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

Context

- 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.

Context

- 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 $\sum_{i=1}^{q} \theta_i f_i$.

Context

- 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.

Contents

Contents

Directional derivative (of Clarke)

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

Subdifferential (of Clarke)

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

Remark

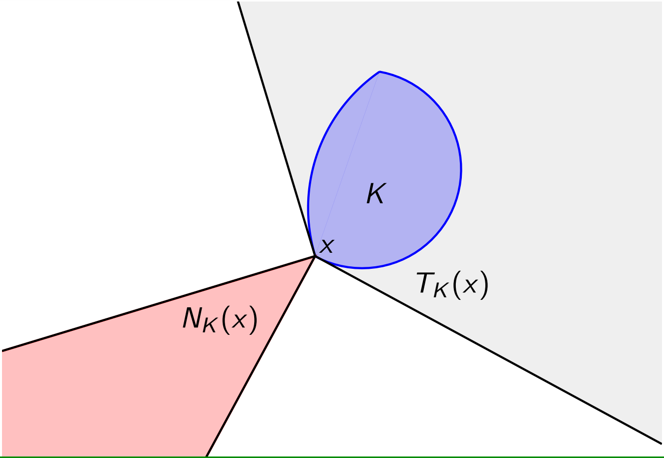
If f is of class C^1, then
$\partial f(x) = \{\nabla f(x)\}$ and $df(x;d) = \langle \nabla f(x), d \rangle$.

Nonsmooth analysis tools

Tangent and normal cones

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

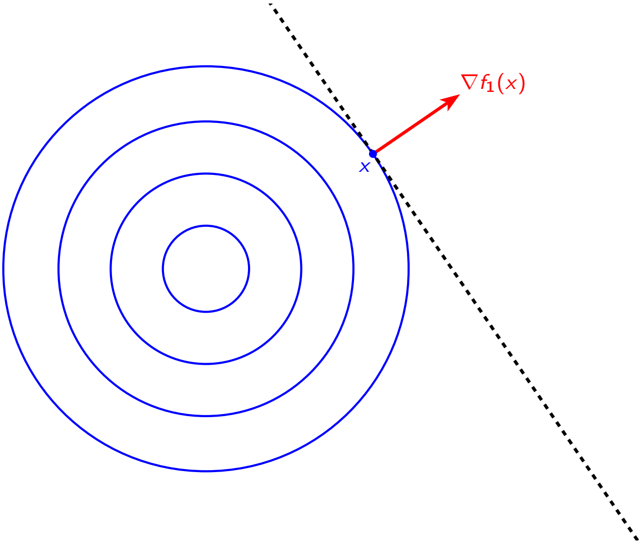
$$N_K(x) := \{p \in H \mid \langle p, d \rangle \leq 0\ \forall d \in T_K(x)\}.$$



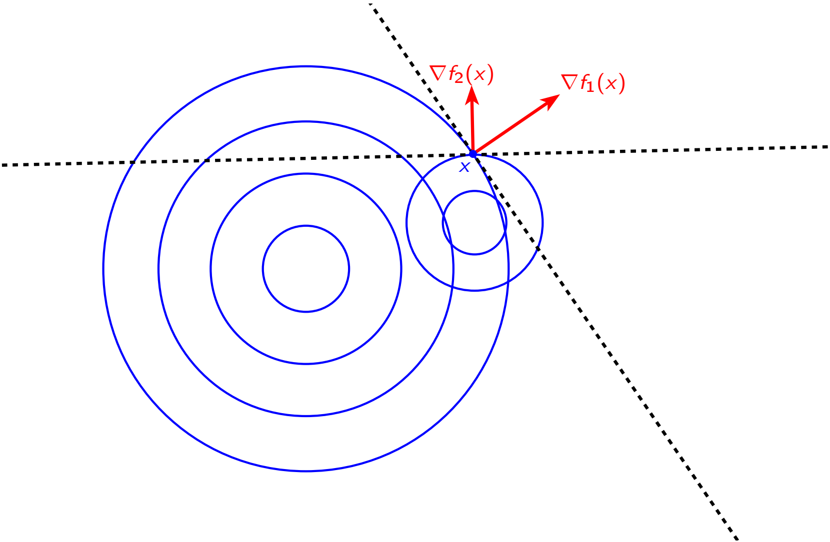
Descent direction

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



Example



Armijo direction

We say that a descent direction $d \in H$ is an *Armijo direction* if $\exists \varepsilon > 0$ s.t. 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 $x + td \in K$.

