Dynamique(s) de descente pour l’optimisation multi-objectif

Guillaume Garrigos

Istituto Italiano di Tecnologia & Massachusetts Institute of Technology
Genova, Italie

Journées SMAI-MODE
24 Mars, 2016
In engineering, decision sciences, it happens that various objective functions shall be minimized simultaneously: $f_1, \ldots, f_m : H \to \mathbb{R}$
In engineering, decision sciences, it happens that various objective functions shall be minimized simultaneously: $f_1, \ldots, f_m : H \rightarrow \mathbb{R}$

→ Needs appropriate tools: multi-objective optimization.
The multi-objective optimization problem

Let $F = (f_1, ..., f_m) : H \rightarrow \mathbb{R}^m$ locally Lipschitz, $H$ Hilbert.

Solve $\text{MIN} (f_1(x), ..., f_m(x)) : x \in C \subset H \text{ convex.}$
The multi-objective optimization problem

Let \( F = (f_1, \ldots, f_m) : H \to \mathbb{R}^m \) locally Lipschitz, \( H \) Hilbert.

Solve \( \text{MIN} (f_1(x), \ldots, f_m(x)) : x \in C \subset H \) convex.

We consider the usual order(s) on \( \mathbb{R}^m \):

\[
\begin{align*}
    a \leq b & \iff a_i \leq b_i \text{ for all } i = 1, \ldots, m, \\
    a < b & \iff a_i < b_i \text{ for all } i = 1, \ldots, m.
\end{align*}
\]

\[x\] is a Pareto point if \( \not\exists y \in C \) such that \( F(y) \nleq F(x) \).

\[x\] is a weak Pareto point if \( \not\exists y \in C \) such that \( F(y) \aleq F(x) \).
The multi-objective optimization problem

Let $F = (f_1, \ldots, f_m) : H \to \mathbb{R}^m$ locally Lipschitz, $H$ Hilbert.

Solve $\text{MIN} \ (f_1(x), \ldots, f_m(x)) : x \in C \subset H$ convex.

We consider the usual order(s) on $\mathbb{R}^m$:

- $a \leq b \iff a_i \leq b_i$ for all $i = 1, \ldots, m$,
- $a < b \iff a_i < b_i$ for all $i = 1, \ldots, m$.

$x$ is a **Pareto point** if
$\nexists y \in C$ such that $F(y) \not\leq F(x)$

$x$ is a **weak Pareto point** if
$\nexists y \in C$ such that $F(y) < F(x)$
The multi-objective optimization problem

Let $F = (f_1, \ldots, f_m) : H \rightarrow \mathbb{R}^m$ locally Lipschitz.

Solve $\text{MIN } f_1(x), \ldots, f_m(x) : x \in C \subset H$ convex.

How to solve it?
The multi-objective optimization problem

Let \( F = (f_1, \ldots, f_m) : H \to \mathbb{R}^m \) locally Lipschitz.

Solve \( \text{MIN } f_1(x), \ldots, f_m(x) : x \in C \subset H \) convex.

How to solve it?

- genetic algorithm \( \rightarrow \) no theoretical guarantees.
The multi-objective optimization problem

Let \( F = (f_1, ..., f_m) : H \to \mathbb{R}^m \) locally Lipschitz.

\[
\text{Solve } \quad \text{MIN } f_1(x), ..., f_m(x) : x \in C \subset H \text{ convex.}
\]

How to solve it?

- genetic algorithm \( \longrightarrow \) no theoretical guarantees.
- scalarization method:

\[
\bigcup_{\theta \in \Delta^m} \text{argmin}_{x \in H} f_\theta(x) \subset \{\text{weak Paretos}\} \subset \{\text{Paretos}\},
\]

where \( \Delta^m \) is the simplex unit and \( f_\theta(x) := \sum_{i=1}^{m} \theta_i f_i(x) \).
Let $F = (f_1, ..., f_m) : H \to \mathbb{R}^m$ locally Lipschitz.

