

A DYNAMIC GRADIENT APPROACH TO PARETO OPTIMIZATION WITH NONSMOOTH CONVEX OBJECTIVE FUNCTIONS

HÉDY ATTOUCH, GUILLAUME GARRIGOS, AND XAVIER GOUDOU

ABSTRACT. In a general Hilbert framework, we consider continuous gradient-like dynamical systems for constrained multiobjective optimization involving non-smooth convex objective functions. Based on the Yosida regularization of the subdifferential operators involved in the system, we obtain the existence of strong global trajectories. We prove a descent property for each objective function, and the convergence of trajectories to weak Pareto minima. This approach provides a dynamical endogenous weighting of the objective functions. Applications are given to cooperative games, inverse problems, and numerical multiobjective optimization.

Key words: multiobjective optimization; Pareto optima; multiobjective steepest descent; convex objective functions; subdifferential operators; Yosida approximation; asymptotic behavior; cooperative games; sparse optimization; inverse problems; gradient methods.

AMS subject classification. 34E10, 37L05, 37L65, 90B50, 90C29, 90C31, 91A12, 91A35, 91B06, 91B55.

INTRODUCTION

Throughout the paper, \mathcal{H} is a real Hilbert space with scalar product and norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ respectively. We are interested with a gradient-like dynamical approach to the Pareto optima of the Constrained Multiobjective Optimization problem ((CMO) for short)

$$(CMO) \quad \min \{F(v) : v \in K\}$$

where $F : \mathcal{H} \rightarrow \mathbb{R}^q$, $F(v) = (f_i(v))_{i=1, \dots, q}$, $q \in \mathbb{N}^*$. Working in a general Hilbert space (possibly infinite dimensional) covers both applications in decision sciences and engineering. We make the following standing assumptions on the multiple objective functions $(f_i)_{i=1, 2, \dots, q}$, and constraint K :

H0) $K \subset \mathcal{H}$ is a closed convex nonempty set.

For each $i = 1, 2, \dots, q$, $f_i : \mathcal{H} \rightarrow \mathbb{R}$ is a real-valued function which satisfies:

H1) f_i is convex continuous. It is supposed to be Lipschitz continuous on bounded sets. Equivalently, its subdifferential $\partial f_i : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ is bounded on bounded sets;

H2) f_i is bounded from below on \mathcal{H} .

We are interested in this paper with the *lazy* solutions (also called slow solutions, see [10, Ch. 6, section 8]) of the differential inclusion

$$(1) \quad \dot{u}(t) + N_K(u(t)) + \text{Conv} \{ \partial f_i(u(t)) \} \ni 0,$$

which is governed by the sum of the two set-valued operators $u \mapsto N_K(u)$, and $u \mapsto \text{Conv} \{ \partial f_i(u) \}$. For $u \in K$, $N_K(u)$ is the normal cone to K at u , a closed convex cone modeling the *contact forces* which are attached to the constraint K . Besides, $\text{Conv} \{ \partial f_i(u) \}$ denotes the closed convex hull of the sets $\{ \partial f_i(u) ; i = 1, \dots, q \}$, and models the *driving forces* which govern our system.

Lazy solution means that the trajectory chooses a velocity which has minimal norm among all possible directions offered by the differential inclusion. This type of differential inclusion occurs in various domains (mechanics, economics, control...), and has subsequently been the object of active research, see for example [5], [9], [13], [20]. Precisely, for any $u \in K$, the set $-N_K(u) - \text{Conv} \{\partial f_i(u)\}$ is a closed convex set, therefore it has a unique element of minimal norm, denoted as usual $(-N_K(u) - \text{Conv} \{\partial f_i(u)\})^0$. The direction $s(u) := (-N_K(u) - \text{Conv} \{\partial f_i(u)\})^0$ is called the *multiobjective steepest descent direction* at u , and the associated dynamical system

$$(2) \quad (\text{MOG}) \quad \dot{u}(t) + \left(N_K(u(t)) + \text{Conv} \{\partial f_i(u(t))\} \right)^0 = 0,$$

is called the Multi-Objective Gradient system, (MOG) for short. It was first investigated by Henry [31], Cornet [21], [22], [23], and Smale [38] in the seventies (in the finite dimensional case, and in the case where the objective functions f_i are smooth), as a dynamical model of allocation of resources in economics (planification procedure).

From the point of view of modeling, we will show that the (MOG) system has the following properties:

- a) It is a descent method, i.e., for *each* $i = 1, \dots, q$, $t \mapsto f_i(u(t))$ is nonincreasing.
- b) Its trajectories converge to weak Pareto optimal points.
- c) The scalarization of the multiobjective optimization problem is done endogeneously. At time t , the vector field which governs the system involves a convex combination $\sum_{i=1}^q \theta_i(t) \partial f_i(\cdot)$ of the subdifferential $\partial f_i(\cdot)$, with scalars $\theta_i(t)$ which are not fixed in advance. They are part of the process, whence the decentralized features of this dynamic.

Indeed, this system provides a weighting of the different criteria which offers applications in various domains, and which are still largely to explore. In inverse problems, signal/imaging processing, putting convenient weights on the data fitting term, and the regularization, or sparsity term is a central question. In game theory, economics, social science, management, the multiobjective steepest descent direction has attractive properties: it improves each of the objective functions, while putting a higher weight on the “weakest” agents, a key property of the interaction between cooperating agents. In addition, for each trajectory, the Pareto equilibrium which is finally reached, is not too far from the initial Cauchy data. In some particular situations, it is the projection of the initial data on the Pareto set.

Mathematical analysis of (MOG) gives rise to general statements whose formulation is simple, but some proofs are quite technical. The dynamic is governed by a vector field, $u \mapsto s(u)$, which is discontinuous. Indeed, the multivalued operators $u \mapsto N_K(u)$ and $u \mapsto \text{Conv} \{\partial f_i(u)\}$ are only upper semicontinuous (with closed graphs). Moreover, in general, $u \mapsto s(u)$ is not a gradient vector field, nor Lipschitz continuous, and $u \mapsto -s(u)$ is not a monotone operator. Let us list our main results concerning the (MOG) dynamical system. A section is devoted to each of them.

i) In Theorem 1.9, section 1, based on von Neumann’s minmax duality theorem, we provide equivalent formulations of the multiobjective steepest descent direction: instead of the subdifferential operators, they make use of the directional derivatives of the objective functions.

ii) In Theorem 2.2, section 2, we prove the weak convergence of the trajectories of (MOG) to weak Pareto optimal solutions of the constrained multiobjective optimization problem (CMO). Our proof is in the line of the proof of convergence of the steepest descent by Goudou and Munier [30] in the case of a single (quasi-convex) objective function; it makes use of Lyapunov analysis and Opial’s lemma.

iii) In Theorem 3.1, section 3, assuming further that \mathcal{H} is finite dimensional, we prove the existence of strong global solutions to (MOG). This is the more delicate part of the mathematical analysis. We provide a constructive proof which is based on the Yosida approximation of ∂f_i , and Peano existence theorem for differential equations. The regularized equations are relevant of the existence results which have been obtained in [6]. The difficult point is to pass to the limit on the

regularized differential equations, as the regularization parameter goes to zero, because the vector field which governs our dynamic is not continuous, nor monotone.

iv) In section 4, some modeling and numerical aspects are discussed for the (MOG) system. We first consider some connections between (MOG) and modeling in cooperative games, and inverse problems (signal/imaging processing). Then, by time discretization of (MOG), we introduce numerical algorithms for nonsmooth constrained multiobjective optimization, and make the link with the recent studies of Fliege and Svaiter [25], Grana Drummond and Svaiter [27], Bonnel, Iusem and Svaiter [14].

We end with a conclusion and some perspectives.

1. PARETO OPTIMALITY AND MULTIOBJECTIVE STEEPEST DESCENT

As a preliminary, let us make precise some classical notions of variational analysis. Given a proper lower-semicontinuous convex function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, its *subdifferential* $\partial f(u)$ is the closed convex subset of \mathcal{H} defined for any $u \in \mathcal{H}$ by

$$\partial f(u) = \{p \in \mathcal{H} : f(v) \geq f(u) + \langle p, v - u \rangle \quad \forall v \in \mathcal{H}\}.$$

In the special case where $f = \delta_K$ is the indicator function of a nonempty closed convex set $K \subset \mathcal{H}$, the subdifferential of δ_K at $u \in K$ is the *normal cone* to K at u , denoted $N_K(u)$. The subdifferential operator enjoys the following additivity rule: let f and g be two proper lower-semicontinuous convex functions such that one of them is continuous at a point belonging to the domain of the other, then

$$\partial(f + g)(v) = \partial f(v) + \partial g(v) \quad \text{for all } v \in \mathcal{H}.$$

Suppose now that f is locally Lipschitz continuous. Then the subdifferential of f at $u \in \mathcal{H}$ is a nonempty closed convex and bounded set. It is also interesting to consider the *directional derivative* of f at $u \in \mathcal{H}$ in the direction $d \in \mathcal{H}$, defined by

$$df(u, d) = \lim_{t \downarrow 0} \frac{f(u + td) - f(u)}{t}.$$

It is in duality with the subdifferential since $df(u, d)$ is equal to $\sup\{\langle p, d \rangle : p \in \partial f(u)\}$. Thus, in our context, this directional derivative takes only finite values. Furthermore, for all $u \in \mathcal{H}$, $df(u, \cdot)$ is convex and Lipschitz continuous on \mathcal{H} .

1.1. Pareto optimality. When considering problem (CMO), which is to minimize various cost functions on K , we seek a solution in the sense of Pareto, i.e., none of the objective functions can be improved in value without degrading some of the other objective values. It is a cooperative approach, the mathematical formulation is described below.

Definition 1.1. (*Pareto optimality*)

i) An element $u \in K$ is called *Pareto optimal* if there does not exist $v \in K$ such that $f_i(v) \leq f_i(u)$ for all $i = 1, \dots, q$, and $f_j(v) < f_j(u)$ for one $j \in 1, \dots, q$.

ii) An element $u \in K$ is called *weak Pareto optimal* if there does not exist $v \in K$ such that $f_i(v) < f_i(u)$ for all $i = 1, \dots, q$.

We equip \mathbb{R}^q with the order $y \preceq z \Leftrightarrow y_i \leq z_i$ for all $i = 1, \dots, q$, and the strict order relation $y \prec z \Leftrightarrow y_i < z_i$ for all $i = 1, \dots, q$. Then Pareto optimality admits an equivalent formulation: $u \in K$ is Pareto optimal iff there does not exist $v \in K$ such that $F(v) \preceq F(u)$ and $F(v) \neq F(u)$, while u is a weak Pareto optimum if there does not exist $v \in K$ such that $F(v) \prec F(u)$. These optimality notions can be generalized by considering orders generated by convex cones (see [39]).

In the case of a single objective function f , Pareto and weak Pareto optima coincide with the notion of global minimizer. Writing the necessary optimality condition leads to the notion of critical point, namely $N_K(u) + \partial f(u) \ni 0$. A similar approach exists for Pareto optimality:

Definition 1.2. Denote by $\mathcal{S}^q = \{\theta = (\theta_i) \in \mathbb{R}^q : 0 \leq \theta_i \leq 1, \sum_{i=1}^q \theta_i = 1\}$ the unit simplex in \mathbb{R}^q . We say that $u \in K$ is a Pareto critical point of the constrained multiobjective optimization problem (CMO) if there exists $(\theta_i) \in \mathcal{S}^q$ such that

$$(3) \quad N_K(u) + \sum_{i=1}^q \theta_i \partial f_i(u) \ni 0.$$

In the differentiable case, this notion has been considered by Smale in [38], Cornet in [21], see [11], [27], [39] for recent account of this notion, and various extensions of it. It is a multiobjective extension of the Fermat rule, and (see below) a first-order necessary optimality condition for (local) vectorial optimization. Note that equivalent formulations of this notion can be given, thanks to the positive homogeneity property of the formula: the condition $\sum_{i=1}^q \theta_i = 1$ can be dropped, just assuming the θ_i to be nonnegative, and at least one of them positive.

Let us respectively denote by \mathcal{P} , \mathcal{P}_w , and \mathcal{P}_c the set of Pareto optima, weak Pareto optima, and Pareto critical points. Clearly, $\mathcal{P} \subset \mathcal{P}_w$ always holds. In the case of convex multi-objective optimization, critical Pareto optimality is a necessary *and sufficient* condition for weak Pareto optimality. Let us state it in a precise way.

Lemma 1.3. Let $f_i : \mathcal{H} \rightarrow \mathbb{R}$, $i = 1, \dots, q$ be convex objective functions. Then $\mathcal{P}_w = \mathcal{P}_c$. Assuming further that the objective functions are strictly convex, then all these concepts of Pareto optimality coincide, i.e., $\mathcal{P} = \mathcal{P}_w = \mathcal{P}_c$.

Proof. The inclusion $\mathcal{P}_w \subset \mathcal{P}_c$ is obtained for convex functions by a direct application of the Hahn-Banach separation theorem (see for example [21, Proposition 1.1], [39], [11]). Let us prove the reverse inclusion. Let $u \in \mathcal{P}_c$. Then, u is a (global) solution of the convex minimization problem

$$(4) \quad \min \left\{ \sum_{i=1}^q \theta_i f_i(v) : v \in K \right\}$$

for some $\theta_i \in [0, 1]$ which are all nonnegative, and at least one of them positive. Indeed, (4) forces u to be a weak Pareto minimum. Otherwise, there would exist some $v \in K$ such that $f_i(v) < f_i(u)$ for all $i = 1, \dots, q$, which would imply (one uses the fact that at least one of the θ_i is positive) $\sum_{i=1}^q \theta_i f_i(v) < \sum_{i=1}^q \theta_i f_i(u)$, a clear contradiction. Now suppose that the objective functions are strictly convex, then $\sum_{i=1}^q \theta_i f_i$ is also strictly convex (we use again the fact that at least one of the θ_i is positive), and u is its unique minimizer over K . If we assume the existence of $v \in K$ such that $f_i(v) \leq f_i(u)$ for all $i = 1, \dots, q$, this would imply that v is also a minimizer of $\sum_{i=1}^q \theta_i f_i$ over K . Hence $v = u$, and u is Pareto optimal. \square

1.2. Multiobjective steepest descent direction. We discuss the concept of multiobjective descent direction, and present a multiobjective steepest descent direction, by analogy with the case of a single criterion. Considering the problem of minimizing a single objective f over K , we say that d is a descent direction at $u \in K$ when $df(u, d) < 0$, and d lies in the closed convex *tangent cone* to K at u , which is defined as the *polar cone* of $N_K(u)$

$$T_K(u) := \{v \in \mathcal{H} : \langle v, \eta \rangle \leq 0 \text{ for all } \eta \in N_K(u)\}.$$

Following Smale [38], let us generalize this notion of descent direction for the multiobjective optimization problem (CMO).

Definition 1.4. Considering the problem (CMO), we say that $d \in \mathcal{H}$ is a multiobjective descent direction at $u \in K$ if $df_i(u, d) < 0$ for each $i \in \{1, \dots, q\}$, and $d \in T_K(u)$.

