Foreword

Our friend Jacques Périaux likes particularly:
His family, his friends, Paris, Normandy, carrot-free salad, Barcelona-Houston-Jyväskylä-Nanjing, . . . , the rest of the world, airplanes, fishing, playing bridge, holding the podium, etc

But, also:

- PDE's
- compressible aerodynamics
- domain-decomposition methods
- optimization, particularly with evolutionary algorithms
- cooperation (among algorithms, but also people)
- competition (among algorithms, but also people)
- HPC algorithms
- etc
MULTIPLE GRADIENT DESCENT ALGORITHM (MGDA) FOR MULTIOBJECTIVE OPTIMIZATION

Application to domain partitioning

Jean-Antoine Désidéri
Equipe-Projet INRIA OPALÉ
Centre de Sophia Antipolis Méditerranée
http://www-sop.inria.fr/opale

Optimization and partial-differential equations for industrial systems
Dedicated to the jubilee of Jacques Périaux
June 14-15, 2012, University of Jyväskylä, Finland
Outline

1 MGDA
   Pareto-stationarity
   Descent direction
   Main results
   Practicalities
   Mathematical test-cases
   Aerodynamics

2 Application to domain-partitioning model problem
   Model problem
   Quasi-Newton Method
   Basic MGDA
   MGDA-II

3 Conclusions

4 Cooperation and Competition
Constructing a gradient-based cooperative algorithm

The basic question:
Knowing the gradients, $\nabla J_i(Y^0)$, of $n$ criteria, that are smooth functions of the design vector $Y \in \mathbb{R}^N$ ($n \leq N$) at a given starting design-point $Y^0$, can we define a vector $\omega \in \mathbb{R}^N$, in the direction of which the Frechet derivatives of all criteria have the same sign,

$$\nabla J_i(Y^0).\omega \geq 0 \quad (\forall i = 1, 2, \ldots, n)$$?

If so, $-\omega$ is locally a descent direction common to all criteria.

Answer:
Yes!, if $Y^0$ is not Pareto-optimal
Notion of Pareto-stationarity

Classically

For real-valued functions of $N$ variables:

Regularity + Optimality $\implies$ Stationarity

Proposition 1: Pareto Stationarity

For $\mathbb{R}^n$-valued functions of $N$ variables ($n \leq N$):

If the design point $Y^0$ is Pareto-optimal, there exists a convex combination of the gradients equal to 0:

$$\sum_{i=1}^{n} \alpha_i \nabla J_i(Y^0) = 0, \quad \alpha_i \geq 0 \ (\forall i), \quad \sum_{i=1}^{n} \alpha_i = 1$$

(assuming regularity about $Y^0$). Thus:

Regularity + Pareto-optimality $\implies$ Pareto-stationarity
Hypothesis:
Design-point $Y^0$ is at the center of an open ball in design space in which all criteria are smooth; in this ball, there is no design-point that dominates $Y^0$ in efficiency.

\[ u_i = \nabla J_i(Y^0) \quad (i = 1, \ldots, n) \]
\[ Y^0, u_i \in \mathbb{R}^N \quad (n \leq N) \]
\[ r = \text{rank} \{u_1, u_2, \ldots, u_n\} \]

Trivial cases first

- $r = 0$: $u_i = 0$ ($\forall i$)
- $r = 1$: $u_i = \beta_i u$ ($\forall i$)

Pareto-optimality then requires that for at least 2 indices, say 1 and 2, the coefficients are nonzero and of opposite sign, and:

\[ \alpha_1 u_1 + \alpha_2 u_2 = 0, \quad \alpha_1 = -\beta_2/(\beta_1 - \beta_2), \quad \alpha_2 = \beta_1/(\beta_1 - \beta_2) \]

so that $0 \leq \alpha_1 \leq 1, 0 \leq \alpha_2 \leq 1, \alpha_1 + \alpha_2 = 1$. 

Proof
Proof (cont’d)

General case: $2 \leq r \leq n - 1$

Possibly after reordering:

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

To establish that $\mu_k \geq 0 \ (\forall k \geq 2)$, assume inversely that $\mu_2 < 0$, and define:

$$V = \text{Sp}\left(\{u_i\}_{3\leq i \leq r+1}\right).$$

Then $\dim V \leq r - 1 \leq n - 2 \leq N - 2$, and $\dim V^\perp \geq 2$, and

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

- $0 = \gamma \times 0$: cannot be the case for all $\omega$
- Thus pick $\omega$ s.t.: $J'_{1,\omega} \neq 0, J'_{2,\omega} \neq 0$: contradiction with Pareto-optimality assumption.

