Multicriteria Optimization
Some continuous and discrete dynamics

Guillaume Garrigos

Institut de Mathématiques et de Modélisation de Montpellier
Universidad Tecnica Federico Santa Maria

Sestri-Levante: Franco/Italian workshop
8-12 September 2014
- $H$ is an Hilbert space,
- $f_i : H \rightarrow \mathbb{R}$ are Lipschitz continuous on bounded sets.
- $K \subset H$ is a closed convex non empty set of constraints,
- One of the objective functions is bounded from below.
- $H$ is an Hilbert space,
- $f_i : H \rightarrow \mathbb{R}$ are Lipschitz continuous on bounded sets.
- $K \subset H$ is a closed convex non empty set of constraints,
- One of the objective functions is bounded from below.

One approach, the scalarization method:
chose $0 \leq \theta_i \leq 1$, $\sum_{i=1}^{q} \theta_i = 1$, and minimize $\sum_{i=1}^{q} \theta_i f_i$. 
- $H$ is an Hilbert space,
- $f_i : H \to \mathbb{R}$ are Lipschitz continuous on bounded sets.
- $K \subset H$ is a closed convex non empty set of constraints,
- One of the objective functions is bounded from below.

One approach, the scalarization method:

Chose $0 \leq \theta_i \leq 1$, $\sum_{i=1}^{q} \theta_i = 1$, and minimize $\sum_{i=1}^{q} \theta_i f_i$.

We are looking for the **simultaneous** minimization of the $f_i$'s.
1 Multicriteria analysis

2 Continuous steepest descent dynamic
1 Multicriteria analysis

2 Continuous steepest descent dynamic
Nonsmooth analysis tools

**Directional derivative (of Clarke)**

\[
df(x; d) := \limsup_{t \downarrow 0} \frac{f(x' + td) - f(x')}{t}.
\]

**Subdifferential (of Clarke)**

\[
\partial f(x) := \{ p \in H \mid \langle p, d \rangle \leq df(x; d) \ \forall \ d \in H \}.
\]

**Remark**

If \( f \) is of class \( C^1 \), then
\[
\partial f(x) = \{ \nabla f(x) \} \ \text{and} \ df(x; d) = \langle \nabla f(x), d \rangle.
\]
Nonsmooth analysis tools

Tangent and normal cones

\[ T_K(x) := \text{cl} \{ d \in H \mid \exists \epsilon > 0, \forall t \in ]0, \epsilon[, \ x + td \in K \}. \]

\[ N_K(x) := \{ p \in H \mid \langle p, d \rangle \leq 0 \ \forall d \in T_K(x) \}. \]
We say that $d \in H$ is a descent direction at $x$ if $df_i(x; d) < 0$ holds for all $i = 1..q$.
We say that it is an admissible descent direction if moreover $d \in T_K(x)$. 
Example
We say that a descent direction $d \in H$ is an Armijo direction if there exists $\varepsilon > 0$ such that for all $t \in ]0, \varepsilon [$:

$$\forall i, f_i(x + td) < f_i(x) + \frac{t}{2} df_i(x; d).$$

We say that it is an admissible Armijo direction if moreover $x + td \in K$. 
We say that $x \in K$ is a Pareto if there is no $y \in K$ such that $\forall i f_i(y) \leq f_i(x)$ and $\exists I f_I(y) < f_I(x)$.

We say that $x \in K$ is a weak Pareto if there is no $y \in K$ such that $\forall i f_i(y) < f_i(x)$. 
Pareto equilibrium(s)

We say that $x \in K$ is a Pareto if there is no $y \in K$ such that
$\forall i \ f_i(y) \leq f_i(x)$ and $\exists I \ f_I(y) < f_I(x)$. 

### Pareto equilibrium(s)

- We say that \( x \in K \) is a Pareto if there is no \( y \in K \) such that \( \forall i \, f_i(y) \leq f_i(x) \) and \( \exists i \, f_i(y) < f_i(x) \).