Remark 1.5. Define the closed convex hull of the subdifferentials at $u \in \mathcal{H}$

$$\text{Conv} \{ \partial f_i(u); i = 1, \dots, q \} := \left\{ \sum_{i=1}^q \theta_i p_i : p_i \in \partial f_i(u), (\theta_i) \in \mathcal{S}^q \right\}.$$

To simplify the notation we just write $\text{Conv} \{ \partial f_i(u) \}$. Then, from the dual characterization $df_i(u, d) = \sup \{ \langle p, d \rangle : p \in \partial f_i(u) \}$ and the definition of $T_K(u)$, d is a multiobjective descent direction at u iff

$$\langle p, d \rangle < 0 \text{ and } \langle \eta, d \rangle \leq 0 \quad \forall p \in \text{Conv} \{ \partial f_i(u) \}, \forall \eta \in N_K(u).$$

It is therefore clear that no multiobjective descent direction can be found at a critical Pareto $u \in \mathcal{P}_c$, since this is equivalent to $0 \in N_K(u) + \text{Conv} \{ \partial f_i(u) \}$.

Let us define the vector field that governs our dynamical system.

Definition 1.6. For any $u \in K$, the unique element of minimal norm of the closed convex set $-N_K(u) - \text{Conv} \{ \partial f_i(u) \}$ is called the multiobjective steepest descent direction at u . It is denoted by

$$(5) \quad s(u) := \left(-N_K(u) - \text{Conv} \{ \partial f_i(u) \} \right)^0.$$

Note that, for any $u \in K$, the set $-N_K(u) - \text{Conv} \{ \partial f_i(u) \}$ is a closed convex set, as being equal to the vectorial sum of two closed convex sets, one of them being bounded. Hence, it has a unique element of minimal norm, and $s(u)$ is well defined. See Theorem 1.9 for an equivalent formulation of the multiobjective steepest descent direction that makes use of dual notions, namely the directional derivatives of the objective functions, and the tangent cone to K . This vector field clearly satisfies $u \in \mathcal{P}_c \Leftrightarrow s(u) = 0$. Furthermore, for any $u \notin \mathcal{P}_c$, $s(u)$ is a multiobjective descent direction:

Proposition 1.7. For all $u \in K$ we have

$$(6) \quad \langle s(u), p \rangle \leq -\|s(u)\|^2 \quad \text{for all } p \in \text{Conv} \{ \partial f_i(u) \}.$$

In particular, $s(u)$ is a multiobjective descent direction at any $u \in K \setminus \mathcal{P}_c$.

The above result has a simple geometrical interpretation. Take for simplicity the unconstrained problem, i.e., $K = \mathcal{H}$ and two criteria f_1, f_2 . Then $-s(u)$ is the orthogonal projection of the origin on the vectorial segment $[\nabla f_1(u), \nabla f_2(u)]$. By the classical result on the sum of the angles of a triangle, this forces the angles between $-s(u)$ and $\nabla f_i(u)$, $i = 1, 2$ to be acute, and hence $\langle s(u), \nabla f_i(u) \rangle \leq 0$. In order to prove Proposition 1.7, and in the following, we will make frequent use of the Moreau decomposition theorem [32].

Theorem 1.8. (Moreau) Let T be a closed convex cone of a real Hilbert space \mathcal{H} , and N be its polar cone, i.e., $N = \{ v \in \mathcal{H} : \langle v, \xi \rangle \leq 0 \text{ for all } \xi \in T \}$. Then, for all $v \in \mathcal{H}$ there exists a unique decomposition

$$\begin{aligned} v &= v_T + v_N, \quad v_T \in T, \quad v_N \in N; \\ \langle v_T, v_N \rangle &= 0. \end{aligned}$$

Moreover, $v_T = \text{proj}_T(v)$, and $v_N = \text{proj}_N(v)$.

Proof of Proposition 1.7. We introduce the notation $\mathcal{C}(u) = \text{Conv} \{ \partial f_i(u); i = 1, \dots, q \}$. By definition, $-s(u)$ is the projection of the origin onto the closed convex set $N_K(u) + \mathcal{C}(u)$. Hence, using that $0 \in N_K(u)$, we have for any $p \in \mathcal{C}(u)$

$$(7) \quad \langle 0 - (-s(u)), p - (-s(u)) \rangle \leq 0,$$

that is

$$(8) \quad \|s(u)\|^2 + \langle s(u), p \rangle \leq 0$$

which is the desired inequality. Verify now that $s(u)$ is a multiobjective descent direction at $u \in K \setminus \mathcal{P}_c$. By definition of $s(u)$, we can write $s(u) = \left(z - N_K(u)\right)^0$ for some $z \in -\mathcal{C}(u)$. Since $T_K(u)$ is the polar cone of $N_K(u)$, then by Moreau decomposition theorem,

$$\begin{aligned} \left(z - N_K(u)\right)^0 &= z - \text{proj}_{N_K(u)} z \\ &= \text{proj}_{T_K(u)} z \end{aligned}$$

which shows that $s(u) \in T_K(u)$, and concludes the proof. \square

Now that we have established that $s(u)$ is a multiobjective descent direction, one may wonder why it is called the *steepest* descent direction. Observe first that in the case of a single differentiable objective function f , and a constraint K , the direction $s(u)$ at $u \in K$ is given by

$$(9) \quad s(u) = (-N_K(u) - \nabla f(u))^0,$$

$$(10) \quad = \text{proj}_{T_K(u)}(-\nabla f(u)).$$

It is known that the normalized vector $\frac{s(u)}{\|s(u)\|}$ is the solution, when $\nabla f(u) \neq 0$, of the minimization problem

$$\min \{ df(u, d) : d \in T_K(u), \|d\| = 1 \},$$

whence the name of *steepest descent direction* for $\text{proj}_{T_K(u)}(-\nabla f(u))$. As shown below, this steepest descent property can be extended to the multiobjective case:

$$(11) \quad \frac{s(u)}{\|s(u)\|} = \text{argmin} \left\{ \max_{i=1, \dots, q} df_i(u, d) : d \in T_K(u), \|d\| = 1 \right\}.$$

Moreover, in the case of a single differentiable objective function, it can be easily verified that

$$(12) \quad s(u) = \text{argmin} \left\{ \frac{1}{2} \|v\|^2 + df(u, d) : d \in T_K(u) \right\},$$

and this further characterization will also be generalized to the multiobjective case. In addition, we make a link between the formulations (12) and (11), by introducing a continuum of characterizations based on the use of $\|\cdot\|^r$, these two situations corresponding to $r = 2$, and the limiting case $r = +\infty$.

Theorem 1.9. *Let $u \in K \setminus \mathcal{P}_c$. Then $s(u)$ can be formulated in the following equivalent forms:*

$$(13) \quad 1. \quad s(u) = \left(-N_K(u) - \text{Conv} \{ \partial f_i(u) \} \right)^0$$

$$(14) \quad 2. \quad \frac{s(u)}{\|s(u)\|^{\frac{r-2}{r-1}}} = \text{argmin}_{d \in T_K(u)} \left\{ \frac{1}{r} \|d\|^r + \max_{i=1, \dots, q} df_i(u, d) \right\} \text{ for all } r \in]1, +\infty[$$

$$(15) \quad 3. \quad \frac{s(u)}{\|s(u)\|} = \text{argmin}_{\substack{v \in T_K(u) \\ \|v\|=1}} \left\{ \max_{i=1, \dots, q} df_i(u, d) \right\}.$$

Remark 1.10. The equivalence between formulations 1. and 3. of the steepest descent direction has been first obtained, in the finite dimensional and smooth case, by Cornet in [21, Proposition 3.1]. Formulation 2 appears in Fliege and Svaiter [25] in the smooth case, for $r = 2$:

$$s(u) = \text{argmin}_{d \in T_K(u)} \left\{ \frac{1}{2} \|d\|^2 + \max_{i=1, \dots, q} df_i(u, d) \right\},$$

but the equivalence between formulations 1. and 2. is seemingly new. The second formulation for $r \neq 2$ is new, although it was stressed in [25] that $\frac{1}{2} \|d\|^2$ could be replaced by any positive proper l.s.c strictly convex function which is dominated by the norm around the origin. The interest of

considering r arbitrary large is that we can see -at least formally- the third formulation as the limit of the second when $r \rightarrow +\infty$: $\frac{r-2}{r-1}$ tends to 1, while the function $\frac{1}{r}\|\cdot\|^r$ is pointwise converging to the indicator function of the unit ball $\delta_{\mathbb{B}(0,1)}(\cdot)$.

The following proof is based on duality arguments (von Neumann's min-max theorem) which were first introduced in the smooth differentiable case in [21]. The extension of these results to the non-smooth case is nontrivial and requires some adjustments. In addition, the third formulation is obtained from the second, using an epiconvergence (Γ -convergence) argument.

Proof of Theorem 1.9. In all that follows, $\mathcal{C}(u)$ denotes $\text{Conv}\{\partial f_i(u)\}$, where u is a fixed element of $K \setminus \mathcal{P}_c$. As a consequence, $s(u) \neq 0$.

Let us start by proving item 2. Because of the powered norm term, $d \mapsto \frac{1}{r}\|d\|^r + \max_{i=1,\dots,q} df_i(u, d)$ is a coercive strictly convex function. Therefore, there exists a unique solution \bar{d} to the minimization problem

$$(16) \quad \min_{d \in T_K(u)} \left\{ \frac{1}{r}\|d\|^r + \max_{i=1,\dots,q} \max_{p_i \in \partial f_i(u)} \langle p_i, d \rangle \right\}.$$

Let us show that $\bar{d} = \frac{s(u)}{\|s(u)\|^{\frac{r-2}{r-1}}}$. We use a duality argument which relies on the equivalent formulation of (16) as the convex-concave saddle value problem

$$(17) \quad \min_{d \in T_K(u)} \max_{p \in \mathcal{C}(u)} \left\{ \frac{1}{r}\|d\|^r + \langle p, d \rangle \right\}.$$

It is associated with the convex-concave Lagrangian function

$$L(d, p) = \frac{1}{r}\|d\|^r + \langle p, d \rangle$$

defined on $T_K(u) \times \mathcal{C}(u)$. Since L is convex and coercive with respect to the first variable, and $\mathcal{C}(u)$ is bounded, by the von Neumann's minimax theorem (see [3, Theorem 9.7.1]) there exists $\bar{p} \in \mathcal{C}(u)$ such that (\bar{d}, \bar{p}) is a saddle point of (17), that is

$$(18) \quad \inf_{d \in T_K(u)} L(d, \bar{p}) = L(\bar{d}, \bar{p}) = \sup_{p \in \mathcal{C}(u)} L(\bar{d}, p).$$

For any $p \in \mathcal{C}(u)$ let us define

$$(19) \quad d(p) := \operatorname{argmin}_{d \in T_K(u)} \left\{ \frac{1}{r}\|d\|^r + \langle p, d \rangle \right\}.$$

Writing down the optimality condition for the above primal problem gives

$$(20) \quad d(p) = \operatorname{proj}_{T_K(u)} \left(\frac{-p}{\|d(p)\|^{r-2}} \right),$$

which, by Moreau's theorem, can be rewritten as

$$(21) \quad d(p) = \frac{1}{\|d(p)\|^{r-2}} (-p - N_K(u))^0.$$

Observe that (\bar{d}, \bar{p}) being a saddle point of L implies $\bar{d} = d(\bar{p})$. Thus we just need to prove that $d(\bar{p}) = \frac{s(u)}{\|s(u)\|^{\frac{r-2}{r-1}}}$. To identify \bar{p} , we use the dual formulation

$$(22) \quad \bar{p} = \operatorname{argmax}_{p \in \mathcal{C}(u)} \min_{d \in T_K(u)} \left\{ \frac{1}{r}\|d\|^r + \langle p, d \rangle \right\},$$

which, by (19) and (20), can be rewritten as

$$(23) \quad \bar{p} = \operatorname{argmax}_{p \in \mathcal{C}(u)} \frac{1}{r} \|d(p)\|^r - \|d(p)\|^{r-2} \left\langle \frac{-p}{\|d(p)\|^{r-2}}, \operatorname{proj}_{T_K(u)} \left(\frac{-p}{\|d(p)\|^{r-2}} \right) \right\rangle.$$

Using Moreau's theorem, we obtain

$$(24) \quad \bar{p} = \operatorname{argmax}_{p \in \mathcal{C}(u)} \frac{1}{r} \|d(p)\|^r - \|d(p)\|^{r-2} \left\| \operatorname{proj}_{T_K(u)} \left(\frac{-p}{\|d(p)\|^{r-2}} \right) \right\|^2,$$

which, by (20) and $r \in]1, +\infty]$, is equivalent to

$$(25) \quad \bar{p} = \operatorname{argmin}_{p \in \mathcal{C}(u)} \|d(p)\|^{r-1}.$$

From (21,) we know that $\|d(p)\|^{r-1} = \|(-p - N_K(u))^0\|$. Therefore, $s(u) = (-\bar{p} - N_K(u))^0$ with $\|d(\bar{p})\|^{r-1} = \|s(u)\|$. Using again (21), we obtain $d(\bar{p}) = \frac{s(u)}{\|s(u)\|^{\frac{r-2}{r-1}}}$, as expected.

Let us complete the proof by proving the third characterisation. As we said in Remark 1.10, it relies on a limit argument. Define, for any $r > 1$, the functions $F_r : d \in \mathcal{H} \mapsto \frac{1}{r} \|d\|^r + \max_{i=1, \dots, q} df_i(u, d) + \delta_{T_K(u)}(d)$. It can be easily verified that the sequence $(F_r)_{r>1}$ epiconverges when $r \rightarrow +\infty$ to

$$F : d \in \mathcal{H} \mapsto \delta_{\{\|\cdot\| \leq 1\}}(d) + \max_{i=1, \dots, q} df_i(u, d) + \delta_{T_K(u)}(d).$$

From (14) and [3, Theorem 12.1.1] we can deduce that

$$(26) \quad \frac{s(u)}{\|s(u)\|} = \operatorname{argmin}_{\substack{v \in T_K(u) \\ \|v\| \leq 1}} \left\{ \max_{i=1, \dots, q} df_i(u, d) \right\},$$

where the inequality constraint $\|v\| \leq 1$ can be replaced by $\|v\| = 1$, since $\frac{s(u)}{\|s(u)\|}$ is a normalized vector. □

1.3. The Multi-Objective Gradient dynamic. Here we present and discuss the continuous dynamic governed by the multiobjective steepest descent vector field $u \mapsto s(u)$.