Conclusion:

$$\forall k : \mu_k \geq 0$$

Then let $\alpha_k = \mu_k / \sum_i \mu_i$ so that:

$$\sum_k \alpha_k u_k = 0, \ 0 \leq \alpha_k \leq 1 (\forall k), \ \sum_k \alpha_k = 1$$
Proof (end)

Last case: \( r = n \)
Design-point \( Y^0 \) minimizes at least one criterion subject to the constraints of the others; let \( C_k = J_k(Y^0) \):

\[
Y^0 = \text{Argmin}_Y J_i(Y) \quad \text{subject to: } g_k(Y) := J_k(Y) - C_k = 0 \ (\forall k \neq i)
\]

Assume \( i = 1 \), and express stationarity of the Lagrangian
\[
L = J_1(Y) + \sum_{k=2}^{n} \lambda_k g_k(Y):
\]

\[
\begin{align*}
    u_1 + \sum_{k=2}^{n} \lambda_k u_k &= 0
\end{align*}
\]

and this contradicts the assumption on the rank.
This case is incompatible with the assumptions.

Conclusion
In all the cases compatible with the assumptions, a Pareto-stationarity condition has been identified. □
**General principle**

**Proposition 2: Minimum-norm element in the convex hull**

Let \( \{ u_i \}_{i=1}^{n} \) be a family of \( n \) vectors in \( \mathbb{R}^N \), and \( \mathbb{U} \) the so-called *convex hull* of this family, i.e. the following set of vectors:

\[
\mathbb{U} = \left\{ u \in \mathbb{R}^N / u = \sum_{i=1}^{n} \alpha_i u_i ; \alpha_i \geq 0 (\forall i) ; \sum_{i=1}^{n} \alpha_i = 1 \right\}
\]

Then, \( \mathbb{U} \) admits a unique element of minimal norm, \( \omega \), and the following holds:

\[
\forall u \in \mathbb{U} : (u, \omega) \geq \| \omega \|^2
\]

in which \((u, v)\) denotes the usual scalar product of the vectors \( u \) and \( v \).
Existence and uniqueness of $\omega$

Existence $\iff \overline{U}$ closed. Uniqueness $\iff \overline{U}$ convex.

To establish uniqueness, let $\omega_1$ and $\omega_2$ be two realisations of the minimum:

$$\|\omega_1\| = \|\omega_2\| = \text{Argmin}_{u \in \overline{U}} \|u\|$$

Then, since $\overline{U}$ is convex, $\forall \varepsilon \in [0, 1]$, $\omega_1 + \varepsilon r_{12} \in \overline{U}$ ($r_{12} = \omega_2 - \omega_1$) and:

$$\|\omega_1 + \varepsilon r_{12}\| \geq \|\omega_1\|$$

Square both sides, and remplace by scalar products:

$$\forall \varepsilon \in [0, 1]: (\omega_1 + \varepsilon r_{12}, \omega_1 + \varepsilon r_{12}) - (\omega_1, \omega_1) \geq 0$$

Then

$$\forall \varepsilon \in [0, 1]: 2\varepsilon (r_{12}, \omega_1) + \varepsilon^2 (r_{12}, r_{12}) \geq 0$$

As $\varepsilon \to 0^+$, this condition requires that $(r_{12}, \omega_1) \geq 0$; but then, for $\varepsilon = 1$:

$$\|\omega_2\|^2 - \|\omega_1\|^2 = 2(r_{12}, \omega_1) + (r_{12}, r_{12}) > 0$$

unless $r_{12} = 0$, i.e. $\omega_2 = \omega_1$. \[\square\]

Remark: The identification of the element $\omega$ is equivalent to the constrained minimization of a quadratic form in $\mathbb{R}^n$, and not $\mathbb{R}^N$. 

First consequence

\[ \forall u \in \overline{U} : \quad (u, \omega) \geq \|\omega\|^2 \]

Proof:
Let \( u \in \overline{U} \), and \( r = u - \omega \).
\( \forall \varepsilon \in [0, 1], \omega + \varepsilon r \in \overline{U} \); hence:

\[
\|\omega + \varepsilon r\|^2 - \|\omega\|^2 = 2\varepsilon (r, \omega) + \varepsilon^2 (r, r) \geq 0
\]

and this requires that:

\[
(r, \omega) = (u - \omega, \omega) \geq 0
\]
Second consequence

If additionally, $\omega \in U$ (and not only in $\overline{U}$), then:

$$\forall u \in \overline{U} : (u, \omega) = \text{const.} = ||\omega||^2$$

**Proof:**

Identify $\omega$ by the coefficients $\{\alpha_i\}_{i=1,2,...,n}$.

If $\omega \in U$, none of the inequality constraints is saturated ($\forall i, \alpha_i > 0$). Hence the solution corresponds to the minimum of the quadratic form $||\omega||^2 = ||\sum_{i=1}^{n} \alpha_i u_i||^2$ subject to the unique equality constraint: $\sum_{i=1}^{n} \alpha_i = 1$. The Lagrangian is formed with a unique multiplier $\lambda$:

$$L(\alpha, \lambda) = \frac{1}{2} \left( \sum_{i=1}^{n} \alpha_i u_i, \sum_{i=1}^{n} \alpha_i u_i \right) + \lambda \left( \sum_{i=1}^{n} \alpha_i - 1 \right)$$