Solve $\text{MIN } f_1(x), ..., f_m(x) : x \in C \subset H$ convex.

We are going to present a method which:
Let $F = (f_1, \ldots, f_m) : H \rightarrow \mathbb{R}^m$ locally Lipschitz.

Solve \ \text{MIN} \ f_1(x), \ldots, f_m(x) : x \in C \subset H \text{ convex}.

We are going to present a method which:

- generalizes the gradient descent dynamic $\dot{x}(t) + \nabla f(x(t)) = 0$,
The multi-objective optimization problem

Let $F = (f_1, ..., f_m) : H \to \mathbb{R}^m$ locally Lipschitz.

Solve \( \text{MIN } f_1(x), ..., f_m(x) : x \in C \subset H \) convex.

We are going to present a method which:

- generalizes the gradient descent dynamic $\dot{x}(t) + \nabla f(x(t)) = 0$,
- is cooperative, i.e. all objective functions decrease simultaneously,
The multi-objective optimization problem

Let $F = (f_1, ..., f_m) : H \to \mathbb{R}^m$ locally Lipschitz.

Solve \( \text{MIN } f_1(x), ..., f_m(x) : x \in C \subset H \) convex.

We are going to present a method which:

- generalizes the gradient descent dynamic $\dot{x}(t) + \nabla f(x(t)) = 0$,
- is \textit{cooperative}, i.e. all objective functions decrease simultaneously,
- is independent of any choice of parameters.
Single objective optimization:

\[ x_{n+1} = x_n + \lambda_n d_n, \]

where \( d_n \) satisfies \( df(x_n; d_n) < 0 \) (e.g. \( d_n = -\nabla f(x_n) \)).

Multi-objective optimization:
Can we find \( d_n \) such that \( df_i(x_n; d_n) < 0 \) for all \( i \in \{1, ..., m\} \) ?
Towards a descent dynamic for multi-objective optimization

Historical review

Cornet (1981)

\[ s(x) := -[\nabla f_1(x), \nabla f_2(x)]^0 \]

\[ \langle s(x), \nabla f_i(x) \rangle < 0 \]

\[ \nabla f_1(x) \]

\[ \nabla f_2(x) \]

\[ \langle s(x), \nabla f_i(x) \rangle < 0 \]
Let $F = (f_1, \ldots, f_m) : H \to \mathbb{R}^m$ locally Lipschitz, $C = H$ Hilbert.

**Definition**

For all $x \in H$, $s(x) := -(\co \{ \partial^C f_i(x) \}_{i=1,\ldots,m})^0$ is the (common) steepest descent direction at $x$. 
Let $F = (f_1, \ldots, f_m) : H \rightarrow \mathbb{R}^m$ locally Lipschitz, $C = H$ Hilbert.

**Definition**

For all $x \in H$, $s(x) := -\left(\text{co} \left\{ \partial^c f_i(x) \right\}_{i=1,\ldots,m}\right)^0$ is the (common) steepest descent direction at $x$.

**Remarks in the smooth case**

- If $m = 1$ then $s(x) = -\nabla f_1(x)$. 

Multi-objective steepest descent

Let $F = (f_1, \ldots, f_m) : H \rightarrow \mathbb{R}^m$ locally Lipschitz, $C = H$ Hilbert.

**Definition**

For all $x \in H$, $s(x) := -(\text{co} \{ \partial^c f_i(x) \}_{i=1,\ldots,m})^0$ is the (common) steepest descent direction at $x$.

**Remarks in the smooth case**

- If $m = 1$ then $s(x) = -\nabla f_1(x)$.
- At each $x$, $s(x)$ selects a convex combination:

$$s(x) = -\sum_{i=1}^m \theta_i(x) \nabla f_i(x) = -\nabla f_{\theta(x)}(x)$$

where $f_{\theta(x)} = \sum_{i=1}^m \theta_i(x) f_i$. 
Multi-objective steepest descent

Let \( F = (f_1, \ldots, f_m) : H \rightarrow \mathbb{R}^m \) locally Lipschitz, \( C = H \) Hilbert.