In [38], Smale defined the notion of *gradient process* for the multiobjective optimization problem (CMO). It is a differential equation

$$(27) \quad \dot{u}(t) = \phi(u(t))$$

where $\phi : K \rightarrow \mathcal{H}$ is a mapping which satisfies the following properties:

$$(28) \quad \begin{cases} \phi(u) \text{ is a multiobjective descent direction whenever } u \notin \mathcal{P}_c, \\ \phi(u) = 0 \text{ if } u \in \mathcal{P}_c. \end{cases}$$

The interest of such a gradient process is twofold : the stationary points of the dynamic are exactly the critical Pareto points, and as long as $u(t)$ is not a critical Pareto point, all the objective functions are decreasing. Clearly, from its definition and Proposition 1.7, the vector field $u \mapsto s(u)$ induces a gradient process, defined as follows :

Definition 1.11. *The dynamical system which is governed by the vector field $u \mapsto s(u)$, is called the Multi-Objective Gradient system. Its solution trajectories $t \mapsto u(t)$ verify*

$$(29) \quad (\text{MOG}) \quad \dot{u}(t) + \left(N_K(u(t)) + \operatorname{Conv} \{ \partial f_i(u(t)) \} \right)^0 = 0.$$

Remark 1.12. Instead of considering the vector field $u \mapsto s(u)$ to govern our dynamic, we could have chosen one of the directions that appear in Theorem 1.9. In fact, each of them induces a gradient process. From the viewpoint of the dynamic system, these directions generate the same integral curves, with a different time scale.

In [38], the vector field ϕ governing the gradient process is continuous, in a finite dimensional setting. In our context, the corresponding notions have been extended in order to cover dynamical systems governed by a discontinuous vector field on a general Hilbert space, as the (MOG) dynamic. In particular, instead of classical (continuously differentiable) solutions, we will consider strong solutions (absolutely continuous on bounded time intervals), the equality (29) being satisfied almost everywhere. Let us make precise this (see [16, Appendix] for more details):

Definition 1.13. *Given $T \in \mathbb{R}^+$, a function $u : [0, T] \rightarrow \mathcal{H}$ is said to be absolutely continuous if one of the following equivalent properties holds:*

- i) there exists an integrable function $g : [0, T] \rightarrow \mathcal{H}$ such that $u(t) = u(0) + \int_0^t g(s) ds \quad \forall t \in [0, T]$;*
- ii) u is continuous and its distributional derivative belongs to the Lebesgue space $L^1([0, T]; \mathcal{H})$;*
- iii) for every $\epsilon > 0$, there exists $\eta > 0$ such that for any finite family of intervals $I_k = (a_k, b_k)$, $I_k \cap I_j = \emptyset$ for $k \neq j$ and $\sum_k |b_k - a_k| \leq \eta \implies \sum_k \|u(b_k) - u(a_k)\| \leq \epsilon$.*

Moreover, an absolutely continuous function is differentiable almost everywhere, its derivative coincide with its distributional derivative almost everywhere, and one can recover the function from its derivative $u' = g$ using the integration formula (i). We can now make precise the notion of solution for the (MOG) dynamic (recall that \mathcal{S}^q denotes the unit simplex in \mathbb{R}^q).

Definition 1.14. *We say that $u(\cdot)$ is a strong global solution of (MOG) if the following properties are satisfied:*

- (i) $u : [0, +\infty[\rightarrow \mathcal{H}$ is absolutely continuous on each interval $[0, T]$, $0 < T < +\infty$;*
- (ii) there exists $\eta : [0, +\infty[\rightarrow \mathcal{H}$, $v_i : [0, +\infty[\rightarrow \mathcal{H}$, $\theta_i : [0, +\infty[\rightarrow [0, 1]$ $i=1, 2, \dots, q$ which satisfy*

$$(30) \quad \theta_i \in L^\infty(0, +\infty; \mathbb{R}), \quad (\theta_i(t)) \in \mathcal{S}^q \text{ for almost all } t > 0;$$

$$(31) \quad v_i \in L^\infty(0, T; \mathcal{H}), \quad \eta \in L^2(0, T; \mathcal{H}) \text{ for all } T > 0 \text{ and all } i = 1, 2, \dots, q;$$

$$(32) \quad \eta(t) \in N_K(u(t)), \quad v_i(t) \in \partial f_i(u(t)) \text{ for almost all } t > 0;$$

$$(33) \quad \dot{u}(t) + \eta(t) + \sum_i \theta_i(t) v_i(t) = 0 \quad \text{for almost all } t > 0;$$

$$(34) \quad \dot{u}(t) + (N_K(u(t)) + \text{Conv} \{\partial f_i(u(t))\})^0 = 0 \quad \text{for almost all } t > 0.$$

Now we can establish the first qualitative properties of strong solutions of (MOG). First, we show that trajectories satisfy a local Lipschitz continuity property. Second, as announced, we show that the objective functions are decreasing along the trajectories.

Proposition 1.15. *Let us make assumptions H0), H1). Then for any strong global solution $t \in [0, +\infty[\mapsto u(t) \in \mathcal{H}$ of (MOG), the following properties hold:*

- i) Descent property: for each $i = 1, \dots, q$, $t \mapsto f_i(u(t))$ is a nonincreasing absolutely continuous function, and for almost all $t > 0$*

$$(35) \quad \frac{d}{dt} f_i(u(t)) \leq -\|\dot{u}(t)\|^2.$$

- ii) Lipschitz continuity: The trajectory u is Lipschitz continuous on any finite time interval $[0, T]$. If moreover it is bounded, u is Lipschitz continuous on $[0, +\infty[$.*

This Proposition is a direct consequence of Proposition 1.7 and the following generalized chain rule from Brézis:

Lemma 1.16. [16, Lemma 4, p.73] *Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a closed convex proper function. Let $u \in L^2(0, T; \mathcal{H})$ be such that $\dot{u} \in L^2(0, T; \mathcal{H})$, and $u(t) \in \text{dom}(\partial\Phi)$ for a.e. t . Assume that there exists $\xi \in L^2(0, T; \mathcal{H})$ such that $\xi(t) \in \partial\Phi(u(t))$ for a.e. t . Then the function $t \mapsto \Phi(u(t))$ is absolutely continuous, and for every t such that u and $\Phi(u)$ are differentiable at t , and $u(t) \in \text{dom}(\partial\Phi)$, we have*

$$\forall p \in \partial\Phi(u(t)), \quad \frac{d}{dt}\Phi(u(t)) = \langle \dot{u}(t), p \rangle.$$

Proof of Proposition 1.15. i) By definition of (MOG), for almost all $t > 0$, $\dot{u}(t) = s(u(t))$ holds. Hence, using Proposition 1.7, for any $p \in \text{Conv}\{\partial f_i(u(t))\}$

$$(36) \quad \|\dot{u}(t)\|^2 + \langle \dot{u}(t), p \rangle \leq 0.$$

Moreover, for almost all $t > 0$, there exists $v_i(t) \in \partial f_i(u(t))$ with $v_i \in L^2(0, T; \mathcal{H})$. Hence taking $p = v_i(t) \in \text{Conv}\{\partial f_i(u(t))\}$ in (36) yields

$$(37) \quad \|\dot{u}(t)\|^2 + \langle \dot{u}(t), v_i(t) \rangle \leq 0.$$

The derivation chain rule is valid in our situation, see Lemma 1.16. Hence, $f_i(u)$ is absolutely continuous on each bounded interval $[0, T]$, which, by (37), gives for almost all $t > 0$

$$(38) \quad \|\dot{u}(t)\|^2 + \frac{d}{dt}f_i(u(t)) \leq 0.$$

As a consequence, $\frac{d}{dt}f_i(u(t)) \leq 0$, and for each $i = 1, \dots, q$ the function $t \mapsto f_i(u(t))$ is nonincreasing. ii) By Definition 1.14 of a strong global solution, we can write

$$\dot{u}(t) + \eta(t) + \sum_i \theta_i(t)v_i(t) = 0$$

with $\eta(t) \in N_K(u(t))$, $v_i(t) \in \partial f_i(u(t))$, $(\theta_i(t)) \in \mathcal{S}^q$ for almost all $t > 0$. Let us argue on some $[0, T]$. Since ∂f_i is bounded on bounded sets, and $(\theta_i(t)) \in \mathcal{S}^q$, we have

$$\dot{u}(t) + \eta(t) = g(t)$$

with $g := -\sum_i \theta_i v_i \in L^\infty(0, T; \mathcal{H})$. Taking the scalar product with $\dot{u}(t)$, we obtain

$$(39) \quad \|\dot{u}(t)\|^2 + \langle \eta(t), \dot{u}(t) \rangle = \langle g(t), \dot{u}(t) \rangle.$$

Note that the normal cone mapping N_K is the subdifferential of δ_K , the indicator function of K . Using the derivation chain rule, we have

$$(40) \quad \langle \eta(t), \dot{u}(t) \rangle = \frac{d}{dt}\delta_K(u(t)) = 0.$$

Combining (39), (40) and the Cauchy-Schwarz inequality, we obtain

$$\|\dot{u}(t)\| \leq \|g(t)\|.$$

Hence, $\dot{u} \in L^\infty(0, T; \mathcal{H})$, which is equivalent to the Lipschitz continuity of u . If u is bounded, just notice that $g \in L^\infty(0, +\infty; \mathcal{H})$, and conclude in a similar way. \square

1.4. Examples. We now illustrate the (MOG) dynamic through some simple examples in $\mathcal{H} = \mathbb{R} \times \mathbb{R}$. They suggest that its study is nontrivial, because (MOG) is governed by a vector field that is neither monotone, nor locally Lipschitz: without any further assumption, we cannot expect more than the Hölder continuity of this field vector (see [6], and Example 3 below).

Example 1. Take the quadratic functions $f_1(x, y) = \frac{1}{2}(x+1)^2 + \frac{1}{2}y^2$ and $f_2(x, y) = \frac{1}{2}(x-1)^2 + \frac{1}{2}y^2$. The corresponding Pareto set is $\mathcal{P} = \mathcal{P}_w = [-1, +1] \times \{0\}$ and the steepest descent is given by :

$$(41) \quad s(x, y) = \begin{cases} -(x-1, y) & \text{if } x > 1, \\ -(0, y) & \text{if } -1 \leq x \leq 1, \\ -(x+1, y) & \text{if } x < -1. \end{cases}$$

Figure 1 shows some trajectories of the (MOG) dynamic. Trajectories are straight lines connecting the starting point and its projection on \mathcal{P} . On this example, we can observe that (MOG) is different from the descent dynamics associated with a scalarized function $\alpha f_1 + f_2$, $\alpha > 0$. It is neither related to the descent dynamic associated to the max function $f = \max f_i$. Indeed, when starting from some (x_0, y_0) with $x_0 > 1$, the trajectory of the steepest descent for $f = \max f_i$ is first oriented toward the Pareto equilibrium $(-1, 0)$, while the trajectory of (MOG) is oriented toward $(1, 0)$.

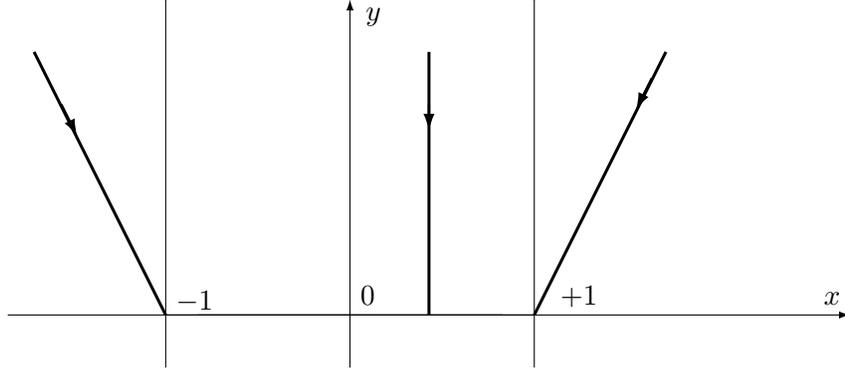


Figure 1

Example 2. Let $f_1(x, y) = \frac{1}{2}x^2$ and $f_2(x, y) = \frac{1}{2}y^2$. Here $\mathcal{P} = \{(0, 0)\}$ and $\mathcal{P}_w = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$. The multiobjective steepest descent vector field, once computed, is:

$$s(x, y) = -\left(\frac{xy^2}{x^2 + y^2}, \frac{yx^2}{x^2 + y^2}\right) \text{ if } (x, y) \neq (0, 0), \quad s(0, 0) = (0, 0).$$

Observe that $(x, y) \mapsto s(x, y)$ is a nonlinear vector field, and it is not a gradient vector field. Moreover, trajectories tend to move away from each other (see Figure 2), which reflects the fact that $(x, y) \mapsto -s(x, y)$ is not a monotone operator. Indeed, for $x > 0$, $y > 0$, $x \neq y$

$$\langle -s(x, y) + s(y, x), (x, y) - (y, x) \rangle = -2 \frac{xy(x-y)^2}{x^2 + y^2} < 0.$$

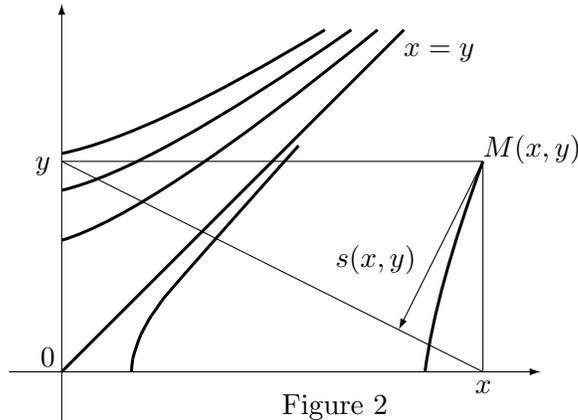


Figure 2

Example 3. Let $f_1(x, y) = \frac{1}{2}(x^2 + y^2)$ and $f_2(x, y) = x$. The corresponding Pareto set is $\mathcal{P} = \mathcal{P}_w =]-\infty, 0] \times \{0\}$. Once computed, we see that the steepest descent vector field is defined according to three areas of the plane (see Figure 3):

$$(42) \quad s(x, y) = \begin{cases} -(1, 0) & \text{if } x \geq 1, \\ -(x, y) & \text{if } (x - \frac{1}{2})^2 + y^2 \leq \frac{1}{4}, \\ \frac{-1}{(x-1)^2 + y^2} (y^2, y(1-x)) & \text{else.} \end{cases}$$

As in the previous example, this vector field is neither linear nor a gradient vector field, or monotonous. Moreover, it is not locally Lipschitz. The lack of Lipschitz continuity occurs at the point $(1, 0)$, where the vector field "splits" into three parts. Figure 4 provides a simple example of parameterized vectors u_θ, v_θ that converge both to $(1, 0)$ when θ goes to zero, but such that $\|s(u_\theta) - s(v_\theta)\| = \sin(\theta)$ and $\|u_\theta - v_\theta\| = \sin(\theta) \tan(\theta)$. As a consequence, $\frac{\|s(u_\theta) - s(v_\theta)\|}{\|u_\theta - v_\theta\|} \simeq \frac{\theta}{\theta^2}$ is unbounded when $\theta \rightarrow 0$.