The stationarity condition requires that:

$$\forall i, \quad \frac{\partial L}{\partial \alpha_i} = (u_i, \omega) + \lambda = 0 \implies (u_i, \omega) = -\lambda = \text{const.} = ||\omega||^2$$

By convex combination, the result extends straightforwardly to the whole $\overline{U}$. $\square$
Application to the case of gradients

Suppose \( u_i = \nabla J_i(Y^0) \ (\forall i) \); then:

- either \( \omega \neq 0 \), and all criteria admit positive Fréchet derivatives in the direction of \( \omega \) (all equal if \( \omega \) belongs to the interior \( U \))
- or \( \omega = 0 \), and the current design point \( Y^0 \) is Pareto stationary:

\[
\exists \{\alpha_i\}_{i=1,2,...,n} \ (\alpha_i \geq 0, \forall i; \sum_{i=1}^{n} \alpha_i = 1) \text{ so that } \sum_{i=1}^{n} \alpha_i u_i = 0
\]
**Proposition 3: Common Descent Direction**

By virtue of Propositions 1 and 2 in the particular case where $u_i = \nabla J_i(Y^0)/K_i$ ($K_i$: adjustable normalization constant; $K_i > 0$), two situations are possible at $Y = Y^0$:

- either $\omega = 0$, and the design point $Y^0$ is Pareto-stationary (or Pareto-optimal);

- or $\omega \neq 0$, and $-\omega$ is a descent direction for all criteria $\{J_i(x)\}_{i=1,...,n}$; additionally, if strict inequalities $\alpha_i > 0$ ($\forall i$) hold, the Frechet derivatives $(u_i, \omega)$ are all equal to $||\omega||^2$.

*MGDA*: substitute vector $\omega$ to the single-criterion gradient in the steepest-descent method.

**Proposition 4: Convergence**

Certain normalization provisions being made, in $\mathbb{R}^N$, the Multiple Gradient Descent Algorithm (MGDA) converges to a Pareto-stationary point.
Define the criteria to be positive (possibly apply an exp-transform), and $\infty$ at $\infty$ (possibly add an exploding term outside of the working ball):

$$\forall i, \quad J_i(Y) \to \infty \text{ as } \|Y\| \to \infty.$$

Assume these criteria to be continuous.

If the iteration stops in a finite number of steps, a Pareto-stationary point has been reached.

Otherwise, the iteration continues indefinitely, generating an infinite sequence of design points, $\{Y^k\}$. The corresponding sequences of criteria are infinite, positive and monotone decreasing. They are bounded. Hence, the sequence of iterates, $\{Y^k\}$, is itself bounded and it admits a subsequence converging to say $Y^\star$. Necessarily, $Y^\star$ is a Pareto-stationary point. To establish this, assume instead that $\omega^\star$, which corresponds to $Y^\star$, is nonzero. Then for each criterion, there exists a stepsize $\rho$ for which, the variation is finite. These criteria are in finite number. Hence, the smallest $\rho$ will cause a finite variation to all criteria, and this is in contradiction with the fact that only infinitely-small variations of the criteria are realized from $Y^\star$. $\square$
Remark 1: practical
determination of vector $\omega$
in the convex hull

Problem to be solved in $\bar{U} \subset \mathbb{R}^N$

$$\omega = \text{Argmin}_{u \in \bar{U}} \|u\|$$

$$\bar{U} = \left\{ u \in \mathbb{R}^N \mid u = \sum_{i=1}^{n} \alpha_i u_i ; \alpha_i \geq 0 (\forall i) ; \sum_{i=1}^{n} \alpha_i = 1 \right\}$$

Usually, but not necessarily: $N \geq n$.

Parameterization

Let

$$\alpha_i = \sigma_i^2 \quad (i = 1, \ldots, n)$$

to satisfy trivially the inequality constraints ($\alpha_i \geq 0$, $\forall i$), and transform the equality constraint, $\sum_{i=1}^{n} \alpha_i = 1$ into

$$\sum_{i=1}^{n} \sigma_i^2 = 1 \iff \sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n) \in S_n$$

where $S_n$ is the unit sphere of $\mathbb{R}^n$ and precisely not $\mathbb{R}^N$. 
Determining vector $\omega$ (cont’d)

The sphere is easily parameterized

using trigonometric functions of $n - 1$ independent arcs $\phi_1, \phi_2, \ldots, \phi_{n-1}$:

$$
\begin{align*}
\sigma_1 &= \cos \phi_1 \cdot \cos \phi_2 \cdot \cos \phi_3 \cdot \ldots \cdot \cos \phi_{n-1} \\
\sigma_2 &= \sin \phi_1 \cdot \cos \phi_2 \cdot \cos \phi_3 \cdot \ldots \cdot \cos \phi_{n-1} \\
\sigma_3 &= \sin \phi_1 \cdot \sin \phi_2 \cdot \cos \phi_3 \cdot \ldots \cdot \cos \phi_{n-1} \\
&\vdots \\
\sigma_{n-1} &= 1 \cdot 1 \cdot \ldots \cdot \sin \phi_{n-2} \cdot \cos \phi_{n-1} \\
\sigma_n &= 1 \cdot 1 \cdot \ldots \cdot 1 \cdot \sin \phi_{n-1}
\end{align*}
$$

(Consider only : $\phi_i \in [0, \pi/2], \forall i$ and set : $\phi_0 = \pi/2$.)