Pareto equilibrium(s)

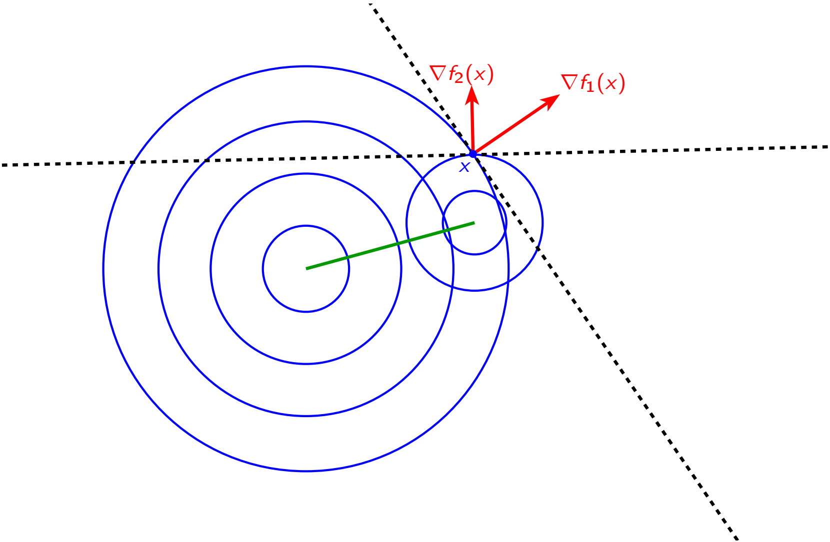
Pareto equilibrium(s)

- 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

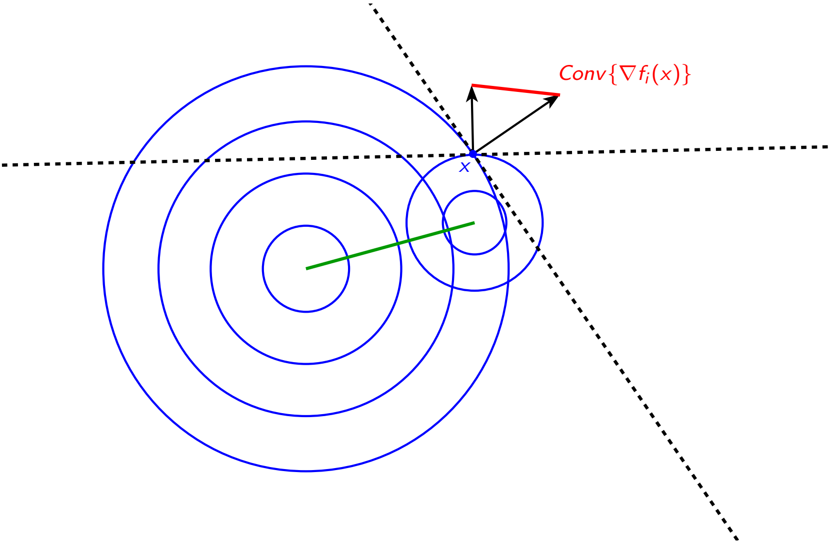


Multicriteria analysis

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 $0 \in N_K(x) + Conv\{\partial f_i(x)\}$.

Example



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 $0 \in N_K(x) + Conv\{\partial f_i(x)\}$.

Multicriteria analysis

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 $0 \in N_K(x) + Conv\{\partial f_i(x)\}$.

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.

Link between descent direction and Pareto equilibrium

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.

Objectif

We will consider

1. a continuous dynamic $\dot{u}(t) = s(u(t))$, where $s : K \to H$ 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.