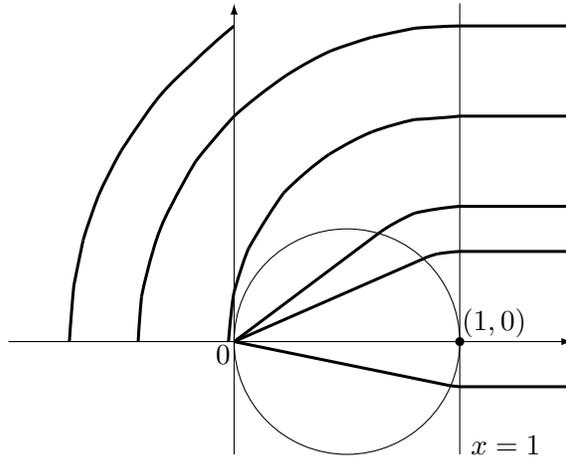


Figure 3

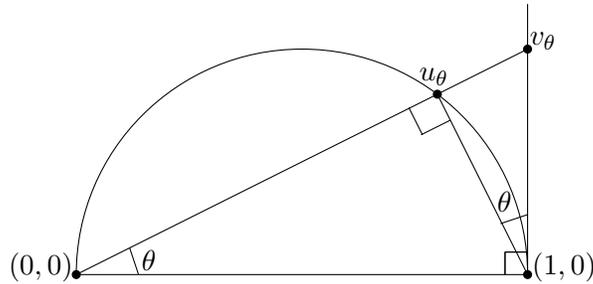


Figure 4

1.5. Related dynamics. When there is just one objective function f , since $\partial f(u(t))$ is a closed convex set, the (MOG) system specializes to

$$\dot{u}(t) + \left(N_K(u(t)) + \partial f(u(t)) \right)^0 = 0.$$

Indeed, this system is equivalent to

$$\dot{u}(t) + N_K(u(t)) + \partial f(u(t)) \ni 0,$$

because, in this case, the lazy solution property is automatically satisfied by the trajectories of the semigroup of contractions generated by the maximal monotone operator $N_K + \partial f$, see [16, Theorem 3.1]. In particular, our existence and asymptotic analysis for (MOG) in Sections 2 and 3 extends the well-known results for the nonsmooth gradient flow (see [16]).

This leads us to ask a natural question, which is the study of the relationship (or differences) between (MOG) and the Multiobjective Differential Inclusion ((MDI) for short)

$$(43) \quad (\text{MDI}) \quad \dot{u}(t) + N_K(u(t)) + \text{Conv} \{ \partial f_i(u(t)) \} \ni 0.$$

It appears that (MDI) enjoys a weaker form of Proposition 1.15 *i* :

Proposition 1.17. *Let $t \in [0, +\infty[\mapsto u(t) \in \mathcal{H}$ be a strong global solution of (MDI) in the sense of Definition 1.14 (except the lazy property). Then for almost all $t \geq 0$, such that $u(t) \notin \mathcal{P}_c$, there exists some $i \in \{1, \dots, q\}$ (which depends on t) such that*

$$\frac{d}{dt} f_i(u(t)) < 0.$$

Proof. Since u is a strong solution of (MDI), there exists $\eta : [0, +\infty[\rightarrow \mathcal{H}$, $v_i : [0, +\infty[\rightarrow \mathcal{H}$, $\theta_i : [0, +\infty[\rightarrow [0, 1]$ $i = 1, 2, \dots, q$ which satisfy

$$\begin{aligned} \theta_i &\in L^\infty(0, +\infty; \mathbb{R}), \quad (\theta_i(t)) \in \mathcal{S}^q \text{ for almost all } t > 0; \\ v_i &\in L^\infty(0, T; \mathcal{H}), \quad \eta \in L^2(0, T; \mathcal{H}) \text{ for all } T > 0 \text{ and all } i = 1, 2, \dots, q; \\ \eta(t) &\in N_K(u(t)), \quad v_i(t) \in \partial f_i(u(t)) \quad \text{for almost all } t > 0; \\ \dot{u}(t) + \eta(t) + \sum_i \theta_i(t) v_i(t) &= 0 \quad \text{for almost all } t > 0; \end{aligned}$$

Taking the scalar product of the above equation with $\dot{u}(t)$, we obtain

$$(44) \quad \|\dot{u}(t)\|^2 + \langle \eta(t), \dot{u}(t) \rangle + \sum_i \theta_i(t) \langle v_i(t), \dot{u}(t) \rangle = 0.$$

The derivation chain rule is valid in our situation, see Lemma 1.16. Hence, $f_i(u)$ is absolutely continuous on each bounded interval $[0, T]$, which gives, for almost all $t > 0$

$$(45) \quad \frac{d}{dt} f_i(u(t)) = \langle v_i(t), \dot{u}(t) \rangle.$$

By a similar argument using the indicator function δ_K of K (recall (40)) we have

$$(46) \quad 0 = \frac{d}{dt} \delta_K(u(t)) = \langle \eta(t), \dot{u}(t) \rangle.$$

Combining (44) with (45), (46), we obtain

$$(47) \quad \|\dot{u}(t)\|^2 + \sum_i \theta_i(t) \frac{d}{dt} f_i(u(t)) \leq 0.$$

Since $u(t) \notin \mathcal{P}_c$, we have $\dot{u}(t) \neq 0$. Hence

$$(48) \quad \sum_i \theta_i(t) \frac{d}{dt} f_i(u(t)) < 0.$$

Since $(\theta_i(t)) \in \mathcal{S}^q$, this clearly implies that at least one of the derivatives $\frac{d}{dt} f_i(u(t))$ is negative. \square

Remark 1.18. Proposition 1.17 tells us that, for any trajectory of (MDI), for almost all $t > 0$, at least one of the objective functions decreases. We will illustrate this on a few examples, and highlight the fact that, by contrast, for (MOG) trajectories, they are all decreasing.

a) Consider the steepest descent dynamic associated to one of the objective functions, say f_i , i being fixed. Clearly, $\partial f_i(u) \subset \text{Conv} \{ \partial f_j(u) \}$, and the corresponding trajectories are solutions of

(MDI). The strategy consisting in taking care of only one objective function f_i , clearly leads to Pareto equilibria, but fails in general to improve all the objective functions.

b) The scalarization approach consists in taking a constant convex combination of the objective functions $f_\theta = \sum_{i=1}^q \theta_i f_i$, with $\theta \in \mathcal{S}^q$. The sum rule for continuous convex functions gives, for any $u \in \mathcal{H}$

$$(49) \quad \partial f_\theta(u) = \sum_{i=1}^q \theta_i \partial f_i(u),$$

and clearly $\partial f_\theta(u) \subset \text{Conv} \{ \partial f_i(u) \}$. By Bruck's theorem [19], any orbit of the generalized gradient flow generated by ∂f_θ converges to a minimizer of f_θ , which, by Lemma 1.3, is a weak Pareto optimal point. But, in general, this approach fails to improve all the objective functions. Take for instance in Example 1 any $\theta = (\lambda, (1 - \lambda))$ for $\lambda \in]0, 1]$. When starting from $(1, 0)$, the trajectory goes straight to $(1 - 2\lambda, 0)$ by decreasing f_1 but increasing f_2 .

c) Consider the steepest descent dynamic associated to the function $f = \max_i f_i$. This dynamic has some similarities with (MOG), but it is different. As a supremum of a finite number of convex continuous functions, f is still convex continuous. The classical subdifferential rule for the supremum of convex functions, see for example [12, Theorem 18.5], gives (in our setting)

$$(50) \quad \partial f(u) = \text{Conv} \{ \partial f_i(u) : i \in I(u) \}$$

where $I(u) = \{ i \in I : f_i(u) = f(u) \}$ is the set of the active indices at u . Clearly $\partial f(u) \subset \text{Conv} \{ \partial f_i(u) \}$. As a consequence, the trajectories of the steepest descent for $f = \max f_i$ are also solutions of (MDI). But, in general, they fail to satisfy that all the objective functions are decreasing. Take for instance Example 1: when starting from some (x_0, y_0) with $x_0 > y_0 > 1$, along the trajectory f_2 is first decreasing, until the current point reaches the projection of $(1, 0)$ on the line segment joining (x_0, y_0) to $(-1, 0)$, then it is increasing.

d) As shown by the above examples, (MDI) provides diversity, an interesting feature for evolutionary processes, and generating the whole Pareto set, see [18].

2. ASYMPTOTIC CONVERGENCE TO A WEAK PARETO MINIMUM

In this section, we study the asymptotic behavior (as $t \rightarrow +\infty$) of the strong global solutions of (MOG). We take for granted their existence, this question being examined into detail in section 3. In order to prove the weak convergence of the trajectories of (MOG), we use the classical Opial's lemma [33]. We recall its statement in its continuous form, and give a short proof of it:

Lemma 2.1. *Let S be a non empty subset of \mathcal{H} , and $u : [0, +\infty[\rightarrow \mathcal{H}$ a map. Assume that*

- (i) *for every $z \in S$, $\lim_{t \rightarrow +\infty} \|u(t) - z\|$ exists;*
- (ii) *every weak sequential cluster point of the map u belongs to S .*

Then

$$w - \lim_{t \rightarrow +\infty} u(t) = u_\infty \quad \text{exists, for some element } u_\infty \in S.$$

Proof. By (i) and $S \neq \emptyset$, the trajectory u is bounded in \mathcal{H} . In order to obtain its weak convergence, we just need to prove that the trajectory has a unique weak sequential cluster point. Let $u(t_n^1) \rightharpoonup z^1$ and $u(t_n^2) \rightharpoonup z^2$, with $t_n^1 \rightarrow +\infty$, and $t_n^2 \rightarrow +\infty$. By (ii), $z^1 \in S$, and $z^2 \in S$. By (i), it follows that $\lim_{t \rightarrow +\infty} \|u(t) - z^1\|$ and $\lim_{t \rightarrow +\infty} \|u(t) - z^2\|$ exist. Hence, $\lim_{t \rightarrow +\infty} (\|u(t) - z^1\|^2 - \|u(t) - z^2\|^2)$ exists. Developing and simplifying this last expression, we deduce that

$$\lim_{t \rightarrow +\infty} \langle u(t), z^2 - z^1 \rangle \quad \text{exists.}$$

Hence

$$\lim_{n \rightarrow +\infty} \langle u(t_n^1), z^2 - z^1 \rangle = \lim_{n \rightarrow +\infty} \langle u(t_n^2), z^2 - z^1 \rangle,$$

which gives $\|z^2 - z^1\|^2 = 0$, and hence $z^2 = z^1$. \square

We can now state our main convergence result.

Theorem 2.2. *Let us make assumptions H0), H1), H2). Then for any strong global solution $t \in [0, +\infty[\mapsto u(t) \in \mathcal{H}$ of (MOG), the following properties hold:*

i) Finite energy property:

$$(51) \quad \int_0^{+\infty} \|\dot{u}(t)\|^2 dt < +\infty.$$

ii) Weak convergence: Assume that the trajectory $t \in [0, +\infty[\mapsto u(t) \in \mathcal{H}$ is bounded in \mathcal{H} . Then $u(t)$ converges weakly in \mathcal{H} as $t \rightarrow +\infty$ to a weak Pareto optimum.

Remark 2.3. a) Since each function $t \mapsto f_i(u(t))$ is nonincreasing (see Proposition 1.15), a natural condition insuring that the trajectory remains bounded is that one of the functions f_i has bounded sublevel sets (see also Remark 2.5).

b) Similarly, if one of the functions f_i has relatively compact sublevel sets (inf-compactness property), then the trajectory is relatively compact, and hence converges strongly in \mathcal{H} . It is an interesting (open) question to find other conditions on the data (f_i and K) which provide strong convergence of trajectories, and extend the well-known conditions in the case of a single criterion.

Proof. From Proposition 1.15 and by integrating (35), along with the fact that f_i is bounded from below on K , we obtain

$$(52) \quad \int_0^{+\infty} \|\dot{u}(t)\|^2 dt \leq f_i(u(0)) - \inf_K f_i.$$

This proves items *i)*.

Let us now prove the weak convergence of any bounded trajectory u of the (MOG) system. To that end we use Opial's Lemma 2.1 with

$$(53) \quad S = \left\{ v \in K : \forall i = 1, \dots, q \quad f_i(v) \leq \inf_{t \geq 0} f_i(u(t)) \right\}.$$

Functions f_i are convex continuous, and hence lower semicontinuous for the weak topology of \mathcal{H} . As well, the closed convex set K is closed for the weak topology of \mathcal{H} . The trajectory $t \in [0, +\infty[\mapsto u(t) \in \mathcal{H}$ has been assumed to be bounded in \mathcal{H} . As a consequence, every weak sequential cluster point of the trajectory belongs to S , which is a closed convex non empty subset of \mathcal{H} .

i) Take $z \in S$ and set, for any $t \geq 0$

$$(54) \quad h_z(t) = \frac{1}{2} \|u(t) - z\|^2.$$

We have

$$(55) \quad \dot{h}_z(t) = \langle \dot{u}(t), u(t) - z \rangle.$$

Since u is a solution of (MOG), for almost all $t > 0$ there exists

$$(56) \quad \eta(t) \in N_K(u(t)), \text{ and } v_i(t) \in \partial f_i(u(t)), \quad (\theta_i(t)) \in \mathcal{S}^q$$

such that,

$$(57) \quad \dot{u}(t) + \sum_{i=1}^q \theta_i(t) v_i(t) + \eta(t) = 0.$$

By combining (55) and (57) we obtain

$$(58) \quad \dot{h}_z(t) + \sum_{i=1}^q \theta_i(t) \langle v_i(t), u(t) - z \rangle + \langle \eta(t), u(t) - z \rangle = 0.$$

On the one hand, since $\eta(t) \in N_K(u(t))$ and $z \in K$

$$(59) \quad \langle \eta(t), u(t) - z \rangle \geq 0.$$

On the other hand, the convex subdifferential inequality at $u(t)$, and $v_i(t) \in \partial f_i(u(t))$ gives

$$(60) \quad f_i(z) \geq f_i(u(t)) + \langle v_i(t), z - u(t) \rangle.$$

Since $z \in S$ we have $f_i(z) \leq f_i(u(t))$, which gives

$$(61) \quad \langle v_i(t), z - u(t) \rangle \leq 0.$$

As a consequence

$$(62) \quad \sum_{i=1}^q \theta_i(t) \langle v_i(t), u(t) - z \rangle \geq 0.$$

Combining (58) with (59) and (62) we obtain

$$(63) \quad \dot{h}_z(t) \leq 0.$$

Hence, h_z is a nonincreasing function, which proves item *i*) of Opial's Lemma 2.1.