Let $c_i = \cos^2 \phi_i$ \quad ($i = 1, \ldots, n$)

and get:

$$
\begin{align*}
\alpha_1 &= c_1 \cdot c_2 \cdot c_3 \cdot \ldots \cdot c_{n-1} \\
\alpha_2 &= (1 - c_1) \cdot c_2 \cdot c_3 \cdot \ldots \cdot c_{n-1} \\
\alpha_3 &= 1 \cdot (1 - c_2) \cdot c_3 \cdot \ldots \cdot c_{n-1} \\
&\vdots \\
\alpha_{n-1} &= 1 \cdot 1 \cdot \ldots \cdot (1 - c_{n-2}) \cdot c_{n-1} \\
\alpha_n &= 1 \cdot 1 \cdot \ldots \cdot 1 \cdot (1 - c_{n-1})
\end{align*}
$$

($c_0 = 0$, and $c_i \in [0, 1]$ for all $i \geq 1$).
Determining vector $\omega$ (end)

The convex hull is thus parameterized in

$$[0, 1]^{n-1} \quad (n: \text{number of criteria})$$

independently of $N$ (dimension of design space).

For example, with $n = 4$ criteria:

$$\begin{align*}
\alpha_1 &= c_1 c_2 c_3 \\
\alpha_2 &= (1 - c_1) c_2 c_3 \\
\alpha_3 &= (1 - c_2) c_3 \\
\alpha_4 &= (1 - c_3)
\end{align*}$$

$$(c_1, c_2, c_3) \in [0, 1]^3$$
Remark 2: appropriate normalization of the gradients may reveal to be essential: \( u_i = \nabla J_i(Y^0)/K_i \)

Case \( n = 2 \)

- **Without gradient normalization**
  
  Then, \( \omega = \omega^\perp \) unless the angle \( \langle u_1, u_2 \rangle \) is acute, and the norms \( ||u_1||, ||u_2|| \) are very different; in that case, \( \omega \) is equal the one of smaller norm.

- **With normalization:** The equality \( \omega = \omega^\perp \) holds automatically \( \implies \) \boxed{\text{EQUAL FRECHET DERIVATIVES}}

**General Case**

**IF** the vectors \( \{u_i\} \) are **NOT NORMALIZED**, those of smaller norms are more influential to determine the direction of vector \( \omega \)
Recall

\[
\begin{align*}
   u_i &= \frac{\nabla J_i(Y^0)}{K_i} \ (K_i > 0) \implies \omega \\
   \delta Y &= -\rho \omega \implies \delta J_i = (\nabla J_i, \delta Y) = -\rho K_i (u_i, \omega) \\
   (u_i, \omega) &= \text{const. if } \omega \text{ does not lie on the "edge" of convex hull}
\end{align*}
\]

- Standard:

\[ u_i = \frac{\nabla J_i(Y^0)}{\|\nabla J_i(Y^0)\|} \]

- Equal logarithmic variations (whenever \( \omega \) is not on edge):

\[ u_i = \frac{\nabla J_i(Y^0)}{J_i(Y^0)} \]

- Newton-inspired when \( \lim J_i = 0 \):

\[ u_i = \frac{J_i}{\|\nabla J_i(Y^0)\|^2} \nabla J_i(Y^0) \]

- Newton-inspired when \( \lim J_i \neq 0 \):

\[ u_i = \frac{\max\left(J_i^{(k-1)} - J_i^{(k)}, \delta\right)}{\|\nabla J_i(Y^0)\|^2} \nabla J_i(Y^0), \ k: \text{iteration no.}, \ \delta > 0 \text{ (small)} \]
Is standard normalization sufficient to guarantee that \( \omega = \omega^\perp \)?

(yielding equal Frechet derivatives)

No, if \( n > 2 \) :
Basic testcase: minimize the functions

\[ f(x, y) = 4x^2 + y^2 + xy \quad g(x, y) = (x - 1)^2 + 3(y - 1)^2 \]
Basic testcase
Convergence from different initial design points

**DESIGN SPACE**

- MGDA
- MGDA-II

**FUNCTION SPACE**

- (0.5, 2)
- (1.5, 2.5)
- (-1.5, -2.5)
Fonseca testcase

Minimize 2 functions of 3 variables:

\[ f_{1,2}(x_1, x_2, x_3) = 1 - \exp \left( - \sum_{i=1}^{3} \left( x_i \pm \frac{1}{\sqrt{3}} \right)^2 \right) \]
Fonseca testcase (cont’d)
Convergence from initial design points over a sphere
Fonseca testcase (end)

Compare MGDA with Pareto Archived Evolution Strategy (PAES)
Wing-shape optimization