**Definition**

For all \( x \in H \), \( s(x) := -\left( \text{co} \left\{ \partial^c f_i(x) \right\} \right)_{i=1,\ldots,m}^0 \) is the (common) steepest descent direction at \( x \).

**Remarks in the smooth case**

- If \( m = 1 \) then \( s(x) = -\nabla f_1(x) \).
- At each \( x \), \( s(x) \) selects a convex combination:

  \[
  s(x) = -\sum_{i=1}^{m} \theta_i(x) \nabla f_i(x) = -\nabla f_{\theta(x)}(x) \quad \text{where} \quad f_{\theta(x)} = \sum_{i=1}^{m} \theta_i(x) f_i.
  \]

- \( s(x) \) is the steepest descent:

  \[
  \frac{s(x)}{\|s(x)\|} = \arg\min_{d \in B_H} \left\{ \max_{i=1,\ldots,m} \langle \nabla f_i(x), d \rangle \right\}.
  \]
The (multi-objective) Steepest Descent dynamic

Algorithm:

\[ x_{n+1} = x_n + \lambda_n s(x_n). \]

Studied in the 2000’s by Svaiter, Fliege, Iusem, ...

Continuous dynamic:

\[ (SD) \quad \dot{x}(t) = s(x(t)), \]

i.e. \[ (SD) \quad \dot{x}(t) + \left( \text{co} \left\{ \partial^c f_i(x(t)) \right\}_i \right)^0 = 0 \]
The (multi-objective) Steepest Descent dynamic example

\[
(SD) \quad \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = \|x\|^2 \quad \text{and} \quad f_2(x) = x_1.
\]
The (multi-objective) Steepest Descent dynamic

Example

\[
(\text{SD}) \quad \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = \|x\|^2 \quad \text{and} \quad f_2(x) = x_1.
\]
The (multi-objective) Steepest Descent dynamic
Example

(SD) \( \dot{x}(t) = s(x(t)) \) with \( f_1(x) = \|x\|^2 \) and \( f_2(x) = x_1 \).
The (multi-objective) Steepest Descent dynamic

Example

\( (SD) \quad \dot{x}(t) = s(x(t)) \) with \( f_1(x) = \|x\|^2 \) and \( f_2(x) = x_1 \).
The (multi-objective) Steepest Descent dynamic

Example

$$(SD) \quad \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = \|x\|^2 \quad \text{and} \quad f_2(x) = x_1.$$
The (multi-objective) Steepest Descent dynamic

Example

\[ (SD) \quad \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = \|x\|^2 \quad \text{and} \quad f_2(x) = x_1. \]
The (multi-objective) Steepest Descent dynamic example

\[ \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = x_1^2 \quad \text{and} \quad f_2(x) = x_2^2. \]
The (multi-objective) Steepest Descent dynamic

Example

\[ (SD) \quad \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = x_1^2 \quad \text{and} \quad f_2(x) = x_2^2. \]
The (multi-objective) Steepest Descent dynamic example

\[(SD) \quad \dot{x}(t) = s(x(t)) \quad \text{with} \quad f_1(x) = x_1^2 \quad \text{and} \quad f_2(x) = x_2^2.\]
A cooperative dynamic

Let $x : \mathbb{R}_+ \rightarrow H$ be a solution of (SD) $\dot{x}(t) = s(x(t))$.
For all $i = 1, ..., m$, the function $t \mapsto f_i(x(\cdot))$ is decreasing.
A cooperative dynamic

Let $x : \mathbb{R}_+ \rightarrow H$ be a solution of (SD) $\dot{x}(t) = s(x(t))$.
For all $i = 1, \ldots, m$, the function $t \mapsto f_i(x(\cdot))$ is decreasing.

Convergence in the convex case

Assume that the objective functions are convex. Then any bounded trajectory weakly converges to a weak Pareto point.
A cooperative dynamic

Let $x : \mathbb{R}_+ \rightarrow H$ be a solution of (SD) $\dot{x}(t) = s(x(t))$.
For all $i = 1, \ldots, m$, the function $t \mapsto f_i(x(\cdot))$ is decreasing.