Let us verify item *ii*) of Opial's Lemma 2.1. Let $w - \lim u(t_n) = z$ for some sequence $t_n \rightarrow +\infty$. Since $u(t_n) \in K$ and K is a closed convex subset of \mathcal{H} , we have $z \in K$. Moreover

$$(64) \quad \inf_{t \geq 0} f_i(u(t)) = \lim_{t \rightarrow +\infty} f_i(u(t))$$

$$(65) \quad = \lim_{n \rightarrow +\infty} f_i(u(t_n))$$

$$(66) \quad \geq f_i(z)$$

where the last inequality follows from the fact that f_i is convex continuous, and hence lower semicontinuous for the weak topology of \mathcal{H} . This being true for each $i = 1, \dots, q$ we conclude that $z \in S$. The two conditions of Opial's Lemma 2.1 are satisfied, which gives the weak convergence of each bounded trajectory of the (MOG) dynamic. Set

$$(67) \quad u(t) \rightharpoonup u_\infty \quad \text{weakly in } \mathcal{H}, \text{ as } t \rightarrow +\infty,$$

and show that u_∞ is a Pareto critical point, and hence a weak Pareto optimum (Lemma 1.3). The finite energy property (51)

$$\int_0^{+\infty} \|\dot{u}(t)\|^2 dt < +\infty$$

implies

$$(68) \quad \liminf_{t \rightarrow +\infty} \|\dot{u}(t)\| = 0.$$

Since relations (32) and (33) are satisfied for almost all $t > 0$, (68) implies the existence of a sequence $t_n \rightarrow +\infty$ such that

$$(69) \quad \dot{u}(t_n) \rightarrow 0 \text{ strongly in } \mathcal{H}$$

$$(70) \quad -\dot{u}(t_n) \in N_K(u(t_n)) + \text{Conv} \{ \partial f_i(u(t_n)) \} \text{ for each } n \in \mathbb{N}.$$

Moreover by (67)

$$(71) \quad u(t_n) \rightharpoonup u_\infty \quad \text{weakly in } \mathcal{H}.$$

We conclude using (69), (70), (71), and the following lemma which establishes a closure property for the operator $N_K(\cdot) + \text{Conv} \{ \partial f_i(\cdot) \}$. \square

Lemma 2.4. *Under assumptions H0), H1), the multi-application*

$$\begin{aligned} \mathcal{H} &\rightrightarrows \mathcal{H} \\ u &\longmapsto N_K(u) + \text{Conv} \{ \partial f_i(u) \} \end{aligned}$$

is demiclosed, i.e., its graph is sequentially closed for the weak – \mathcal{H} \times strong – \mathcal{H} topology.

Proof. Let (u_n, w_n) be a sequence in the graph of $N_K + \mathcal{C}$ where $\mathcal{C}(u) = \text{Conv} \{ \partial f_i(u) \}$. Suppose that u_n converges weakly to $\bar{u} \in K$, that w_n converges strongly to \bar{w} , and prove that $\bar{w} \in N_K(\bar{u}) + \mathcal{C}(\bar{u})$. For each $n \in \mathbb{N}$, there exists $q_n \in N_K(u_n)$, $p_{i,n} \in \partial f_i(u_n)$, $\lambda_{i,n} \in [0, 1]$, $i = 1, \dots, q$, such that

$$(72) \quad w_n = q_n + \sum_{i=1}^q \lambda_{i,n} p_{i,n} \quad \text{and} \quad \sum_{i=1}^q \lambda_{i,n} = 1.$$

For each n , $(\lambda_{i,n})_{i=1, \dots, q}$ belongs to the unit simplex in \mathbb{R}^q , which is a compact set. Hence we can extract a subsequence (still noted $(\lambda_{i,n})$ to simplify the notation) such that, for each $i = 1, \dots, q$

$$(73) \quad \lambda_{i,n} \rightarrow \bar{\lambda}_i,$$

with

$$(74) \quad 0 \leq \bar{\lambda}_i \leq 1, \quad \sum_{i=1}^q \bar{\lambda}_i = 1.$$

Noticing that the functions f_i are convex continuous, thanks to the Moreau-Rockafellar additivity rule for the subdifferential of a sum of convex functions, we can rewrite (72) as follows

$$(75) \quad w_n \in \partial \left(\delta_K + \sum_{i=1}^q \lambda_{i,n} f_i \right) (u_n),$$

where δ_K is the indicator function of K . Equivalently, for any $\xi \in \mathcal{H}$

$$(76) \quad \sum_{i=1}^q \lambda_{i,n} f_i(\xi) + \delta_K(\xi) \geq \sum_{i=1}^q \lambda_{i,n} f_i(u_n) + \delta_K(u_n) + \langle w_n, \xi - u_n \rangle.$$

Let us pass to the lower limit in (76). By using (73), the lower semicontinuity property of the f_i and δ_K for the weak topology of \mathcal{H} (K is closed convex and hence weakly closed), and the weak (resp. strong) convergence of u_n (resp. w_n), we obtain

$$(77) \quad \sum_{i=1}^q \bar{\lambda}_i f_i(\xi) + \delta_K(\xi) \geq \sum_{i=1}^q \bar{\lambda}_i f_i(\bar{u}) + \delta_K(\bar{u}) + \langle \bar{w}, \xi - \bar{u} \rangle.$$

In the above limit process, we use the fact that the functions f_i are finitely valued (otherwise we would face the delicate question concerning the product $0 \times \infty$). Using again the Moreau-Rockafellar additivity rule, we equivalently obtain

$$\bar{w} \in \sum_{i=1}^q \bar{\lambda}_i \partial f_i(\bar{u}) + N_K(\bar{u}),$$

which, with (74), expresses that (\bar{u}, \bar{w}) is in the graph of $N_K + \mathcal{C}$. □

Remark 2.5. Suppose that there exists an ideal solution \bar{z} to (CMO). Then, for any solution trajectory of (MOG), $\bar{z} \in S$, where S has been defined in (53). Following the proof of Theorem 2.2, the function $h_{\bar{z}}(\cdot) = \frac{1}{2} \|u(\cdot) - \bar{z}\|^2$ is nonincreasing. Thus, in that case, any trajectory of (MOG) is bounded. We recover the fact that, in the case of a single convex objective function, the trajectories of the steepest descent equation are bounded iff the solution set is not empty.

3. EXISTENCE OF STRONG GLOBAL SOLUTIONS

In this section, it is assumed that $\mathcal{H} = \mathbb{R}^d$ is a finite dimensional Euclidean space. This is because our proof of the existence of solutions to the (MOG) dynamic is based on the Peano theorem, and not on the Cauchy-Lipschitz. It is likely that the proof can be adapted to the case of infinite dimension by making ad hoc assumption on the data (as inf-compactness). This is an interesting topic for further studies, particularly involving applications to PDEs. Our approach is based on the regularization of the non-smooth functions f_i by the Moreau-Yosida approximation. This approximation brings us back to the situation studied in [6], which considers the case of differentiable functions.

3.1. Statement of the result.

Theorem 3.1. *Let \mathcal{H} be a finite dimensional Hilbert space. Let us make assumptions H0), H1), H2). Then, for any initial data $u_0 \in K$, there exists a strong global solution $u : [0, +\infty[\rightarrow \mathcal{H}$ of (MOG) system (2), which satisfies $u(0) = u_0$.*

Remark 3.2. In Theorem 3.1, for any $u_0 \in K$, we claim the existence of a strong global solution $u : [0, +\infty[\rightarrow \mathcal{H}$ of (MOG) system, satisfying the Cauchy data $u(0) = u_0$. By definition of a strong solution, u is absolutely continuous on any finite time interval $[0, T]$, but from Proposition 1.15 we know it is moreover Lipschitz continuous.

In the above theorem, we only claim existence. Without further assumptions, uniqueness is not guaranteed. Indeed, the following proof of existence relies on Peano, not Cauchy-Lipschitz theorem. Before entering the proof of existence, we will briefly discuss the question of uniqueness which remains an open question.

Remark 3.3. In the unconstrained case, and for convex differentiable objective functions, illustrative examples of the (MOG) dynamic were given in Section 1.4. In these elementary situations, we have been able to explicitly compute the vector field $v \mapsto s(v)$. We observed that it can be Lipschitz continuous (Example 1 and 2) or only Hölder continuous (Example 3). This naturally raises the following question: in the unconstrained case, and for differentiable objective functions, what are the assumptions ensuring that the vector field $v \mapsto s(v)$ is Lipschitz continuous (recall that it is Hölder continuous, see [6])? This is clearly a key property for uniqueness for (MOG).

The end of this section is devoted to the proof of Theorem 3.1, which is quite technical. To make reading easier, the proof has been divided into several stages. First of all, let us bring some additional results to [6], which concern the smooth case, and which will be useful for our study.

3.2. The smooth case, complements. Let us suppose that the f_i are convex differentiable functions. Following [6], for any $u_0 \in K$, there exists a strong global solution $u : [0, +\infty[\rightarrow \mathcal{H}$ of the Cauchy problem

$$(78) \quad \begin{cases} \dot{u}(t) + \left(N_K(u(t)) + \text{Conv} \{ \nabla f_i(u(t)) \} \right)^0 = 0, \\ u(0) = u_0. \end{cases}$$

The concept of solution u is as follows.

- (i) $u : [0, +\infty[\rightarrow \mathcal{H}$ is absolutely continuous on each interval $[0, T]$, $0 < T < +\infty$;

(ii) there exists $\eta : [0, +\infty[\rightarrow \mathcal{H}$ and $w : [0, +\infty[\rightarrow \mathcal{H}$ which satisfy

$$(79) \quad \eta \in L^2(0, T; \mathcal{H}), \quad w \in L^\infty(0, T; \mathcal{H}) \quad \text{for all } T > 0;$$

$$(80) \quad \eta(t) \in N_K(u(t)), \quad w(t) \in \text{Conv} \{ \nabla f_i(u(t)) \} \quad \text{for almost all } t > 0;$$

$$(81) \quad \eta(t) + w(t) = \left(N_K(u(t)) + \text{Conv} \{ \nabla f_i(u(t)) \} \right)^0 \quad \text{for almost all } t > 0;$$

$$(82) \quad \dot{u}(t) + \eta(t) + w(t) = 0 \quad \text{for almost all } t > 0.$$

Let us make precise (80).

Lemma 3.4. *Let u be a solution of (78), and η, w the associated functions satisfying (79)-(80)-(81)-(82). Then $w(t) \in \text{Conv} \{ \nabla f_i(u(t)) \}$ can be written as follows:*

$$(83) \quad w(t) = \sum_i \theta_i(t) \nabla f_i(u(t))$$

with $\theta_i \in L^\infty(0, +\infty)$, $i = 1, 2, \dots, q$, and for almost all $t > 0$, $(\theta_i(t)) \in \mathcal{S}^q$.

Proof. From (81) we see that for almost all $t > 0$

$$(84) \quad w(t) = \text{proj}_{\mathcal{C}(u(t))}(-\eta(t)),$$

where $\mathcal{C}(u(t)) = \text{Conv} \{ \nabla f_i(u(t)) \}$. Equivalently, $w(t) = \sum_i \theta_i(t) \nabla f_i(u(t))$ for any $\theta(t) = (\theta_i(t))$ such that

$$(85) \quad \theta(t) \in \text{argmin} \{ j(t, \theta) : \theta \in \mathbb{R}^q \}$$

where

$$(86) \quad j(t, \theta) = \|\eta(t) + \sum_i \theta_i \nabla f_i(u(t))\| + \delta_{\mathcal{S}^q}(\theta),$$

where $\delta_{\mathcal{S}^q}$ is the indicator function of \mathcal{S}^q . The crucial point is to prove that we can take the $\theta_i(t)$ measurable. Since $j : [0, +\infty[\times \mathbb{R}^q \rightarrow \mathbb{R} \cup \{+\infty\}$ is a positive (convex) normal integrand, the mapping $t \mapsto \text{argmin} j(t, \cdot)$ is measurable, and hence admits a measurable selection $t \mapsto \theta(t)$, see [37, Corollary 14.6; Theorem 14.37]. Hence, we can write $w(t) = \sum_i \theta_i(t) \nabla f_i(u(t))$, with θ_i measurable, and $\theta(t) \in \mathcal{S}^q$. Since θ_i is bounded, we have

$$w(t) = \sum_i \theta_i(t) \nabla f_i(u(t)) \quad \text{and } \theta_i \in L^\infty(0, +\infty), \quad \theta(t) \in \mathcal{S}^q \quad \text{a.e. } t > 0.$$

□

Let us now return to our setting involving non-smooth objective functions f_i .

3.3. Approximate equations. The main difficulty comes from the discontinuity of the vector field which governs the (MOG) dynamic (2). As a main ingredient of our approach, we use the Moreau-Yosida approximation of the convex functions f_i (equivalently the Yosida approximation of the maximal monotone operators ∂f_i), $i = 1, \dots, q$. This regularization method is widely used in nonsmooth convex analysis, see [1], [10], [12], [16], [42] for a detailed presentation. Its main properties are summarized in the following statement.

Proposition 3.5. *Let $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a closed convex proper function. The Moreau-Yosida approximation of index $\lambda > 0$ of Φ is the function $\Phi_\lambda : \mathcal{H} \rightarrow \mathbb{R}$ which is defined for all $v \in \mathcal{H}$ by*

$$(87) \quad \Phi_\lambda(v) = \inf \left\{ \Phi(\xi) + \frac{1}{2\lambda} \|v - \xi\|^2 : \xi \in \mathcal{H} \right\}.$$

(1) *The infimum in (87) is attained at a unique point $J_\lambda v \in \mathcal{H}$, which satisfies*

$$(88) \quad \Phi_\lambda(v) = \Phi(J_\lambda v) + \frac{1}{2\lambda} \|v - J_\lambda v\|^2;$$

$$(89) \quad J_\lambda v + \lambda \partial \Phi(J_\lambda v) \ni v.$$

$J_\lambda = (I + \lambda \partial \Phi)^{-1} : \mathcal{H} \rightarrow \mathcal{H}$ is everywhere defined and nonexpansive. It is called the resolvent of index λ of $A = \partial \Phi$.