Wave-drag minimization in conjunction with lift maximization

Metamodel-assisted MGDA (Adrien Zerbinati)

• Wing-shape geometry
  extruded from airfoil geometry (initial airfoil : NACA0012), 10 parameters

• Two-point/two-objective optimization
  1. Subsonic Eulerian flow : \( M_\infty = 0.3, \alpha = 8^\circ \), for \( C_L \) maximization
  2. Transonic Eulerian flow : \( M_\infty = 0.83, \alpha = 2^\circ \), for \( C_D \) minimization

• Algorithm - Initial database of 40 design points evolves as follows :
  1. Evaluate each new point by two 3D Eulerian flow computations
  2. Construct surrogate models for \( C_L \) and \( C_D \) (using the entire dataset), and perform MGDA to convergence
  3. Enrich the database with new points, if appropriate

After 6 loops, the database is made of 227 design points
(554 flow computations)
Wing-shape optimization

Convergence of dataset

6 loops

-LIFT (subsonic)

DRAG (transonic)
Wing-shape optimization
Low-drag quasi-Pareto-optimal shape

$M_\infty = 0.3, \alpha = 8^o$

$M_\infty = 0.83, \alpha = 2^o$
Wing-shape optimization

High-lift quasi-Pareto-optimal shape

\( M_\infty = 0.3, \alpha = 8^\circ \)

\( M_\infty = 0.83, \alpha = 2^\circ \)
Wing-shape optimization

Intermediate quasi-Pareto-optimal shape

SUBSONIC

M_{\infty} = 0.3, \alpha = 8^o

TRANSONIC

M_{\infty} = 0.83, \alpha = 2^o
Some references

Reports from INRIA Open Archive:


Other

MGDA and domain-partitioning

J.-A. Désidéri

Outline

1. MGDA
   - Pareto-stationarity
   - Descent direction
   - Main results
   - Practicalities
   - Mathematical test-cases
   - Aerodynamics

2. Application to domain-partitioning model problem
   - Model problem
   - Quasi-Newton Method
   - Basic MGDA
     - MGDA -II

3. Conclusions

4. Cooperation and Competition
Dirichlet problem

Discrete solution by standard 2nd-order finite-difference method over $40 \times 40$ quadrangular mesh

$$-\Delta u = f \quad \text{over } \Omega = [-1, 1] \times [-1, 1]$$

$$u = 0 \quad (\Gamma = \partial \Omega)$$
Domain partitioning

Function value controls at 4 interfaces

yielding 4 sub-domain Dirichlet problems with 2 controlled interfaces each
Interface jumps

Formal expression

- over $\gamma_1$ ($0 \leq x \leq 1$; $y = 0$):
  \[ s_1(x) = \frac{\partial u}{\partial y}(x, 0^+) - \frac{\partial u}{\partial y}(x, 0^-) = \left[ \frac{\partial u_1}{\partial y} - \frac{\partial u_4}{\partial y} \right](x, 0); \]

- over $\gamma_2$ ($x = 0$; $0 \leq y \leq 1$):
  \[ s_2(y) = \frac{\partial u}{\partial x}(0^+, y) - \frac{\partial u}{\partial x}(0^-, y) = \left[ \frac{\partial u_1}{\partial x} - \frac{\partial u_2}{\partial x} \right](0, y); \]

- over $\gamma_3$ ($-1 \leq x \leq 0$; $y = 0$):
  \[ s_3(x) = \frac{\partial u}{\partial y}(x, 0^+) - \frac{\partial u}{\partial y}(x, 0^-) = \left[ \frac{\partial u_2}{\partial y} - \frac{\partial u_3}{\partial y} \right](x, 0); \]

- over $\gamma_4$ ($x = 0$; $-1 \leq y \leq 0$):
  \[ s_4(y) = \frac{\partial u}{\partial x}(0^+, y) - \frac{\partial u}{\partial x}(0^-, y) = \left[ \frac{\partial u_4}{\partial x} - \frac{\partial u_3}{\partial x} \right](0, y). \]

Approximation

by 2nd-order one-sided finite-differences
Functionals and matching condition

Interface functionals

\[ J_i = \int_{\gamma_i} \frac{1}{2} s_i^2 w \, d\gamma_i := J_i(v) \]

that is, explicitly:

\[ J_1 = \int_0^1 \frac{1}{2} s_1(x)^2 w(x) \, dx \]
\[ J_2 = \int_0^1 \frac{1}{2} s_2(y)^2 w(y) \, dy \]
\[ J_3 = \int_{-1}^0 \frac{1}{2} s_3(x)^2 w(x) \, dx \]
\[ J_4 = \int_{-1}^0 \frac{1}{2} s_4(y)^2 w(y) \, dy \]

\((w(t) \, (t \in [0, 1]))\) is an optional weighting function, and \(w(-t) = w(t)\).

