by actually finding numerically the minimum of $\|u\|^2$ in $\overline{\mathsf{U}}$ by optimizing the spherical coordinates, or more precisely, their cosines squared that are $n-1$ independent parameters varying in the interval [0,1]. This minimization can be conducted trivially when n is small, but can become difficult for large n.

In this new report, we propose a variant of *MGDA* in which the direction ω is found by a direct process, assuming the family of gradients $\{J_i'\}_{(i=1,\dots,n)}$ is linearly independent.

2 Direct calculation of the minimum-norm element

In [2], the following remark was made:

Remark 1
If the gradients are not normalized, 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 Figure 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. Some on-going research is focused on analyzing various normalization procedures to circumvent this undesirable trend. In these alternatives, the gradient J_i' is replaced by one of the following formulas:

$$\frac{J_i'}{\left\|J_i'\right\|},\ \frac{J_i'}{J_i(Y^0)},\ \frac{J_i(Y^0)}{\left\|J_i'\right\|^2} J_i',\ \text{or}\ \frac{\max\left(J_i^{(k-1)}(Y^0) - J_i^{(k)}(Y^0), \delta\right)}{\left\|J_i'\right\|^2} J_i' \tag{6}$$

(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 logorithmic first variations of the criteria whenever ω belongs to the interior U of the convex hull since then, the Fréchet derivatives $\left(u_i, \omega\right)$ are equal; the last two are inspired from Newton's method (assuming $\lim J_i = 0$ for the first). This question is still open.

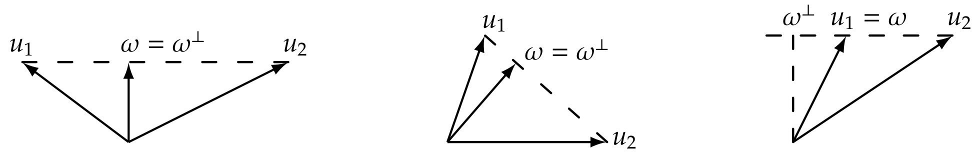

Figure 1: Case $n = 2$: possible positions of vector ω with respect to the two gradients u_1 and u_2

Thus, accounting for the above remark, we consider a given set of strictly-positive "scaling factors" $\{S_i\}_{(i=1,\dots,n)}$, intended to normalize the gradients appropriately. We apply the Gram-Schmidt orthogonalization process to the gradients with the following special calibration of the normalization:

$$u_1 = \frac{J_1'}{A_1} \tag{7}$$

where $A_1 = S_1$, and, for $i = 2, 3, \dots, n$:

$$u_i = \frac{J_i' - \sum_{k<i} c_{i,k} u_k}{A_i} \tag{8}$$

where:

$$\forall k < i\ :\ c_{i,k} = \frac{\left(J_i', u_k\right)}{\left(u_k, u_k\right)} \tag{9}$$

and

$$A_i = \begin{cases} S_i - \sum_{k<i} c_{i,k} & \text{if nonzero} \\ \varepsilon_i S_i & \text{otherwise} \end{cases} \tag{10}$$

for some arbitrary, but small ε_i ($0 < |\varepsilon_i| \ll 1$).

Then :

In general, the vectors $\{u_i\}_{(i=1,\dots,n)}$ are not of norm unity. But they are orthogonal, and this property makes the calculation of the minimum-norm element of the convex hull, ω, direct. For this, note that:

$$\omega = \sum_{i=1}^{n} \alpha_i u_i \tag{11}$$

and

$$\|\omega\|^2 = \sum_{i=1}^{n} \alpha_i^2 \, \|u_i\|^2 \; . \tag{12}$$

To determine the coefficients $\{\alpha_i\}_{(i=1,\ldots,n)}$, anticipating that ω belongs to the interior of the convex hull, the inequality constraints are ignored, and the following Lagrangian is made stationary:

$$\mathsf{L}(\alpha,\lambda) = \|\omega\|^2 - \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right) = \sum_{i=1}^{n} \alpha_i^2 \, \|u_i\|^2 - \lambda \left(\sum_{i=1}^{n} \alpha_i - 1 \right) \tag{13}$$