(2) Φ_λ is convex, and continuously differentiable. Its gradient at $v \in \mathcal{H}$ is equal to

$$(90) \quad \nabla \Phi_\lambda(v) = \frac{1}{\lambda} (v - J_\lambda v).$$

(3) *The operator $A_\lambda = \nabla \Phi_\lambda = \frac{1}{\lambda} (I - J_\lambda)$ is called the Yosida approximation of index λ of the maximal monotone operator $A = \partial \Phi$. It is Lipschitz continuous with Lipschitz constant $\frac{1}{\lambda}$.*

(4) *For any $v \in \text{dom} A$, $\|A_\lambda v\| \leq \|A^0(v)\|$, ($A^0(v)$ is the element of minimal norm of $A(v)$).*

(5) *For any $v \in \mathcal{H}$, $\Phi_\lambda(v) \uparrow \Phi(v)$ as $\lambda \downarrow 0$.*

We are going to adapt to our situation the classical proof of the existence of strong solutions to evolution equations governed by subdifferentials of convex lower semicontinuous functions, see [16]. For each $\lambda > 0$, we set $f_{i,\lambda} = (f_i)_\lambda$ the Moreau-Yosida approximation of index λ of f_i . We consider the Cauchy problem which is obtained by replacing each ∂f_i by its Yosida approximation $\nabla f_{i,\lambda}$, in (MOG). So doing, we are in the situation studied in [6], which treats the case of differentiable objective functions. Precisely, by [6, Theorem 3.5], for each $\lambda > 0$ there exists of a strong global solution $u_\lambda : [0, +\infty[\rightarrow \mathcal{H}$ of the Cauchy problem

$$(91) \quad (\text{MOG})_\lambda \begin{cases} \dot{u}_\lambda(t) + \left(N_K(u_\lambda(t)) + \text{Conv} \{ \nabla f_{i,\lambda}(u_\lambda(t)) \} \right)^0 = 0, \\ u_\lambda(0) = u_0. \end{cases}$$

By Lemma 3.4 and (82), there exists $\theta_{i,\lambda} \in L^\infty(0, +\infty)$, and $\eta_\lambda \in L^2(0, T; \mathcal{H})$ for all $T > 0$, such that, for almost all $t > 0$

$$(92) \quad \dot{u}_\lambda(t) + \sum_{i=1}^q \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)) + \eta_\lambda(t) = 0,$$

and

$$(93) \quad \eta_\lambda(t) \in N_K(u_\lambda(t)), \quad (\theta_{i,\lambda}(t)) \in \mathcal{S}^q.$$

3.4. Estimations on the sequence (u_λ) . Let us establish bounds for the net $(u_\lambda)_\lambda$, which are independent of λ . Let us make a similar argument to that used in Theorem 2.2, just replacing f_i by $f_{i,\lambda}$. We obtain

$$(94) \quad \int_0^{+\infty} \|\dot{u}_\lambda(t)\|^2 dt \leq f_{i,\lambda}(u_0) - \inf_{\mathcal{H}} f_{i,\lambda}.$$

Then notice that $f_{i,\lambda}(u_0) \leq f_i(u_0)$, and $\inf_{\mathcal{H}} f_{i,\lambda} = \inf_{\mathcal{H}} f_i$. Hence

$$(95) \quad \int_0^{+\infty} \|\dot{u}_\lambda(t)\|^2 dt \leq f_i(u_0) - \inf_{\mathcal{H}} f_i,$$

and

$$(96) \quad \sup_{\lambda} \int_0^{+\infty} \|\dot{u}_\lambda(t)\|^2 dt < +\infty.$$

From

$$(97) \quad u_\lambda(t) = u_0 + \int_0^t \dot{u}_\lambda(\tau) d\tau,$$

and Cauchy-Schwarz inequality, we obtain

$$(98) \quad \|u_\lambda(t)\| \leq \|u_0\| + \sqrt{t} \left(\int_0^t \|\dot{u}_\lambda(\tau)\|^2 d\tau \right)^{\frac{1}{2}}.$$

Combining (96) with (98) we deduce that, for any $T > 0$

$$(99) \quad \sup_\lambda \|u_\lambda\|_{L^\infty([0,T];\mathcal{H})} < +\infty.$$

Let us now consider the gradients terms $\nabla f_{i,\lambda}(u_\lambda)$ which appear in (92). By Proposition 3.5, item 4., for any $v \in \mathcal{H}$, $\lambda > 0$, and $i = 1, 2, \dots, q$

$$(100) \quad \|\nabla f_{i,\lambda}(v)\| \leq \|(\partial f_i)^0(v)\|.$$

Combining (99) with (100), and using assumption H1), which tells us that f_i is a convex continuous function whose subdifferential ∂f_i is bounded on bounded sets, we obtain that, for any $T > 0$

$$(101) \quad \sup_\lambda \|\nabla f_{i,\lambda}(u_\lambda)\|_{L^\infty([0,T];\mathcal{H})} < +\infty.$$

3.5. Passing to the limit ($\lambda \rightarrow 0$). As we have already pointed out, the difficulty comes from the discontinuous nature of the multivalued operators ∂f_i and N_K , and hence of the vector field which governs the differential equation (2). Indeed, we are going to use the monotonicity property of these operators, and the demiclosedness property (closedness for the strong \times weak product topology) of their graphs in the associated functional spaces.

By (96), (99), the generalized sequence (u_λ) is uniformly bounded and equi-continuous on $[0, T]$. Since \mathcal{H} is finite dimensional, we deduce from Ascoli's theorem that, for any $0 < T < +\infty$, the generalized sequence (u_λ) is relatively compact for the uniform convergence topology on $[0, T]$. Thus, by a diagonal argument (we keep the notation (u_λ) for simplicity), we obtain the existence of $u \in \mathcal{C}([0, +\infty[; \mathcal{H})$, and $v_i, \eta \in L^2_{loc}(0, +\infty; \mathcal{H})$, $\theta_i \in L^\infty(0, +\infty)$ such that, for any $0 < T < +\infty$,

$$(102) \quad u_\lambda \rightarrow u \quad \text{strong} - \mathcal{C}(0, T; \mathcal{H})$$

$$(103) \quad \dot{u}_\lambda \rightharpoonup \dot{u} \quad \text{weak} - L^2(0, T; \mathcal{H})$$

$$(104) \quad \nabla f_{i,\lambda}(u_\lambda) \rightharpoonup v_i \quad \sigma(L^\infty(0, T; \mathcal{H}), L^1(0, T; \mathcal{H}))$$

$$(105) \quad \theta_{i,\lambda} \rightharpoonup \theta_i \quad \sigma(L^\infty(0, T), L^1(0, T))$$

$$(106) \quad \eta_\lambda \rightharpoonup \eta \quad \text{weak} - L^2(0, T; \mathcal{H}).$$

The last statement comes from the following observation: by (92)

$$(107) \quad \eta_\lambda(t) = -\dot{u}_\lambda(t) - \sum_{i=1}^q \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)),$$

which implies that the net (η_λ) remains bounded in $L^2(0, T; \mathcal{H})$ for any $T > 0$. Let us complete this list with the convergence of the net $(f_{i,\lambda}(u_\lambda))$.

Lemma 3.6. *The following convergence result holds: for any $0 < T < +\infty$*

$$(108) \quad f_{i,\lambda}(u_\lambda) \rightarrow f_i(u) \text{ uniformly on } [0, T] \text{ as } \lambda \rightarrow 0.$$

Proof. Let us fix $T > 0$, and work on the bounded interval $[0, T]$. Let us write the triangle inequality

$$(109) \quad |f_{i,\lambda}(u_\lambda) - f_i(u)| \leq |f_{i,\lambda}(u_\lambda) - f_{i,\lambda}(u)| + |f_{i,\lambda}(u) - f_i(u)|.$$

On the one hand, by the Mean value theorem, (100), and (102)

$$(110) \quad |f_{i,\lambda}(u_\lambda(t)) - f_{i,\lambda}(u(t))| \leq \left(\sup_{\xi \in [u_\lambda(t), u(t)]} \|\nabla f_{i,\lambda}(\xi)\| \right) \|u_\lambda(t) - u(t)\|$$

$$(111) \quad \leq \left(\sup_{\xi \in [u_\lambda(t), u(t)]} \|(\partial f_i)^0(\xi)\| \right) \|u_\lambda(t) - u(t)\|$$

$$(112) \quad \leq C \|u_\lambda(t) - u(t)\|,$$

and hence,

$$(113) \quad f_{i,\lambda}(u_\lambda) - f_{i,\lambda}(u) \rightarrow 0 \quad \text{uniformly on } [0, T] \text{ as } \lambda \rightarrow 0.$$

On the other hand, the net $(f_{i,\lambda}(u))_\lambda$ is equi-continuous. This results from the following inequalities

$$(114) \quad \left| \frac{d}{dt} f_{i,\lambda}(u(t)) \right| = | \langle \nabla f_{i,\lambda}(u(t)), \dot{u}(t) \rangle |$$

$$(115) \quad \leq \|(\partial f_i)^0(u(t))\| \|\dot{u}(t)\|$$

$$(116) \quad \leq C \|\dot{u}(t)\|,$$

and

$$(117) \quad |f_{i,\lambda}(u(t)) - f_{i,\lambda}(u(s))| \leq \int_s^t \left| \frac{d}{d\tau} f_{i,\lambda}(u(\tau)) \right| d\tau$$

$$(118) \quad \leq \sqrt{t-s} \left(\int_0^T \left| \frac{d}{d\tau} f_{i,\lambda}(u(\tau)) \right|^2 d\tau \right)^{\frac{1}{2}}$$

$$(119) \quad \leq C \sqrt{t-s} \left(\int_0^T \|\dot{u}(\tau)\|^2 d\tau \right)^{\frac{1}{2}}.$$

Hence, the net $(f_{i,\lambda}(u))_\lambda$ is equi-continuous. Since it converges pointwise to $f_i(u)$, by Ascoli Theorem, we obtain

$$(120) \quad f_{i,\lambda}(u) - f_i(u) \rightarrow 0 \quad \text{uniformly on } [0, T] \text{ as } \lambda \rightarrow 0.$$

Combining (109), (113), (120), we obtain (108). \square

Technically, the most difficult point is to pass to the limit in (92) on the product of the two weakly converging sequences $(\theta_{i,\lambda})$ and $(\nabla f_{i,\lambda}(u_\lambda))$. In order to circumvent this difficulty, we use a variational argument based on the convex differential inequality: for any $\xi \in L^\infty(0, T; \mathcal{H})$,

$$(121) \quad \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(\xi(t)) \geq \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) + \left\langle \sum_{i=1}^q \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)), \xi(t) - u_\lambda(t) \right\rangle.$$

After integration on $[0, T]$, we obtain

$$(122) \quad \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(\xi(t)) dt \geq \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) dt$$

$$(123) \quad + \int_0^T \left\langle \sum_{i=1}^q \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)), \xi(t) - u_\lambda(t) \right\rangle dt.$$

By (92), $\sum_{i=1}^q \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)) = -\dot{u}_\lambda(t) - \eta_\lambda(t)$. Replacing in (122)-(123), we obtain

$$(124) \quad \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(\xi(t)) dt \geq \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) dt$$

$$(125) \quad + \int_0^T \langle -\dot{u}_\lambda(t) - \eta_\lambda(t), \xi(t) - u_\lambda(t) \rangle dt.$$

Since $f_{i,\lambda}(\xi(t)) \leq f_i(\xi(t))$, and $\theta_{i,\lambda}(t) \geq 0$, we obtain

$$(126) \quad \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_i(\xi(t)) dt \geq \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) dt$$

$$(127) \quad + \int_0^T \langle -\dot{u}_\lambda(t) - \eta_\lambda(t), \xi(t) - u_\lambda(t) \rangle dt.$$

For any $\xi \in L^\infty(0, T; \mathcal{H})$, since f_i is continuous and bounded on bounded sets (assumption H1), we have $f_i(\xi(\cdot)) \in L^\infty(0, T)$. Moreover $\theta_{i,\lambda} \rightarrow \theta_i$ for the topology $\sigma(L^\infty(0, T), L^1(0, T))$. Therefore, by passing to the limit on the left member of (126), we obtain

$$\lim_{\lambda} \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_i(\xi(t)) dt = \int_0^T \sum_{i=1}^q \theta_i(t) f_i(\xi(t)) dt.$$

Let us now pass to the limit on the right member of (126)-(127). For the first term, we use Lemma 3.6. For the second term, we notice that this expression involves duality products of nets which are respectively converging for the strong and weak topologies of a duality pairing. More precisely $\dot{u}_\lambda + \eta_\lambda$ converges weakly in $L^2(0, T; \mathcal{H})$ to $\dot{u} + \eta$, and $\xi - u_\lambda$ converges uniformly, and hence strongly in $L^2(0, T; \mathcal{H})$ to $\xi - u$. Hence, by passing to the limit as λ goes to zero, we obtain

$$(128) \quad \int_0^T \sum_{i=1}^q \theta_i(t) f_i(\xi(t)) dt \geq \int_0^T \sum_{i=1}^q \theta_i(t) f_i(u(t)) dt$$

$$+ \int_0^T \langle -\dot{u}(t) - \eta(t), \xi(t) - u(t) \rangle dt.$$

Let us interpret this inequality in the duality pairing between the functional spaces $L^\infty(0, T; \mathcal{H})$ and $L^1(0, T; \mathcal{H}) \subset (L^\infty(0, T; \mathcal{H}))^*$. For this, introduce I , the integral functional on $L^\infty(0, T; \mathcal{H})$ which is defined by

$$(129) \quad I(\xi) = \int_0^T \sum_{i=1}^q \theta_i(t) f_i(\xi(t)) dt.$$

We observe that $I : L^\infty(0, T; \mathcal{H}) \rightarrow \mathbb{R}$ is convex and continuous on $L^\infty(0, T; \mathcal{H})$. Hence, inequality (128) can be rewritten as

$$(130) \quad -\dot{u} - \eta \in \partial I(u).$$

According to the duality theorem of Rockafellar for convex functional integrals, see [36, Theorem 4], for almost all $t > 0$

$$(131) \quad -\dot{u}(t) - \eta(t) \in \partial \left(\sum_{i=1}^q \theta_i(t) f_i \right) (u(t))$$

$$(132) \quad = \sum_{i=1}^q \partial (\theta_i(t) f_i) (u(t)),$$

where the last equality comes from the additivity rule for the subdifferential of the sum of convex continuous functions on \mathbb{R}^q . Indeed we need to prove a slighter more precise result:

$$-\dot{u}(t) - \eta(t) = \sum_{i=1}^q \theta_i(t) v_i(t)$$

with *measurable* functions $v_i \in L^\infty(0, T; \mathcal{H})$ such that

$$(133) \quad v_i(t) \in \partial f_i(u(t)) \quad \text{for almost all } t > 0.$$

This can be proved by a precise analysis of the duality theorem from [36]. Since it is quite technical, the proof is stated in Lemma 3.7, at the end of this section. Assuming this result, we obtain by combination with (130) that

$$(134) \quad \dot{u}(t) + \eta(t) + \sum_i \theta_i(t) v_i(t) = 0 \quad \text{for almost all } t > 0,$$

with

$$(135) \quad \theta_i \in L^\infty(0, +\infty; \mathbb{R}), \quad v_i \in L^\infty(0, T; \mathcal{H}), \quad \eta \in L^2(0, T; \mathcal{H}), \quad \text{for all } T > 0, \quad \text{and all } i = 1, 2, \dots, q;$$

$$(136) \quad (\theta_i(t)) \in \mathcal{S}^q \quad \text{and } v_i(t) \in \partial f_i(u(t)) \quad \text{for almost all } t > 0;$$

On the other hand, from $u_\lambda \rightarrow u$ strong- $\mathcal{C}(0, T; \mathcal{H})$, $\eta_\lambda \rightharpoonup \eta$ weak- $L^2(0, T; \mathcal{H})$, $\eta_\lambda(t) \in N_K(u_\lambda(t))$, and from the demi-closedness property of the extension to $L^2(0, T; \mathcal{H})$ of the maximal monotone normal cone mapping (N_K is the subdifferential of the indicator function fo K), we obtain

$$(137) \quad \eta(t) \in N_K(u(t)).$$

Thus

$$(138) \quad \dot{u}(t) + N_K(u(t)) + \text{Conv} \{ \partial f_i(u(t)) \} \ni 0.$$