Matching condition

\[ J_1 = J_2 = J_3 = J_4 = 0 \]
Adjoint problems

Eight Dirichlet sub-problems

\[
\begin{align*}
\Delta p_i &= 0 \quad (\Omega_i) \\
p_i &= 0 \quad (\partial \Omega_i \setminus \gamma_i) \\
p_i &= s_i \mathbf{w} \quad (\gamma_i)
\end{align*}
\]

\[
\begin{align*}
\Delta q_i &= 0 \quad (\Omega_i) \\
q_i &= 0 \quad (\partial \Omega_i \setminus \gamma_{i+1}) \\
q_i &= s_{i+1} \mathbf{w} \quad (\gamma_{i+1})
\end{align*}
\]

Green's formula

\[
\int_{\gamma_i} s_i \mathbf{u}_{in} \mathbf{w} = \int_{\gamma_i} p_i \mathbf{n} \mathbf{v}_i' + \int_{\gamma_{i+1}} p_i \mathbf{n} \mathbf{v}_{i+1}'
\]

and

\[
\int_{\gamma_{i+1}} s_{i+1} \mathbf{u}_{in} \mathbf{w} = \int_{\gamma_i} q_i \mathbf{n} \mathbf{v}_i' + \int_{\gamma_{i+1}} q_i \mathbf{n} \mathbf{v}_{i+1}'
\]
Functional gradients

\[
\begin{align*}
J'_1 &= \int_0^1 s_1(x) s'_1(x) w(x) \, dx = \int_0^1 \left[ \frac{\partial u'_1}{\partial y} - \frac{\partial u'_4}{\partial y} \right] (x, 0) w(x) \, dx \\
&= \int_0^1 \frac{\partial (p_1 - q_4)}{\partial y} (x, 0) v'_1(x) \, dx + \int_0^1 \frac{\partial p_1}{\partial x} (0, y) v'_2(y) \, dy + \int_{-1}^0 \frac{\partial q_4}{\partial x} (0, y) v'_4(y) \, dy \\
J'_2 &= \int_0^1 s_2(y) s'_2(y) w(y) \, dy = \int_0^1 \left[ \frac{\partial u'_1}{\partial x} - \frac{\partial u'_2}{\partial x} \right] (0, y) w(y) \, dy \\
&= \int_0^1 \frac{\partial q_1}{\partial y} (x, 0) v'_1(x) \, dx + \int_0^1 \frac{\partial (q_1 - p_2)}{\partial x} (0, y) v'_2(y) \, dy + \int_{-1}^0 \frac{\partial p_2}{\partial y} (x, 0) v'_3(x) \, dx \\
J'_3 &= \int_{-1}^0 s_3(x) s'_3(x) w(x) \, dx = \int_{-1}^0 \left[ \frac{\partial u'_2}{\partial y} - \frac{\partial u'_3}{\partial y} \right] (x, 0) w(x) \, dx \\
&= - \int_0^1 \frac{\partial q_2}{\partial x} (0, y) v'_2(y) \, dy + \int_{-1}^0 \frac{\partial (q_2 - p_3)}{\partial y} (x, 0) v'_3(x) \, dx - \int_{-1}^0 \frac{\partial p_3}{\partial x} (0, y) v'_4(y) \, dy \\
J'_4 &= \int_{-1}^0 s_4(y) s'_4(y) w(y) \, dy = \int_{-1}^0 \left[ \frac{\partial u'_4}{\partial x} - \frac{\partial u'_3}{\partial x} \right] (0, y) w(y) \, dy \\
&= - \int_0^1 \frac{\partial p_4}{\partial y} (x, 0) v'_1(x) \, dx - \int_{-1}^0 \frac{\partial q_3}{\partial y} (x, 0) v'_3(x) \, dx + \int_{-1}^0 \frac{\partial (p_4 - q_3)}{\partial x} (0, y) v'_4(y) \, dy
\end{align*}
\]

Conclusion:

\[
J'_i = \sum_{j=1}^{4} \int_{\gamma_j} G_{i,j} v'_j \, d\gamma_j \quad (i = 1, \ldots, 4); \quad \{ G_{i,j} \} : \text{partial gradients.}
\]
Other technical details

**Dirichlet sub-problems**

All 12 (4 direct, 4 × 2 adjoint) sub-problems are solved by direct inverse (discrete sine-transform)

\[
 u_h = (\Omega_X \otimes \Omega_Y) (\Lambda_X \oplus \Lambda_Y)^{-1} (\Omega_X \otimes \Omega_Y) f_h
\]

**Integrals**

are approximated by the trapezoidal rule
Reference method

Quasi-Newton Method applied to agglomerated criterion

\[ J = \sum_{i=1}^{4} J_i \]

**Gradient**

\[ \nabla J = \sum_{i=1}^{4} \nabla J_i \]

**Iteration**

\[ \nu^{(\ell+1)} = \nu^{(\ell)} - \rho_{\ell} \nabla J^{(\ell)} \]