- We say that \( x \in K \) is a weak Pareto if there is no \( y \in K \) s.t. \( \forall i \, f_i(y) < f_i(x) \).
Example

\[ \nabla f_1(x) \nabla f_2(x) \]

Multicriteria optimization - Guillaume Garrigos - Franco/Italian workshop
We say that \( x \in K \) is a Pareto if there is no \( y \in K \) such that
\[
\forall i \ f_i(y) \leq f_i(x) \quad \text{and} \quad \exists l \ f_l(y) < f_l(x).
\]

We say that \( x \in K \) is a weak Pareto if there is no \( y \in K \) s.t.
\[
\forall i \ f_i(y) < f_i(x).
\]

We say that \( x \in K \) is a critical Pareto if
\[
0 \in N_K(x) + \text{Conv}\{\partial f_i(x)\}.
\]
Example

Conv\{\nabla f_i(x)\}
Pareto equilibrium(s)

- We say that \( x \in K \) is a Pareto if there is no \( y \in K \) such that \( \forall i \ f_i(y) \leq f_i(x) \) and \( \exists i \ f_i(y) < f_i(x) \).

- We say that \( x \in K \) is a weak Pareto if there is no \( y \in K \) s.t. \( \forall i \ f_i(y) < f_i(x) \).

- We say that \( x \in K \) is a critical Pareto if \( 0 \in N_K(x) + \text{Conv}\{\partial f_i(x)\} \).
Pareto equilibrium(s)

- We say that \( x \in K \) is a Pareto if there is no \( y \in K \) such that \( \forall i \ f_i(y) \leq f_i(x) \) and \( \exists l \ f_l(y) < f_l(x) \).
- We say that \( x \in K \) is a weak Pareto if there is no \( y \in K \) s.t. \( \forall i \ f_i(y) < f_i(x) \).
- We say that \( x \in K \) is a critical Pareto if \( 0 \in N_K(x) + \text{Conv}\{\partial f_i(x)\} \).

Properties

- Pareto \( \Rightarrow \) weak Pareto \( \Rightarrow \) critical Pareto.
- If the \( f_i \) are convex, then weak Pareto \( \Leftrightarrow \) critical Pareto.
- If the \( f_i \) are strictly convex, then the 3 notions both coincide.
Proposition

The following statements are equivalent:

- $x$ is a critical Pareto point,
- There is no admissible descent direction at $x$,
- There is no admissible Armijo direction at $x$. 
We will consider

1. a continuous dynamic $\dot{u}(t) = s(u(t))$, where $s : K \rightarrow H$ verify
   - $s(u) = 0$ if $u$ is a critical Pareto point,
   - $s(u)$ is an admissible descent direction else.

2. an algorithm $u_{n+1} = u_n + t_n d_n$ where $d_n$ is an admissible Armijo direction.
1 Multicriteria analysis

2 Continuous steepest descent dynamic
Definition

Given $x \in K$, the \textit{multiobjective steepest descent direction} is

$$s(x) := - (N_K(x) + \text{Conv}\{\partial f_i(x)\})^0.$$
Example

\[ -s(x) \text{ Conv}\{\nabla f_i(x)\} \]
Given $x \in K$, the multiobjective steepest descent direction is

$$s(x) := - (N_K(x) + \text{Conv}\left\{\partial f_i(x)\right\})^0.$$
The multiobjective steepest descent direction

Definition

Given $x \in K$, the multiobjective steepest descent direction is

$$s(x) := - (N_K(x) + \text{Conv}\{\partial f_i(x)\})^0.$$

Obviously, $x$ is a Pareto critical iff $s(x) = 0$.
The multiobjective steepest descent direction

Definition

Given $x \in K$, the multiobjective steepest descent direction is

\[ s(x) := - (N_K(x) + \text{Conv}\{\partial f_i(x)\})^0. \]

Obviously, $x$ is a Pareto critical iff $s(x) = 0$.

In a sense, $s(x)$ selects itself a different convex combination of the functions at each $x$. 
The multiobjective steepest descent direction

Definition

Given $x \in K$, the multiobjective steepest descent direction is

$$s(x) := - (N_K(x) + \text{Conv}\{\partial f_i(x)\})^0.$$  

Obviously, $x$ is a Pareto critical iff $s(x) = 0$.

In a sense, $s(x)$ selects itself a different convex combination of the functions at each $x$.

Example

If $q = 1$, then $s(x) = \text{proj } T_K(x)(-\nabla f(x))$. 
The multiobjective steepest descent direction

**Definition**

Given \( x \in K \), the **multiobjective steepest descent direction** is

\[
 s(x) := - (N_K(x) + \text{Conv}\{\partial f_i(x)\})^0.
\]

Obviously, \( x \) is a Pareto critical iff \( s(x) = 0 \).

In a sense, \( s(x) \) selects itself a different convex combination of the functions at each \( x \).

**Example**

If \( q = 1 \), then \( s(x) = \text{proj}_{T_K(x)}(-\nabla f(x)) \).