Contents

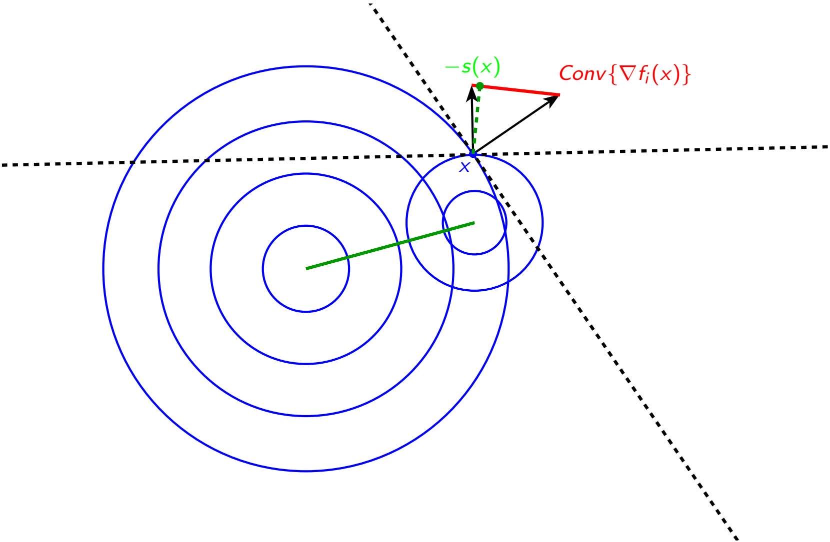
The multiobjective steepest descent direction

Definition

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

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

Example



The multiobjective steepest descent direction

Definition

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

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

The multiobjective steepest descent direction

Definition

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

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

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

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

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

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

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 $s(x) = \mathrm{proj}_{\,T_K(x)}(-\nabla f(x))$.

The multiobjective steepest descent direction

Definition

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

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

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.

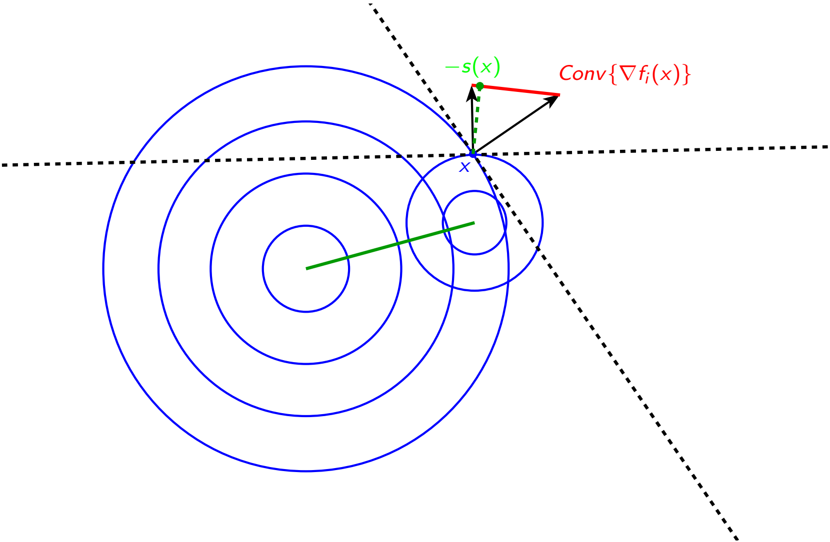
Example

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

Property

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

Example



The multiobjective steepest descent direction

Why $s(x)$ is called the **steepest** descent ?

The multiobjective steepest descent direction

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

$$\frac{-\nabla f(x)}{\|\nabla f(x)\|} = \underset{\|d\|\leq 1}{\operatorname{argmin}} \ df(x,d).$$

The multiobjective steepest descent direction

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

$$\frac{-\nabla f(x)}{\|\nabla f(x)\|} = \underset{\|d\| \leq 1}{\text{argmin}} \;\; df(x, d).$$

The multiobjective steepest descent direction generalizes this fact :

Theorem (Attouch, Garrigos, Goudou, 2014)

$$\frac{s(x)}{\|s(x)\|} = \underset{\|d\| \leq 1, d \in T_K(x)}{\text{argmin}} \;\; \max_i df_i(x, d).$$