**Stepsize**

\[ \delta J = \nabla J^{(\ell)} \cdot \delta \nu^{(\ell)} = -\rho_{\ell} \left\| \nabla J^{(\ell)} \right\|^2 \]

is set to \(-\varepsilon J^{(\ell)}\) by fixing

\[ \rho_{\ell} = \frac{\varepsilon J^{(\ell)}}{\left\| \nabla J^{(\ell)} \right\|^2} \]
Quasi-Newton steepest-descent

\( \epsilon = 1 \)

Convergence history

Asymptotic

Global
Quasi-Newton steepest descent

\[ \varepsilon = 1 \]

History of gradients over 200 iterations

\[ \frac{\partial J}{\partial v_1}, \quad \frac{\partial J}{\partial v_2} \]

\[ \text{DISCRETIZED FUNCTIONAL GRADIENT OF } J = \sum_j J_j \text{ W.R.T. } V_1 \]

\[ \text{DISCRETIZED FUNCTIONAL GRADIENT OF } J = \sum_j J_j \text{ W.R.T. } V_2 \]
Quasi-Newton steepest descent

\( \varepsilon = 1 \)

History of gradients over 200 iterations

\( \frac{\partial J}{\partial v_3} \)

\[ \text{DISCRETIZED FUNCTIONAL GRADIENT OF } J = \sum_i J_i \text{ W.R.T. } V_3 \]

\( \frac{\partial J}{\partial v_4} \)

\[ \text{DISCRETIZED FUNCTIONAL GRADIENT OF } J = \sum_i J_i \text{ W.R.T. } V_4 \]
Quasi-Newton steepest descent

\[ \varepsilon = 1 \]

Discrete solution
Basic *MGDA*

Practical determination of minimum-norm element $\omega$

$$\omega = \sum_{i=1}^{4} \alpha_i u_i$$

$$u_i = \nabla J_i(Y^0)$$

using the following parameterization of the convex hull:

$$
\begin{align*}
\alpha_1 &= c_1 c_2 c_3 \\
\alpha_2 &= (1 - c_1)c_2 c_3 \\
\alpha_3 &= (1 - c_2)c_3 \\
\alpha_4 &= (1 - c_3)
\end{align*}
$$

and

$$(c_1, c_2, c_3) \in [0, 1]^3$$

are discretized by step of 0.01. Best set of coefficients retained.
Basic MGDA

Stepsiz

\[ v^{(\ell+1)} = v^{(\ell)} - \rho^{\ell} \omega^{(\ell)} \]

Iteration

\[ \delta J = \nabla J^{(\ell)} \cdot \delta v^{(\ell)} = -\rho^{\ell} \nabla J^{(\ell)} \cdot \omega \]

Stepsiz

is set to \(-\varepsilon J^{(\ell)}\) by fixing

\[ \rho^{\ell} = \frac{\varepsilon J^{(\ell)}}{\nabla J^{(\ell)} \cdot \omega} \]

In practice: \(\varepsilon = 1\).
Basic MGDA

\[ \epsilon = 1 \]

Asymptotic Convergence

Unscaled

\[ u_i = \nabla J_i \]

Scaled

\[ u_i = \frac{\nabla J_i}{J_i} \]
**Basic MGDA**

$\epsilon = 1$

---

**Global Convergence**

**Unscaled**

$$u_i = \nabla J_i$$

**Scaled**

$$u_i = \frac{\nabla J_i}{J_i}$$

---

**Mathematical test-cases**

Application to domain-partitioning model problem

Model problem

Quasi-Newton Method

Basic MGDA

MGDA-II

Conclusions

Cooperation and Competition
First conclusions

Basic MGDA

- Scaling essential
- Somewhat deceiving convergence

Who is to blame?

- the insufficiently accurate determination of $\omega$;
- the non-optimality of the scaling of gradients;
- the non-optimality of the step-size, the parameter $\varepsilon$ being maintained equal to 1 throughout;
- the large dimension of the design space, here 76 (4 interfaces associated with 19 d.o.f.’s).
**MGDA-II : a direct algorithm**

Valid when gradients are linearly independent

**Prescription of scaling factors for the gradients**

\[ S_i > 0 \quad (i = 1, \ldots, n) \]

In practice: \( S_i = 1/J_i \) for logarithmic gradients.

**Apply Gram-Schmidt in a special way**

\[ u_1 = \frac{J'_1}{A_1} \]

where \( A_1 = S_1 \), and, for \( i = 2, 3, \ldots, n \):

\[ u_i = \frac{J'_i - \sum_{k<i} c_{i,k} u_k}{A_i} \]

where:

\[ \forall k < i : c_{i,k} = \frac{(J'_i, u_k)}{(u_k, u_k)} \]

and

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

\( \varepsilon_i \) arbitrary, but small.
Consequences

Element $\omega$ always in the interior of convex hull

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

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

This implies equal projections of $\omega$ onto $u_k$

$$\forall k : (u_k, \omega) = \alpha_k \|u_k\|^2 = \frac{\lambda}{2}$$

Finally

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

(1)

$(S_i' = (1 + \varepsilon_i) S_i)$, that is, the same positive constant.