Convergence in the convex case

Assume that the objective functions are convex. Then any bounded trajectory weakly converges to a weak Pareto point.

Existence in the convex case

Suppose that $H$ is finite dimensional. Then, for any initial data, there exists a global solution to (SD).
In case of convex constraint $C \subset H$:

\[
(SD) \quad \dot{x}(t) + (N_C(x(t)) + \text{co} \{\partial^c f_i(x(t))\}_i)^0 = 0.
\]

How to discretize it properly?
The (multi-objective) Steepest Descent dynamic

Going further

In case of convex constraint $C \subset H$:

\[
(SD) \quad \dot{x}(t) + (N_C(x(t)) + \text{co } \{\partial^c f_i(x(t))\})^0_i = 0.
\]

How to discretize it properly?

Uniqueness? Yes, if \(\{\nabla f_i(x(\cdot))\}_{i=1,\ldots,m}\) are affinely independants.
In case of convex constraint $C \subset H$:

$$(SD) \quad \dot{x}(t) + (N_C(x(t)) + \text{co} \{ \partial^c f_i(x(t)) \}_i) ^0 = 0.$$ 

How to discretize it properly?

- **Uniqueness?** Yes, if $\{ \nabla f_i(x(\cdot)) \}_{i=1,...,m}$ are affinely independants.

- **Convergence to Pareto points?** Guaranteed by endowing $\mathbb{R}^m$ with a different order (but some of the Paretos might be lost in the operation).
Numerical results

Recovering the Pareto front

\[
\begin{align*}
  f_1(x, y) &= x + y \\
  f_2(x, y) &= x^2 + y^2 + \frac{1}{x} + 3e^{-100(x-0.3)^2} + 3e^{-100(x-0.6)^2} \\
  (x, y) &\in C = [0.1, 1]^2
\end{align*}
\]

Plot of \( F(C), F = (f_1, f_2) : C \rightarrow \mathbb{R}^2 \).
Numerical results
Recovering the Pareto front

\[
\begin{align*}
\mathbf{f}_1(x, y) &= x + y \\
\mathbf{f}_2(x, y) &= x^2 + y^2 + \frac{1}{x} + 3e^{-100(x-0.3)^2} + 3e^{-100(x-0.6)^2}
\end{align*}
\]

\[(x, y) \in C = [0.1, 1]^2\]

Plot of \( F(C) \), \( F = (f_1, f_2) : C \rightarrow \mathbb{R}^2 \) and its pareto front.
Numerical results
Recovering the Pareto front

\[ f_1(x, y) = x + y \]
\[ f_2(x, y) = x^2 + y^2 + \frac{1}{x} + 3e^{-100(x-0.3)^2} + 3e^{-100(x-0.6)^2} \]

\((x, y) \in C = [0.1, 1]^2\)

Gradient method (Right) vs Scalar method (Left). 100 samples.
Can we select, among the weak Paretos (= the zeros of $x \mapsto s(x)$) the closest to a desired state?
Can we select, among the weak Paretos (= the zeros of $x \mapsto s(x)$) the closest to a desired state?

→ Tikhonov regularization

$$\dot{x}(t) - s(x(t)) + \varepsilon (x(t) - x_d) = 0, \varepsilon > 0.$$
Numerical results
Pareto selection with Tikhonov penalization
Can we select, among the weak Paretos (= the zeros of $x \mapsto s(x)$) the closest to a desired state?