3.6. Lazy solution. Let us complete the proof of Theorem 3.1 by showing that u is a lazy solution of the differential inclusion (138). Let us start from the lazy solution property satisfied by the approximate solutions u_λ

$$(139) \quad -\dot{u}_\lambda(t) = \left(N_K(u_\lambda(t)) + \text{Conv} \{ \nabla f_{i,\lambda}(u_\lambda(t)) \} \right)^0.$$

By the obtuse angle property, since $0 \in N_K(u_\lambda(t))$ we have

$$(140) \quad \left\langle \dot{u}_\lambda(t), \dot{u}_\lambda(t) + \sum_{i=1}^q \theta_i(t) \nabla f_{i,\lambda}(u_\lambda(t)) \right\rangle \leq 0,$$

for all $\theta_i \in L^\infty(0, +\infty)$, $i = 1, 2, \dots, q$ that satisfy $(\theta_i(t)) \in \mathcal{S}^q$. After developing, and using the classical derivation chain rule, we obtain

$$(141) \quad \|\dot{u}_\lambda(t)\|^2 + \sum_{i=1}^q \theta_i(t) \frac{d}{dt} f_{i,\lambda}(u_\lambda(t)) \leq 0.$$

In order to pass to the limit on (141), take α a nonnegative test function (a function of t which is regular, and with compact support in $]0, T[$). After multiplication of (141) by α , and integration on $[0, T]$, we obtain

$$(142) \quad \int_0^T \alpha(t) \|\dot{u}_\lambda(t)\|^2 dt + \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_{i,\lambda}(u_\lambda(t)) dt \leq 0.$$

The convex function $v \mapsto \int_0^T \alpha(t) \|v(t)\|^2 dt$ is continuous on $L^2(0, T; \mathcal{H})$, and hence lower semicontinuous for the weak topology of $L^2(0, T; \mathcal{H})$. Since $\dot{u}_\lambda \rightharpoonup \dot{u}$ weakly in $L^2(0, T; \mathcal{H})$, we have

$$(143) \quad \int_0^T \alpha(t) \|\dot{u}(t)\|^2 dt \leq \liminf \int_0^T \alpha(t) \|\dot{u}_\lambda(t)\|^2 dt.$$

In order to pass to the limit on the second term of (142), we use a density argument. First assume that the θ_i are Lipschitz continuous on bounded sets. Since $f_{i,\lambda}(u_\lambda)$ and $\alpha\theta_i$ are absolutely continuous functions of a real variable, their product is still absolutely continuous (see [17, Corollary VIII.9]), and integration by part formula is valid. Hence

$$(144) \quad \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_{i,\lambda}(u_\lambda(t)) dt = - \sum_{i=1}^q \int_0^T \frac{d}{dt} (\alpha\theta_i)(t) f_{i,\lambda}(u_\lambda(t)) dt.$$

By Lemma 3.6,

$$f_{i,\lambda}(u_\lambda) \rightarrow f_i(u) \text{ uniformly on } [0, T], \text{ as } \lambda \rightarrow 0.$$

Moreover $\frac{d}{dt}(\alpha\theta_i) \in L^\infty(0, T; \mathbb{R})$. Thus, as $\lambda \rightarrow 0$

$$(145) \quad \sum_{i=1}^q \int_0^T \frac{d}{dt} (\alpha\theta_i)(t) f_{i,\lambda}(u_\lambda(t)) dt \rightarrow \sum_{i=1}^q \int_0^T \frac{d}{dt} (\alpha\theta_i)(t) f_i(u(t)) dt.$$

Since $f_i(u)$ is absolutely continuous, using again integration by part formula

$$(146) \quad - \sum_{i=1}^q \int_0^T \frac{d}{dt} (\alpha\theta_i)(t) f_i(u(t)) dt = \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_i(u(t)) dt.$$

From (144), (145), and (146) we obtain

$$(147) \quad \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_{i,\lambda}(u_\lambda(t)) dt \rightarrow \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_i(u(t)) dt.$$

Combining (142), (143), and (147) we obtain, for θ_i that satisfy $(\theta_i(t)) \in \mathcal{S}^q$, and are Lipschitz continuous on bounded sets,

$$(148) \quad \int_0^T \alpha(t) \|\dot{u}(t)\|^2 dt + \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_i(u(t)) dt \leq 0.$$

Let us show that, by density, (148) can be extended to arbitrary $\theta_i \in L^\infty(0, +\infty; \mathbb{R})$ that satisfy $(\theta_i(t)) \in \mathcal{S}^q$. Given such functions $(\theta_i)_{i=1,\dots,q}$, by regularization by convolution, we can find a sequence of regular functions $\theta_{i,n} \in \mathcal{C}^\infty(0, +\infty)$, such that

$$(149) \quad \theta_{i,n} \rightarrow \theta_i \text{ a.e. } t \in (0, +\infty) \text{ when } n \text{ goes to } +\infty.$$

Let $T : \mathbb{R}^q \rightarrow \mathbb{R}^q$ be the projection onto the unit simplex $\mathcal{S}^q \subset \mathbb{R}^q$. T is a nonexpansive mapping. By (149), and $\theta(t) = (\theta_i(t))_i \in \mathcal{S}^q$ for almost all $t > 0$, we see that $T \circ \theta_n$ is Lipschitz continuous on any interval $[0, T]$, and satisfies, for almost all $t > 0$

$$(150) \quad T \circ \theta_n(t) \in \mathcal{S}^q,$$

$$(151) \quad (T \circ \theta_n)_i(t) \rightarrow \theta_i(t) \text{ for almost all } t > 0.$$

By (148), for each $n \in \mathbb{N}$, we have

$$(152) \quad \int_0^T \alpha(t) \|\dot{u}(t)\|^2 dt + \sum_{i=1}^q \int_0^T \alpha(t) (T \circ \theta_n)_i(t) \frac{d}{dt} f_i(u(t)) dt \leq 0.$$

On the other hand, by Lemma 1.16, for any $\xi \in L^2(0, T; H)$ such that $\xi(t) \in \partial f_i(u(t))$ (there exists such elements, for example take v_i obtained in (136)), we have

$$(153) \quad \frac{d}{dt} f_i(u(t)) = \langle \dot{u}(t), \xi(t) \rangle,$$

and $\frac{d}{dt} f_i(u)$ is integrable on $[0, T]$ ($t \mapsto f_i(u(t))$ is absolutely continuous on $[0, T]$). From (151), by applying Fatou's lemma, (note that $\frac{d}{dt} f_i(u) \in L^1([0, T])$, which allows us to reduce to the case of non-negative functions), we obtain

$$(154) \quad \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_i(u(t)) dt \leq \liminf_n \sum_{i=1}^q \int_0^T \alpha(t) (T \circ \theta_n)_i(t) \frac{d}{dt} f_i(u(t)) dt.$$

From (152) and (154) we deduce that

$$(155) \quad \int_0^T \alpha(t) \|\dot{u}(t)\|^2 dt + \sum_{i=1}^q \int_0^T \alpha(t) \theta_i(t) \frac{d}{dt} f_i(u(t)) dt \leq 0.$$

Since α is an arbitrary positive test function, we deduce from (155) that

$$(156) \quad \|\dot{u}(t)\|^2 + \sum_{i=1}^q \theta_i(t) \frac{d}{dt} f_i(u(t)) dt \leq 0.$$

Take arbitrary $\eta \in L^2(0, T; \mathcal{H})$, $\xi_i \in L^2(0, T; \mathcal{H})$ $i = 1, \dots, q$, such that $\eta(t) \in N_K(u(t))$, $\xi_i(t) \in \partial f_i(u(t))$ for almost all $t > 0$. Since $u(t) \in K$, we have $\dot{u}(t) \in T_K(u(t))$, and since $\eta(t) \in N_K(u(t))$

$$(157) \quad \langle \dot{u}(t), \eta(t) \rangle \leq 0.$$

Combining (153), (156), and (157) we obtain

$$(158) \quad \|\dot{u}(t)\|^2 + \langle \dot{u}(t), \eta(t) \rangle + \sum_{i=1}^q \theta_i(t) \langle \dot{u}(t), \xi_i(t) \rangle \leq 0.$$

Equivalently, for any $z(t) \in N_K(u(t)) + \text{Conv} \{ \partial f_i(u(t)) \}$

$$(159) \quad \langle 0 - (-\dot{u}(t)), z(t) - \dot{u}(t) \rangle \leq 0.$$

Combining this property with (138) we obtain

$$\dot{u}(t) + \left(N_K(u(t)) + \text{Conv} \{ \partial f_i(u(t)) \} \right)^0 = 0 \quad \text{for almost all } t > 0,$$

which ends the proof.

Lemma 3.7. *Let $I : L^\infty(0, T; \mathcal{H}) \rightarrow \mathbb{R}$ be defined by $I(\xi) = \int_0^T \sum_{i=1}^q \theta_i(t) f_i(\xi(t)) dt$, with $\theta_i \in L^\infty(0, T; \mathcal{H})$ for $i = 1, \dots, q$, and $(\theta_i(t)) \in \mathcal{S}^q$ for almost all $t > 0$. Let $\xi \in L^\infty(0, T; \mathcal{H})$ and $z \in L^1(0, T; \mathcal{H})$ such that $z \in \partial I(\xi)$. Then for all $i = 1, \dots, q$ there exists $v_i \in L^\infty(0, T; \mathcal{H})$ such that*

$$\text{for almost all } t > 0, v_i(t) \in \partial f_i(\xi(t)), \text{ and } z(t) = \sum_{i=1}^q \theta_i(t) v_i(t).$$

Proof. By the Fenchel extremality relation,

$$(160) \quad z \in \partial I(\xi) \Leftrightarrow I(\xi) + I^*(z) - \langle \xi, z \rangle_{(L^\infty(0, T; \mathcal{H}), L^1(0, T; \mathcal{H}))} = 0.$$

By [36, Theorem 2], we have

$$I^*(z) = \int_0^T \left(\sum_{i=1}^q \theta_i(t) f_i \right)^* (z(t)) dt.$$

Let us analyze this last expression. Since the f_i are convex continuous functions, their conjugate are coercive functions, and

$$\left(\sum_{i=1}^q \theta_i(t) f_i \right)^* (z(t)) = \min \left(\sum_{i=1}^q (\theta_i(t) f_i)^* (z_i) : \sum_i z_i = z(t) \right).$$

The same measurable selection argument as the one used in Lemma 3.4 gives the existence of measurable functions $z_i(\cdot)$ such that

$$(161) \quad \left(\sum_{i=1}^q \theta_i(t) f_i \right)^* (z(t)) = \sum_{i=1}^q (\theta_i(t) f_i)^* (z_i(t)) \quad \text{with} \quad \sum_i z_i(t) = z(t).$$

Returning to (160) we obtain

$$(162) \quad z \in \partial I(\xi) \Leftrightarrow \int_0^T \sum_{i=1}^q (\theta_i(t) f_i(\xi(t)) + (\theta_i(t) f_i)^* (z_i(t))) dt - \int_0^T \langle \xi(t), z(t) \rangle dt = 0$$

$$\Leftrightarrow \int_0^T \sum_{i=1}^q (\theta_i(t) f_i(\xi(t)) + (\theta_i(t) f_i)^* (z_i(t)) - \langle \xi(t), z_i(t) \rangle) dt = 0.$$

Since each of the elements of this last sum expression is nonnegative, we deduce that, for each $i = 1, 2, \dots, q$, and for almost all $t > 0$

$$\theta_i(t) f_i(\xi(t)) + (\theta_i(t) f_i)^* (z_i(t)) - \langle \xi(t), z_i(t) \rangle = 0.$$

Equivalently $z_i(t) \in \partial(\theta_i(t) f_i)(\xi(t))$. Let us now verify that

$$\partial(\theta_i(t) f_i)(u(t)) = \theta_i(t) \partial f_i(u(t)).$$

Take some $\tilde{z}_i \in L^\infty(0, T; \mathcal{H})$ such that $\tilde{z}_i(t) \in \partial f_i(\xi(t))$ for almost all $t > 0$, (there exists such element, take for example $\tilde{z}_i(t) = (\partial f_i)^0(\xi(t))$). We have

$$(163) \quad z_i(t) = \theta_i(t) v_i(t) \quad \text{for almost all } t > 0,$$

where

$$(164) \quad v_i(t) = \begin{cases} \frac{z_i(t)}{\theta_i(t)} & \text{if } \theta_i(t) > 0, \\ \tilde{z}_i(t) & \text{if } \theta_i(t) = 0. \end{cases}$$

Moreover v_i is measurable, and $v_i(t) \in \partial f_i(\xi(t))$ for almost all $t > 0$. By continuity of f_i , we conclude that $v_i \in L^\infty(0, T; \mathcal{H})$. \square

4. SOME MODELING AND NUMERICAL ASPECTS, PERSPECTIVES

4.1. Cooperative games. In this section, we consider some modeling aspects concerning the multiobjective steepest descent for cooperative games. This completes [6], where was considered the smooth case. Indeed, for applications, it is quite useful to consider objective functions which are not differentiable (like the $\|\cdot\|_1$ norm for sparse optimization).