A continuous dynamic

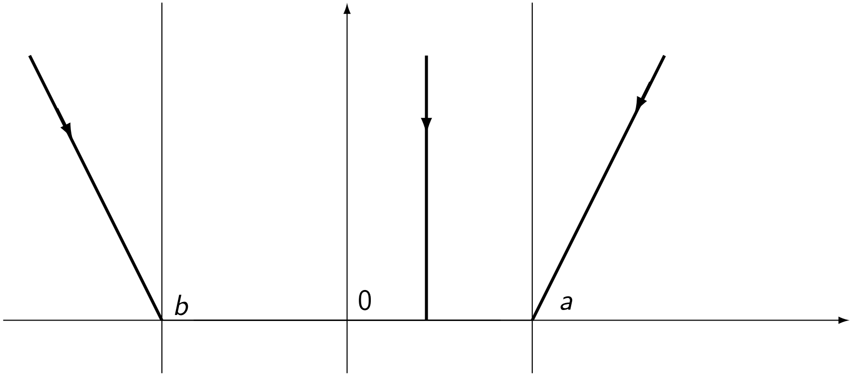
The Multi-Objective Gradient dynamic :

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

A continuous dynamic : example 1

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

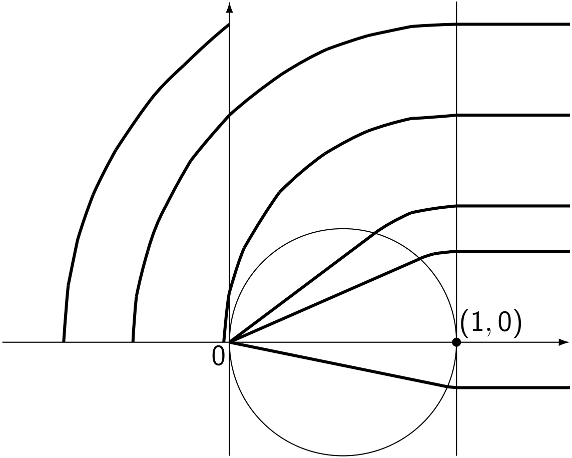
$f_1(x) = \|x - a\|^2$ and $f_2(x) = \|x - b\|^2$



A continuous dynamic : example 2

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

$f_1(x) = \frac{1}{2}\|x\|^2$ and $f_2(x) = \langle a, x \rangle$



A continuous dynamic : Existence and uniqueness

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

Theorem (Attouch, Garrigos, Goudou, 2014)

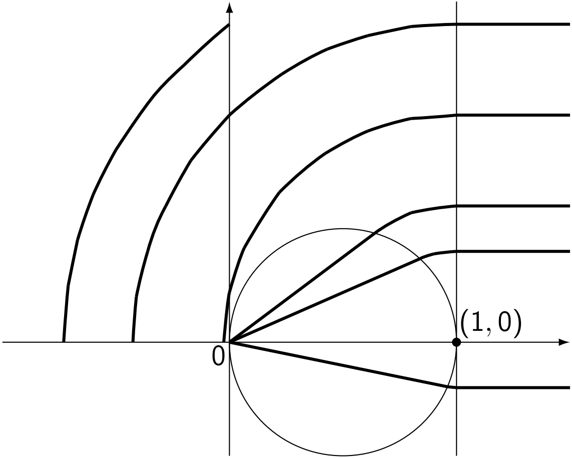
Suppose that H is finite-dimentional, and that the functions are convex and bounded from below. Then for any $u_0 \in K$, there exists a strong solution $u : [0, +\infty[\to K$ of (MOG), such that $u(0) = u_0$.

Strong solution essentially means an absolutely continuous trajectory u satisfying (MOG) for a.e. $t > 0$.

A continuous dynamic : example 2

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

$f_1(x) = \frac{1}{2}\|x\|^2$ and $f_2(x) = \langle a, x \rangle$



A continuous dynamic : Existence and ~~uniqueness~~

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

Theorem (Attouch, Garrigos, Goudou, 2014)

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

The proof cannot rely on Cauchy-Lipschitz because of lack of Lipschitz regularity.
$\rightarrow$ Use Morau-Yoshida's regularization onto the f_i's and the indicator function.
$\rightarrow$ Use Peano's existence theorem on the regularized system : it asks only continuity but do not guarantee uniqueness.
$\rightarrow$ Pass to the limit. Hard.

About uniqueness

The problem of uniqueness is still open.
Can we find hypotheses ensuring Lipschitz continuity of $s(u)$?