This gives:

$$\frac{\partial \mathsf{L}}{\partial \alpha_i} = 2\, \|u_i\|^2 \, \alpha_i - \lambda \Longrightarrow \alpha_i = \frac{\lambda}{2\, \|u_i\|^2} \tag{14}$$

The equality constraint, $\sum_{i=1}^{n} \alpha_i = 1$, then gives:

$$\frac{\lambda}{2} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\|u_i\|^2}} \tag{15}$$

and finally:

$$\alpha_i = \frac{1}{\|u_i\|^2 \sum_{j=1}^{n} \frac{1}{\|u_j\|^2}} = \frac{1}{1 + \sum_{j \neq i} \frac{\|u_i\|^2}{\|u_j\|^2}} < 1 \tag{16}$$

which confirms that ω does belong to the interior of the convex hull, so that:

$$\forall i \; : \; \alpha_i \, \|u_i\|^2 = \frac{\lambda}{2} \qquad \text{(a constant)}, \tag{17}$$

and:

$$\forall k \; : \; \left(u_k, \omega \right) = \alpha_k \, \|u_k\|^2 = \frac{\lambda}{2} \, . \tag{18}$$

Now:

$$\left(J_i', \omega \right) = \left(A_i u_i + \sum_{k<i} c_{i,k} u_k, \omega \right) = \left(A_i + \sum_{k<i} c_{i,k} \right) \frac{\lambda}{2} = S_i \frac{\lambda}{2} \, , \quad \text{or } S_i (1 + \varepsilon_i) \frac{\lambda}{2} \quad (\forall i) \, . \tag{19}$$

Lastly, convening that $\varepsilon_i = 0$ in the regular case ($S_i \neq \sum_{k<i} c_{i,k}$), and otherwise by modifying slightly the definition of the scaling factor according to

$$S_i' = (1 + \varepsilon_i) S_i \, , \tag{20}$$

the following holds:

$$\boxed{\left(S_i'^{-1} J_i', \omega \right) = \frac{\lambda}{2} \quad (\forall i)} \tag{21}$$

that is, the same positive constant.

3 Conclusion

We have considered the multiobjective optimization problem of the simultaneous minimization of n ($n \geq 2$) criteria, $\{J_i(Y)\}_{(i=1,\dots,n)}$, assumed to be smooth real-valued functions of the design vector $Y \in \Omega \subset \mathbb{R}^N$ ($n \leq N$) where Ω is the (open) admissible domain of $\mathbb{R}^N$ over which these functions admit gradients. Given a design point $Y^0 \in \Omega$ that is not Pareto-stationary, we have introduced the gradients $\{J_i'\}_{(i=1,\dots,n)}$ at $Y = Y^0$, and assumed them to be linearly independent. We have also considered the possible "scaling factors", $\{S_i\}_{(i=1,\dots,n)}$ ($S_i > 0\,,\ \forall i$), as specified appropriate normalization constants for the gradients. Then we have shown that the Gram-Schmidt orthogonalization process, if conducted with a particular calibration of the normalization yields a new set of orthogonal vectors $\{u_i\}_{(i=1,..,n)}$ spanning the same subspace as the original gradients; additionally, the minimum-norm element of the convex hull corresponding to this new family, ω, is calculated explicitly, and the Fréchet derivatives of the criteria in the direction of ω are all equal and positive.

This new result has led us to reformulate the definition of the *MGDA*, in which the descent direction common to all criteria is now calculated by a direct process.

Finally, we make the following two remarks:

Remark 2
In the exception case where $\sum_{k<i} c_{i,k} = S_i$, for some i, the corresponding vector u_i has a large *norm. Consequently, this vector has a* weak *influence on the definition of ω.*

Remark 3
As the Pareto set is approached, the family of gradients becomes closer to linear dependence, as the Pareto-stationarity condition requires. At this stage, this variant may experience numerical difficulties, making the more robust former approach preferable.

References

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

[2] Jean-Antoine Désidéri. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. *Comptes rendus - Mathématique*, 1(4867), 2012. DOI: 10.1016/j.crma.2012.03.014.

[3] Adrien Zerbinati, Jean-Antoine Desideri, and Régis Duvigneau. Comparison between MGDA and PAES for Multi-Objective Optimization. Research Report RR-7667, INRIA, June 2011.

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