Let us consider q agents (consumers, social actors, deciders,...). The agent i acts on a decision space \mathcal{H}_i , and takes decision $v_i \in \mathcal{H}_i$, $i = 1, 2, \dots, q$. Let K be a given closed subset of $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \dots \times \mathcal{H}_q$, which reflects the limitation of ressources, and/or various constraints. Feasible decisions $v \in \mathcal{H}$ satisfy

$$v = (v_1, v_2, \dots, v_q) \in K.$$

Each agent i has a disutility (loss) function $f_i : \mathcal{H} \rightarrow \mathbb{R}$ which associates to each feasible decision $v \in K$ the scalar $f_i(v)$. The game in normal form is given by the triplet $(\mathcal{H}, K, (f_i)_{i=1, \dots, q})$. The (MOG) dynamic has been designed in order to satisfy some desirable properties with respect to Pareto equilibration: each trajectory $t \mapsto u(t)$ of (MOG) satisfies

- i) for each $i = 1, 2, \dots, q$, $t \mapsto f_i(u(t))$ is nonincreasing (Theorem 2.2, item i));

ii) $u(t)$ converges to a Pareto critical point as $t \rightarrow +\infty$ (Theorem 2.2, item iii));

Let us make some further observations:

iii) In (MOG) dynamic there is no a priori or a posteriori scalarization of the original vector optimization problem. Neither ordering information nor weighting factors for the different objective functions are assumed to be known. The scalarization is done dynamically, endogenously ((MOG) is an autonomous dynamical system). Taking into account the worst directional derivative (indeed, in view of minimization, it is the greatest), can make progress all agents, and gives to (MOG) system robustness (minimization in the worst case), and good convergence properties. When it is no longer possible to make progress all the agents, the process stops at a weak Pareto optimal point. It is a natural question whether it is possible to reach a Pareto optimum. Indeed, it depends on the willingness of the agents to cooperate more or less. After reaching a weak Pareto optimum, a natural way is to consider the coalition involving agents that can further enhance their performance. Then we can consider the (MOG) dynamics involving these agents. An additional constraint must be added which states that the performance of the agents who stay at rest is not damaged.

iv) The choice of the metric on the space \mathcal{H} plays a fundamental role in the definition of the gradient-like system (MOG). The metric reflects the friction and inertia that are attached to the changes in dynamical decision processes, see [2], [7], for an account on the notion of costs to change (changing a routine...). The definition of (MOG) involves local notions (subdifferentials of the f_i , and tangent cone to K) which corresponds to the modeling of myopic agents.

v) A central question in Pareto optimization is obtaining a Pareto optimum with desirable properties. A major advantage of the dynamic gradient approach is that we don't need to know the whole Pareto front. The weak Pareto equilibrium finally reached is not too far from the starting point of the dynamics (see Figures 1 and 2), making the process realistic in engineering and human sciences. Moreover, one can select a Pareto optimum which is not too far from a desirable state u_d by using an auxiliary asymptotic hierarchical procedure (see [4] and references therein). For example, according to the method of Tikhonov regularization, we can consider $\epsilon(t) \rightarrow 0$ as $t \rightarrow +\infty$, with $\int_0^\infty \epsilon(t)dt = +\infty$, and the following dynamics

$$\dot{u}(t) + \left(N_K(u(t)) + \text{Conv} \{ \partial f_i(u(t)) + \epsilon(t)(u(t) - u_d) \} \right)^0 = 0.$$

It is a (time)-multiscaled nonautonomous dynamic, an interesting subject for further research.

vi) Hybrid methods combine gradient methods (fast, with low computational cost, but local) with evolutionary computation methods (global, but with high computational cost). They have proved to be efficient for the minimization of a single objective function. It would be interesting to develop the same type of idea in order to reach the whole Pareto set, see [15], [18] for some first results in this direction.

4.2. Inverse problems. As a model situation, let us consider the computation of sparse solutions for underdetermined systems of equations. It is an important problem in signal compression and statistics (see [24, 40]). It leads to the following nonsmooth convex minimization problem

$$\min \{ \|Ax - b\|_2^2 + \alpha \|x\|_1 : x \in K \subset \mathbb{R}^n \}$$

where $\|Ax - b\|_2^2$ is a least squares data fitting term, and $\|x\|_1$ forces sparsity. There is numerical evidence that a careful weighting of these two terms is important for the effectiveness of the method. Usually it is done by experimental trials. It would be of great interest to develop a numerical method based on a multiobjective optimization approach (with $f_1(x) = \|Ax - b\|_2^2$ and $f_2(x) = \|x\|_1$), where the weighting is done automatically, while giving more weight to the lower term. Indeed, this is what the (MOG) dynamic does.

All these considerations naturally lead us to consider discretized, algorithmic versions of the method.

4.3. Numerical descent methods for nonsmooth multiobjective optimization. In the unconstrained case, an explicit discretization of (MOG) provides an algorithm of the form:

$$(165) \quad \text{At step } k, \text{ compute } u_{k+1} = u_k + \lambda_k d_k,$$

where $d_k = s(u_k)$ is the multiobjective steepest descent direction at u_k , and λ_k is some nonnegative steplength. If we have a constraint K , we can approach $s(u_k)$ by replacing the tangent cone $T_K(u_k)$ with $\frac{C-u_k}{\mu_k}$ (for some small μ_k) in (14). This leads to :

$$(166) \quad u_{k+1} = u_k + \lambda_k d_k,$$

$$(167) \quad \text{where } d_k = \operatorname{argmin}_{d \in C-u_k} \left\{ \frac{1}{2\mu_k} \|d\|^2 + \max_{i=1, \dots, q} \max_{p_i \in \partial f_i(u)} \langle p_i, d \rangle \right\}.$$

Note that the algorithms given in (165) and (166) are equivalent when $K = \mathcal{H}$ and $\mu_k \equiv \mu$. These algorithms have been studied in [25], [27] (unconstrained case), in [29] (constrained case) in a finite-dimensional setting, and assuming that the objective functions are C^1 (not necessarily convex). As a distinctive feature of these algorithms, the steplength λ_k is computed by an Armijo-like rule (to secure a descent property), and directions d_k are computed approximatively (with a given tolerance). They lead to the following results:

- (1) If $\mu_k \equiv \mu$, then any accumulation point is a critical Pareto point.
- (2) If the objective functions are convex, and if $\mu_k = \frac{\alpha_k}{\max_{i=1, \dots, q} \|\nabla f_i(u_k)\|}$ with $\alpha_k \in \ell^2 \setminus \ell^1$, then any

bounded sequence converges to a weak Pareto optimal point.

It appears that these algorithms, which are obtained -at least formally- by the explicit discretization in time of (MOG), share common properties with our continuous dynamic (descent property, convergence to weak Pareto optimal points). It would be interesting to justify mathematically that the continuous and discrete dynamic systems have the same asymptotic behavior, as it was established in the case of a single objective (see [34]). Another challenging aspect of these algorithms is the effective computation of d_k . For instance, in the unconstrained case, we need to solve the minimization problem (13), which can be done by applying a Gauss-Seidel-like method to

$$(168) \quad \begin{aligned} & \text{minimize} \\ & \Lambda = (\lambda_1, \dots, \lambda_q) \in \mathcal{S}^q \\ & (p_1, \dots, p_q) \in \mathcal{H}^q \end{aligned} \quad \frac{1}{2} \left\| \sum_{i=1}^q \lambda_i p_i \right\|^2 + \delta_{\mathcal{S}^p}(\Lambda) + \sum_{i=1}^q \delta_{\partial f_i(u^k)}(p_i).$$

Problem (14) is also well suited for primal-dual methods, and perhaps other methods could be examined and compared. To our knowledge, this work has never been done, and is a subject for further study.

More recently, a trust-region method for unconstrained multiobjective problems involving smooth functions has been developed in [41], which uses the norm of the multiobjective steepest descent vector as a generalized marginal function. In [26, 28], a Newton method for unconstrained strongly convex vector optimization has been developed, with a local superlinear convergence result. Instead of taking d_k as a descent direction computed from first-order quadratic models as in (14), the authors use *second-order* quadratic models to define a multiobjective Newton direction as:

$$(169) \quad d_k = \operatorname{argmin}_{d \in \mathcal{H}} \left\{ \max_{i=1, \dots, q} \langle \nabla f_i(u_k), d \rangle + \frac{1}{2} \langle \nabla^2 f_i(u_k) d, d \rangle \right\}.$$

As in Theorem 1.9, they show that this discrete dynamic corresponds to the classical Newton's method applied to a weighted combination of the objective functions, but with an endogenous scalarization. In other words, at each step, the algorithm provides $(\theta_i^k) \in \mathcal{S}^q$ such that

$$(170) \quad d_k = - \left(\sum_{i=1}^q \theta_i^k \nabla^2 f_i(u_k) \right)^{-1} \left(\sum_{i=1}^q \theta_i^k \nabla f_i(u_k) \right).$$

See also [35] for second-order models built with a BFGS scheme, to avoid the direct computation of the Hessian. These works suggest the existence of corresponding continuous Newton-like dynamics (see for example [8] in the case of a single objective).

5. CONCLUSION, PERSPECTIVES

In this paper, we have shown some remarkable properties of the multiobjective steepest descent direction, and of the dynamical system which is governed by the corresponding vector field: along each trajectory, all the objective functions are decreasing, and there is convergence to a weak Pareto minimum. Working in a general Hilbert space, and with convex continuous functions (not necessarily differentiable) allows us to cover a wide range of applications. However, there are many issues to be resolved. Among the most challenging, let us mention the uniqueness of the solution, for a given Cauchy data, and the dynamical properties of the weighting coefficients. The natural link between the (MOG) dynamic and the theory of gradient flows naturally suggests to study the dynamics for semi-algebraic functions, on the basis of Kurdyka-Lojasiewicz inequality. Obtaining rapid methods based on an analysis of second order in time (inertial aspects), or space (Newton-like methods) is important both from the numerical, and modeling point of view. It would be also interesting to consider interior point methods. Some modeling aspects in game theory, economics, and inverse problems, have been considered in the previous sections. They are still largely unexplored. All these results suggest that there is a broad class of continuous dynamics that contains (MOG), and having similar properties with respect to Pareto equilibration. Enriching this class of dynamics can be useful for numerical purpose, and for understanding the complex interactions in Pareto equilibration (coalitions, negotiation, bargaining, dealing with uncertainty, changes in the environment, psychological aspects). These are interesting topics for further research.

REFERENCES

- [1] H. Attouch, *Variational convergence for functions and operators*, Pitman Advanced Publishing Program, Applicable Mathematics Series, 1984.
- [2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE's*, *J. Convex Analysis*, **15**(3) (2008), pp. 485–506.
- [3] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in Sobolev and BV spaces. Applications to PDE's and optimization*, MPS/SIAM Series on Optimization, 6, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006, 634 pages.
- [4] H. Attouch and M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, *J. Differential Equations*, **248**(6) (2010), pp. 1315–1344.
- [5] H. Attouch and A. Damlamian, *On multivalued evolution equations in Hilbert spaces*, *Israel J. Math.*, **12** (1972), pp. 373–390.
- [6] H. Attouch and X. Goudou, *A continuous gradient-like dynamical approach to Pareto-optimization in Hilbert spaces*, *Set-Valued and Variational Analysis*, **22**(1) (2014), pp. 189–219.
- [7] H. Attouch and A. Soubeyran, *Local search proximal algorithms as decision dynamics with costs to move*, *Set Valued and Variational Analysis*, **19**(1) (2011), pp. 157–177.
- [8] H. Attouch and B. F. Svaiter, *A continuous dynamical Newton-Like approach to solving monotone inclusions*, *SIAM J. Control Optim.*, **49**(2) (2011), pp. 574–598.
- [9] J.P. Aubin and A. Cellina, *Differential inclusions*, Springer, Berlin, 1984.
- [10] J.P. Aubin and I. Ekeland, *Applied nonlinear analysis*, Wiley, 1984.
- [11] T. Q. Bao and B. Mordukhovich, *Sufficient conditions for global weak Pareto solutions in multiobjective optimization*, *Positivity*, **16** (2012), pp. 579–602.
- [12] H. Bauschke and P. Combettes, *Convex analysis and monotone operator theory*, CMS books in Mathematics, Springer, 2011.
- [13] H. Benabdellah, C. Castaing, and A. Salvadori, *Compactness and discretization methods for differential inclusions and evolution problems*, *Atti. Sem. Mat. Univ. Modena*, XLV (1997), pp. 9–51.
- [14] H. Bonnel, A. N. Iusem, and B. F. Svaiter, *Proximal methods in vector optimization*, *SIAM Journal on Optimization*, **15** (2005), pp. 953–970.

- [15] P.A.N. Bosman, *On gradients and hybrid evolutionary algorithms for real-valued multi-objective optimization*, IEEE Transactions on Evolutionary Computation **16**(1), (2012), pp. 51–69.
- [16] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland/Elsevier, New-York, 1973.
- [17] H. Brézis, *Analyse fonctionnelle*, Masson, 1983.
- [18] M. Brown and R. E. Smith, *Directed multi-objective optimisation*, International Journal on Computers, Systems and Signals, **6**(1) (2005), pp. 3–17.
- [19] R.E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, J. Funct. Anal., **18** (1975), pp. 15–26.
- [20] A. Cellina and V. Staicu, *On evolution equations having monotonicities of opposite sign*, Journal Diff. Equations, **90** (1991), pp. 71–80.
- [21] B. Cornet, *Contributions to the mathematical theory of dynamical mechanisms of resource allocation*, Thèse de doctorat d'état, Université Paris IX Dauphine (1981).
- [22] B. Cornet, *Monotone planning procedures and accessibility of Pareto optima*, in New Trends in Dynamic System Theory and Economics, Aoki et Marzollo (eds.), Academic Press (1979), pp. 337–349.
- [23] B. Cornet, *Existence of slow solutions for a class of differential inclusions*, Journal of Mathematical Analysis and Applications, **96** (1983), pp. 130–147.
- [24] D.-L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, **52**(4) (2006), pp. 1289–1306.
- [25] J. Fliege and B. F. Svaiter, *Steepest descent methods for multicriteria optimization*, Mathematical Methods of Operations Research, **51**(3) (2000), pp. 479–494.
- [26] J. Fliege, L. M. Graña Drummond and B. F. Svaiter, *Newton's method for multiobjective optimization*, SIAM Journal on Optimization, **20**(2) (2009), pp.602–626.
- [27] L. M. Graña Drummond and B. F. Svaiter, *A steepest descent method for vector optimization*, Journal of Computational and Applied Mathematics, **175** (2005), pp. 395–414.
- [28] L. M. Graña Drummond, F.M.P. Raupp, and B. F. Svaiter, *A quadratically convergent Newton method for vector optimization*, Optimization, 2012.
- [29] L. M. Graña Drummond and L.M. Iusem, *A projected gradient method for vector optimization problems*, Comput. Optim. Appl., **28**(1) (2004), pp. 5–29.
- [30] X. Goudou and J. Munier, *The gradient and heavy ball with friction dynamical systems: the quasiconvex case*, Math. Program., **116**(1-2) (2009), pp. 173–191.
- [31] Cl. Henry, *An existence theorem for a class of differential equations with multivalued right-hand side*, Journal of Mathematical Analysis and Applications, **41** (1973), pp. 168–179.
- [32] J.J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires*, Comptes Rendus de l'Académie des Sciences Paris, Série A, **225** (1962), pp. 238–240.
- [33] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc. **73** (1967), pp. 591–597.
- [34] J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. of Convex Analysis, **17** (2010), pp. 1113–1163.
- [35] Z. Povalej, *Quasi-Newton method for multiobjective optimization*, Journal of Computational and Applied Mathematics, **255** (2014), pp. 765–777.
- [36] R. T. Rockafellar, *Integrals which are convex functionals*, Pacific J. Math., **24** (1968), pp. 525–539.
- [37] R. T. Rockafellar and R. Wets, *Variational Analysis*, Grundlehren der Mathematischen Wissenschaften, **317**, Springer, 1998.
- [38] S. Smale, *Global analysis and economics I: Pareto optimum and a generalization of Morse theory*, Dynamical Systems, Academic Press, 1973.
- [39] D. The Luc, *Theory of vector optimization*, Springer, Berlin, 1989.
- [40] R. Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society, Series B, **58**(1) (1996), pp. 267–288.
- [41] K. D.V. Villacorta, P.R. Oliveira, and A. Soubeyran, *A trust-region method for unconstrained multiobjective problems with applications in satisficing processes*, J. Optim. Theory Appl., **160** (2014), pp. 865–889.
- [42] E. Zeidler, *Nonlinear functional analysis and its applications, Part II: Monotone operators*, Springer-Verlag, New-York, 1990.