Local Lipschitz property

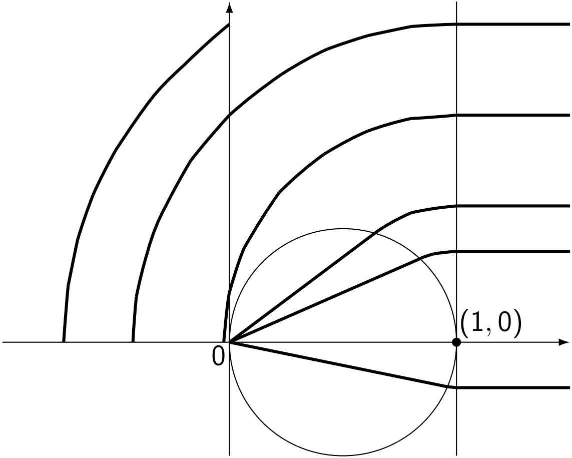
Suppose $K = H$, and that the functions are of class $C^{1,1}$.
The vector field s is Lipschitz continuous at u if :

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

A continuous dynamic : example 2

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

$f_1(x) = \frac{1}{2}\|x\|^2$ and $f_2(x) = \langle a, x \rangle$



A continuous dynamic : Qualitative behaviour

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.

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.

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 $q = 1$.

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 $q = 1$.
- Can we have strong convergence under stronger assumptions ?

Continuous case : What (MOG) is not

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

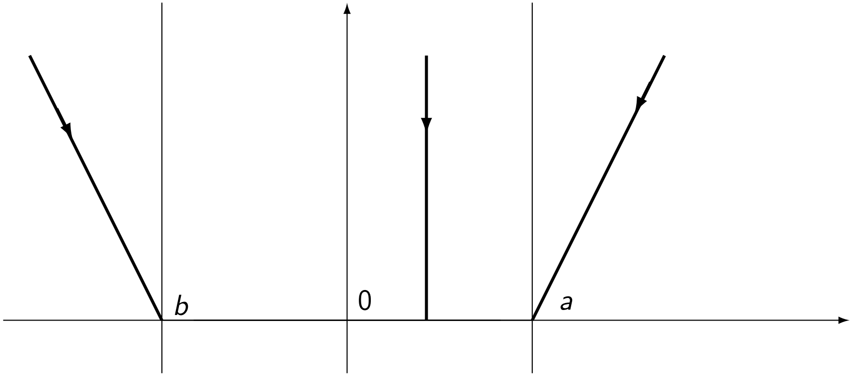
Continuous case : What (MOG) is not

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

A continuous dynamic : example 1

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

$f_1(x) = \|x - a\|^2$ and $f_2(x) = \|x - b\|^2$



Contents

Discrete descent direction

We want to consider an algorithm $u_{n+1} = u_n + t_n d_n$ where d_n is an admissible Armijo direction.
We could take $d_n = s(u_n)$ but it is not Armijo admissible.

Discrete descent direction

We want to consider an algorithm $u_{n+1} = u_n + t_n d_n$ where d_n is an admissible Armijo direction.
We could take $d_n = s(u_n)$ but it is not Armijo admissible.
It can be shown that $s(x) = \underset{d \in T_K(x)}{\operatorname{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i df_i(x, d)$.

Discrete descent direction

We want to consider an algorithm $u_{n+1} = u_n + t_n d_n$ where d_n is an admissible Armijo direction.
We could take $d_n = s(u_n)$ but it is not Armijo admissible.
It can be shown that $s(x) = \underset{d \in T_K(x)}{\operatorname{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i df_i(x, d)$.

An Armijo admissible descent selection

For $\mu > 0$, consider $\sigma_\mu(x) := \underset{d \in \frac{K-x}{\mu}}{\operatorname{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i df_i(x, d)$.

Discrete descent direction

We want to consider an algorithm $u_{n+1} = u_n + t_n d_n$ where d_n is an admissible Armijo direction.
We could take $d_n = s(u_n)$ but it is not Armijo admissible.
It can be shown that $s(x) = \underset{d \in T_K(x)}{\text{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i df_i(x, d)$.

An Armijo admissible descent selection

For $\mu > 0$, consider $\sigma_\mu(x) := \underset{d \in \frac{K-x}{\mu}}{\text{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i df_i(x, d)$.

Remark

- $s(x) = \sigma_\mu(x)$ whenever K is a vectorial subspace.
- In some sense, $\sigma_\mu(x) \to s(x)$ when $\mu \to 0$.

Discrete descent direction

An Armijo admissible descent selection

For $\mu > 0$, consider $\sigma_\mu(x) := \underset{d \in \frac{K-x}{\mu}}{\operatorname{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i df_i(x,d)$.