EQUAL FRÉCHET DERIVATIVES FORMED WITH SCALED GRADIENTS
Geometrical interpretation

\[ u_i = \frac{J'_i - \sum_{k<i} c_{i,k} u_k}{A_i} \]

\[ O \perp \in \overline{U} \implies \omega = \overline{O} \perp \implies \]

\[ (u_1, \omega) = (u_2, \omega) = (u_3, \omega) = \|\omega\|^2 \]

\[ J'_i = A_i u_i + \sum_{k<i} c_{i,k} u_k \implies \]

\[ (J'_i, \omega) = (A_i + \sum_{k<i} c_{i,k}) \|\omega\|^2 \]

adjusted by normalization thru \( A_i \)
MGDA-II \textit{b}: automatic rescale

Do not let constant $A_i$ be negative; instead define

$$A_i = S_i - \sum_{k=1}^{i-1} c_{i,k}$$

only when this number is strictly-positive.
Otherwise, change the scale $S_i$ to become:

$$S_i = \sum_{k=1}^{i-1} c_{i,k}$$

and set $A_i = \varepsilon_i S_i$, for some small $\varepsilon_i$ ("automatic rescaling").

Same formal conclusion:
the Fréchet derivatives are equal; but the value is much larger, and (at least) one criterion has been rescaled.
Global convergence

Unscaled
\[ u_i = \nabla J_i \]

Scaled
\[ u_i = \frac{\nabla J_i}{J_i} \]
Global convergence

Unscaled

\[ u_i = \nabla J_i \]

automatic rescale on

Scaled

\[ u_i = \frac{\nabla J_i}{J_i} \]

automatic rescale on
Convergence over 500 iterations

Unscaled

\[ u_i = \nabla J_i \]

automatic rescale off

Scaled

\[ u_i = \frac{\nabla J_i}{J_i} \]

automatic rescale on
MGDA and domain-partitioning

J.-A. Désidéri

Outline

1. MGDA
   Pareto-stationarity
   Descent direction
   Main results
   Practicalities
   Mathematical test-cases
   Aerodynamics

2. Application to domain-partitioning model problem
   Model problem
   Quasi-Newton Method
   Basic MGDA
   MGDA-II

3. Conclusions

4. Cooperation and Competition
Conclusions

Theory

Multiple-Gradient Descent Algorithm

- Notion of Pareto stationarity introduced to characterize "standard Pareto-optimal points"
- Basic MGDA: permits to identify a descent direction common to arbitrary number of criteria, knowing the local gradients in number at most equal to the number of design parameters
- MGDA-II: a direct procedure, faster and more accurate, when gradients are linearly independent; variant MGDA-II b recommended
- Proof of convergence of MGDA to Pareto-stationary points
- Capability to identify the Pareto front demonstrated for mathematical test-cases; non-convexity of Pareto front not a problem
- On-going research: scaling, preconditioning, step-size adjustment, what to do to extend MGDA-II to cases of linearly-dependent gradients, ...
Conclusions

Aerodynamics

On-going development of meta-model-assisted MGDA (A. Zerbinati, to be presented at ECCOMAS 2012, Vienna, Austria)

Domain-partitioning model problem

The method works, but it is not as cost-efficient as the standard quasi-Newton method applied to the agglomerated criterion; but the test-case was peculiar in several ways:

- Pareto front and set reduced to a singleton
- large dimension of the design space (4 criteria, 76 design variables)
- robust procedure to adjust the step-size needed
- determination of $\omega$ in the basic method should be made more accurately, perhaps using iterative refinement
- scaling of gradients found important but never really analyzed completely (logarithmic gradients appropriate for vanishing criteria)
- general preconditioning not yet clear (on-going)

The MGDA-II direct variant found faster and more robust; with automatic rescaling procedure on, MGDA-II b seems to exhibit quadratic convergence.
<table>
<thead>
<tr>
<th>MGDA and domain-partitioning</th>
<th>J.-A. Désidéri</th>
</tr>
</thead>
</table>

**Outline**

1. **MGDA**
   - Pareto-stationarity
   - Descent direction
   - Main results
   - Practicalities
   - Mathematical test-cases
   - Aerodynamics

2. Application to domain-partitioning model problem
   - Model problem
   - Quasi-Newton Method
   - Basic *MGDA*
   - *MGDA*-II

3. Conclusions

4. Cooperation and Competition
Flirting with the Pareto front
MGDA is combined with a locally-adapted Nash game based on a territory splitting that preserves to second-order the Pareto-stationarity condition

Cooperation **AND** Competition

- **COOPERATION (MGDA)**:
The design point is steered to the Pareto front.

- **COMPETITION (Nash game)**: at start (from the Pareto front):
  \[
  \alpha_1 \nabla J_1 + \alpha_2 \nabla J_2 = 0; \]
  then let:
  \[
  J_A = \alpha_1 J_1 + \alpha_2 J_2, \quad \text{et} \quad J_B = J_2,
  \]
  and adapt territory-splitting to best preserve \( J_A \); then:
  the trajectory remains tangent to the Pareto front.