**Property**

\( s(x) \) is an admissible descent direction at \( x \), whenever \( s(x) \neq 0 \).
Example

\[ -s(x) \operatorname{Conv}\{\nabla f_i(x)\} \]
Why $s(x)$ is called the **steepest** descent?
Why \( s(x) \) is called the **steepest** descent?

Recall that (one objective function, no constraint):

\[
\frac{-\nabla f(x)}{\|\nabla f(x)\|} = \arg\min_{\|d\| \leq 1} df(x, d).
\]
The multiobjective steepest descent direction

Why \( s(x) \) is called the **steepest** descent?
Recall that (one objective function, no constraint):

\[
\frac{-\nabla f(x)}{\|\nabla f(x)\|} = \arg\min_{\|d\| \leq 1} df(x, d).
\]

The multiobjective steepest descent direction generalizes this fact:

**Theorem (Attouch, Garrigos, Goudou, 2014)**

\[
\frac{s(x)}{\|s(x)\|} = \arg\min_{\|d\| \leq 1, d \in T_K(x)} \max_i df_i(x, d).
\]
A continuous dynamic

The Multi-Objective Gradient dynamic:

\[(\text{MOG}) \quad \dot{u}(t) = s(u(t)) \quad \text{i.e.} \quad \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0\]
A continuous dynamic: example 1

(MOG) $\dot{u}(t) = s(u(t))$ i.e $\dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0$

$f_1(x) = \|x - a\|^2$ and $f_2(x) = \|x - b\|^2$
A continuous dynamic: example 2

\[(\text{MOG}) \dot{u}(t) = s(u(t)) \text{ i.e } \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0\]

\[f_1(x) = \frac{1}{2}\|x\|^2 \text{ and } f_2(x) = \langle a, x \rangle\]
A continuous dynamic: Existence and uniqueness

(MOG) \( \dot{u}(t) = s(u(t)) \) i.e \( \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0 \)

**Theorem (Attouch, Garrigos, Goudou, 2014)**

Suppose that \( H \) is finite-dimensional, and that the functions are convex and bounded from below. Then for any \( u_0 \in K \), there exists a strong solution \( u : [0, +\infty] \rightarrow K \) of (MOG), such that \( u(0) = u_0 \).

Strong solution essentially means an absolutely continuous trajectory \( u \) satisfying (MOG) for a.e. \( t > 0 \).
A continuous dynamic : example 2

\( (\text{MOG}) \dot{u}(t) = s(u(t)) \) i.e \( \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0 \)

\[ f_1(x) = \frac{1}{2}\|x\|^2 \text{ and } f_2(x) = \langle a, x \rangle \]
A continuous dynamic: Existence and uniqueness

\[(\text{MOG}) \dot{u}(t) = s(u(t)) \text{ i.e } \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0\]

**Theorem (Attouch, Garrigos, Goudou, 2014)**

Suppose that $H$ is finite-dimensional, and that the functions are convex and bounded from below. Then for any $u_0 \in K$, there exists a strong solution $u : [0, +\infty[ \rightarrow K$ of (MOG), such that $u(0) = u_0$.

The proof cannot rely on Cauchy-Lipschitz because of lack of Lipschitz regularity.

→ Use Morau-Yoshida’s regularization onto the $f_i$’s and the indicator function.

→ Use Peano’s existence theorem on the regularized system: it asks only continuity but do not guarantee uniqueness.

→ Pass to the limit. Hard.
The problem of uniqueness is still open. Can we find hypotheses ensuring Lipschitz continuity of $s(u)$?

**Local Lipschitz property**

Suppose $K = H$, and that the functions are of class $C^{1,1}$. The vector field $s$ is Lipschitz continuous at $u$ if:

- $q = 2$, and $\nabla f_1(u) \neq \nabla f_2(u)$.
- The vectors $\nabla f_i(u)$ are linearly independent.
A continuous dynamic : example 2

\[
\dot{u}(t) = s(u(t)) \text{ i.e } \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0
\]

\[
f_1(x) = \frac{1}{2}||x||^2 \text{ and } f_2(x) = \langle a, x \rangle
\]
A continuous dynamic : Qualitative behaviour

Theorem (Attouch, Garrigos, Goudou, 2014)
Suppose that the objective functions are lower regular (convex, or continuously differentiable ...). Then for all \( i = 1 \ldots q \), the function \( t \mapsto f_i(u(t)) \) is decreasing.