Proposition

For all $\mu > 0$:

- $\sigma_\mu(x) = 0$ if x is a Pareto critical point.
- $\sigma_\mu(x)$ is an admissible Armijo direction else.

Algorithm : steepest descent method with linesearch

Consider a sequence of parameters $(\mu_n) \subset\subset \mathbb{R}_+$ and $x_0 \in K$.
Compute $d_n = \sigma_{\mu_n}(x_n)$.
Select a stepsize t_n by a common Armijo selection over all the f_i's.
Take $x_{n+1} = x_n + t_n d_n$.

Remark

- The algorithm is well defined.
- At each step, all the objective functions decrease.

Algorithm : steepest descent method with linesearch

Theorem

If the functions are quasi-convex, then any bounded sequence weakly converges.

Algorithm : steepest descent method with linesearch

Theorem

If the functions are quasi-convex, then any bounded sequence weakly converges.

Theorem (Drummond, Iusem, 2004) ; (Fliege, Svaiter, 2000)

In finite dimension, if the functions are C^1, then any limit point is Pareto critical.

Algorithm : steepest descent method with linesearch

Theorem

If the functions are quasi-convex, then any bounded sequence weakly converges.

Theorem (Drummond, Iusem, 2004) ; (Fliege, Svaiter, 2000)

In finite dimension, if the functions are C^1, then any limit point is Pareto critical.

Corollary

In finite dimension, if the functions are C^1 quasi-convex, then any bounded sequence converges to a critical Pareto point.

Algorithm : some comments

- Differentiable case, no constraints : the multiobjective descent method is relatively easy to obtain once the gradients $\nabla f_i(x_n)$ are known, mainly if q is small ($q \leq 3\,?$).

Algorithm : some comments

- Differentiable case, no constraints : the multiobjective descent method is relatively easy to obtain once the gradients $\nabla f_i(x_n)$ are known, mainly if q is small ($q \leq 3\,?$).
- Differentiable case, with constraint : the constraint enforces us to transform a projection problem into a minimization subproblem

$$\sigma_\mu(x_n) := \underset{d \in \frac{K-x}{\mu}}{\operatorname{argmin}} \ \frac{1}{2}\|d\|^2 + \max_i \langle \nabla f_i(x_n), d \rangle$$

See also (Fliege, Svaiter, 00) for K defined by inequalites

Algorithm : some comments

- Differentiable case, no constraints : the multiobjective descent method is relatively easy to obtain once the gradients $\nabla f_i(x_n)$ are known, mainly if q is small ($q \leq 3$?).
- Differentiable case, with constraint : the constraint enforces us to transform a projection problem into a minimization subproblem

$$\sigma_\mu(x_n) := \underset{d \in \frac{K-x}{\mu}}{\operatorname{argmin}} \; \frac{1}{2}\|d\|^2 + \max_i \langle \nabla f_i(x_n), d \rangle$$

  See also (Fliege, Svaiter, 00) for K defined by inequalites
- Even with convexity, explicit schemes are bad designed for non-differentiable functions (recall monocriteria case).

Bibliography

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

Steepest descent methods for multicriteria optimization. Fliege, Svaiter, 2000.

A steepest descent method for vector optimization. Drummond, Svaiter, 2005.

Newton's method

Newton's Method for Multiobjective Optimization. Fliege, Drummond, Svaiter, 2009.

A quadratically convergent Newton method for vector optimization. Drummond, Raupp, Svaiter, 2014.

Quasi-Newton's method for multiobjective optimization. Povalej, 2014.

Proximal method

Proximal Methods in Vector Optimization. Bonnel, Iusem, Svaiter, 2005.

Bibliography

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

Steepest descent methods for multicriteria optimization. Fliege, Svaiter, 2000.

A steepest descent method for vector optimization. Drummond, Svaiter, 2005.

Newton's method

Newton's Method for Multiobjective Optimization. Fliege, Drummond, Svaiter, 2009.

A quadratically convergent Newton method for vector optimization. Drummond, Raupp, Svaiter, 2014.

Quasi-Newton's method for multiobjective optimization. Povalej, 2014.

Proximal method

Proximal Methods in Vector Optimization. Bonnel, Iusem, Svaiter, 2005.

Thank you for your attention !