→ Diagonal Tikhonov regularization

$$\dot{x}(t) - s(x(t)) + \varepsilon(t)(x(t) - x_d) = 0,$$

$$\varepsilon(t) \downarrow 0, \int_0^\infty \varepsilon(t) \, dt = +\infty.$$

See the works of Attouch, Cabot, Czarnecki, Peypouquet (...) in the monotone case.
Numerical results
Pareto selection with Tikhonov penalization
Numerical results
Pareto selection with Tikhonov penalization
Numerical results
Pareto selection with Tikhonov penalization
What about inertial dynamics?

\[ \dot{x}(t) + \nabla f(x(t)) = 0 \]

\[ x_{n+1} = x_n - \lambda \nabla f(x_n) \]

\[ \ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0 \]

\[ x_{n+1} = y_n - \lambda \nabla f(y_n) \]

\[ y_{n+1} = x_{n+1} + (1 - \gamma)(x_{n+1} - x_n) \]
What about inertial dynamics?

\[ \dot{x}(t) + \nabla f(x(t)) = 0 \]

\[ x_{n+1} = x_n - \lambda \nabla f(x_n) \]

Inertia promotes Faster trajectories (varying \( \gamma \)), Exploratory properties.

\[ \ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0 \]

\[ x_{n+1} = y_n - \lambda \nabla f(y_n) \]

\[ y_{n+1} = x_{n+1} + (1 - \gamma)(x_{n+1} - x_n) \]
What about inertial dynamics?

\[ \dot{x}(t) + \nabla f(x(t)) = 0 \]

\[ \ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0 \]

Inertia promotes

- Faster trajectories (varying \( \gamma \)),
- Exploratory properties.
Convergence rates: empirical observation

Convergence rate of $\|F(x^n) - F(x^\infty)\|_\infty$:

**Steepest Descent** vs **Inertial Steepest Descent**
Let $f_1, \ldots, f_m$ be smooth, with $L$-Lipschitz gradient.

(ISD) $\ddot{x}(t) = -\gamma \dot{x}(t) + s(x(t))$. 
Inertial (multi-objective) Steepest Descent

Let $f_1, \ldots, f_m$ be smooth, with $L$-Lipschitz gradient.

(ISD) $\ddot{x}(t) = -\gamma \dot{x}(t) + s(x(t))$.

Example: $f_1(x) = \|x\|^2$ and $f_2(x) = x_1$. 
Let $f_1, \ldots, f_m$ be smooth, with $L$-Lipschitz gradient.

\[(\text{ISD}) \quad m\ddot{x}(t) = -\gamma \dot{x}(t) + s(x(t)).\]

Assume that $\gamma \geq L$. 

Inertial (multi-objective) Steepest Descent
Main results (Attouch, G., 2015)

Let $f_1, \ldots, f_m$ be smooth, with $L$-Lipschitz gradient.

\[(ISD) \quad m\ddot{x}(t) = -\gamma \dot{x}(t) + s(x(t)).\]

Assume that $\gamma \geq L$.

Existence

Suppose that $H$ is finite dimensional. Then, for any initial data, there exists a global solution to (ISD).
Let $f_1, \ldots, f_m$ be smooth, with $L$-Lipschitz gradient.

\[(ISD) \quad m\ddot{x}(t) = -\gamma \dot{x}(t) + s(x(t)).\]

Assume that $\gamma \geq L$.

### Existence

Suppose that $H$ is finite dimensional. Then, for any initial data, there exists a global solution to (ISD).

### Convergence in the convex case

Let $f_1, \ldots, f_m$ be convex. Then, any bounded trajectory weakly converges to a weak Pareto point.
The steepest descent provides a flexible tool once adapted to multi-objective optimization problems.
Conclusion

The steepest descent provides a flexible tool once adapted to multi-objective optimization problems.

Open questions:

- Understand the asymptotic behaviour of

  \[ \dot{x}(t) - s(x(t)) + \varepsilon(t)x(t) = 0 \]

  (the set of weak Paretos is non-convex).
The steepest descent provides a flexible tool once adapted to multi-objective optimization problems.

Open questions:

- Understand the asymptotic behaviour of

\[ \dot{x}(t) - s(x(t)) + \varepsilon(t)x(t) = 0 \]

(the set of weak Paretos is non-convex).

- Having convergence rates for first and second-order dynamics (the critical values are not unique).
Thank you for your attention!