Theorem (Attouch, Garrigos, Goudou, 2014)
Suppose that the objective functions are quasi-convex. Then any bounded trajectory is weakly convergent. The limit point is a weak Pareto if the functions are convex. The limit point is a critical Pareto if the functions are \( C^1 \) or convex, and under compact assumption on \( u \).

We recover classic results by taking \( q = 1 \).

Can we have strong convergence under stronger assumptions?
Theorem (Attouch, Garrigos, Goudou, 2014)

Suppose that the objective functions are lower regular (convex, or continuously differentiable ...). Then for all $i = 1..q$, the function $t \mapsto f_i(u(t))$ is decreasing.

Theorem (Attouch, Garrigos, Goudou, 2014)

Suppose that the objective functions are quasi-convex. Then any bounded trajectory is weakly convergent. The limit point is a weak Pareto if the functions are convex. The limit point is a critical Pareto if the functions are $C^1$ or convex, and under compact assumption on $u$. We recover classic results by taking $q = 1$.

Can we have strong convergence under stronger assumptions?
Theorem (Attouch, Garrigos, Goudou, 2014)
Suppose that the objective functions are lower regular (convex, or continuously differentiable ...). Then for all $i = 1..q$, the function $t \mapsto f_i(u(t))$ is decreasing.

Theorem (Attouch, Garrigos, Goudou, 2014)
Suppose that the objective functions are quasi-convex.
- Then any bounded trajectory is weakly convergent.
- The limit point is a weak Pareto if the functions are convex.
- The limit point is a critical Pareto if the functions are $C^1$ or convex, and under compact assumption on $u$. 
A continuous dynamic: Qualitative behaviour

Theorem (Attouch, Garrigos, Goudou, 2014)
Suppose that the objective functions are lower regular (convex, or continuously differentiable ...). Then for all $i = 1..q$, the function $t \mapsto f_i(u(t))$ is decreasing.

Theorem (Attouch, Garrigos, Goudou, 2014)
Suppose that the objective functions are quasi-convex.
- Then any bounded trajectory is weakly convergent.
- The limit point is a weak Pareto if the functions are convex.
- The limit point is a critical Pareto if the functions are $C^1$ or convex, and under compact assumption on $u$.

- We recover classic results by taking $q = 1$. 
Theorem (Attouch, Garrigos, Goudou, 2014)

Suppose that the objective functions are lower regular (convex, or continuously differentiable ...). Then for all $i = 1..q$, the function $t \mapsto f_i(u(t))$ is decreasing.

Theorem (Attouch, Garrigos, Goudou, 2014)

Suppose that the objective functions are quasi-convex.

- Then any bounded trajectory is weakly convergent.
- The limit point is a weak Pareto if the functions are convex.
- The limit point is a critical Pareto if the functions are $C^1$ or convex, and under compact assumption on $u$.

- We recover classic results by taking $q = 1$.
- Can we have strong convergence under stronger assumptions?
A descent method associated to some scalarization $\sum_{i=1}^{q} \theta_i f_i$. In (MOG) the $\theta_i$ are chosen and modified automatically along the time. And ALL the functions decrease.
Continuous case: What (MOG) is not

- A descent method associated to some scalarization $\sum_{i=1}^{q} \theta_i f_i$. In (MOG) the $\theta_i$ are chosen and modified automatically along the time. And ALL the functions decrease.
- A descent method associated to $\max f_i$. 

Multicriteria optimization - Guillaume Garrigos - Franco/Italian workshop
A continuous dynamic: example 1

\[(\text{MOG}) \dot{u}(t) = s(u(t)) \text{ i.e. } \dot{u}(t) + (N_K(u(t)) + \text{Conv}\{\partial f_i(u(t))\})^0 = 0\]

\[f_1(x) = \|x - a\|^2 \text{ and } f_2(x) = \|x - b\|^2\]
The Multi-Objective Gradient dynamic

A Continuous Gradient-like Dynamical Approach to Pareto-Optimization in Hilbert Spaces. Attouch, Goudou, 2014

A Dynamic Gradient Approach to Pareto Optimization with Nonsmooth (...). Attouch, Garrigos, Goudou, Submitted.

Multi-Objective Gradient algorithm


Newton’s method


Proximal method

The Multi-Objective Gradient dynamic
A Continuous Gradient-like Dynamical Approach to Pareto-Optimization in Hilbert Spaces. Attouch, Goudou, 2014
A Dynamic Gradient Approach to Pareto Optimization with Nonsmooth (...). Attouch, Garrigos, Goudou, Submitted.

Multi-Objective Gradient algorithm

Newton’s method

Proximal method

Thank you for your attention!