Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-024-02389-3

Fast Multiobjective Gradient Methods with Nesterov Acceleration via Inertial Gradient-Like Systems

Konstantin Sonntag[1] · **Sebastian Peitz**[2]

Received: 16 November 2022 / Accepted: 13 January 2024
© The Author(s) 2024

Abstract
We derive efficient algorithms to compute weakly Pareto optimal solutions for smooth, convex and unconstrained multiobjective optimization problems in general Hilbert spaces. To this end, we define a novel inertial gradient-like dynamical system in the multiobjective setting, which trajectories converge weakly to Pareto optimal solutions. Discretization of this system yields an inertial multiobjective algorithm which generates sequences that converge weakly to Pareto optimal solutions. We employ Nesterov acceleration to define an algorithm with an improved convergence rate compared to the plain multiobjective steepest descent method (Algorithm 1). A further improvement in terms of efficiency is achieved by avoiding the solution of a quadratic subproblem to compute a common step direction for all objective functions, which is usually required in first-order methods. Using a different discretization of our inertial gradient-like dynamical system, we obtain an accelerated multiobjective gradient method that does not require the solution of a subproblem in each step (Algorithm 2). While this algorithm does not converge in general, it yields good results on test problems while being faster than standard steepest descent.

Keywords Multiobjective optimization · Gradient methods · Nesterov acceleration · Inertial dynamics · Lyapunov analysis

Communicated by Margaret M. Wiecek.

✉ Konstantin Sonntag
konstantin.sonntag@uni-paderborn.de

Sebastian Peitz
sebastian.peitz@uni-paderborn.de

1 Department of Mathematics, Paderborn University, 33098 Paderborn, Germany

2 Department of Computer Science, Paderborn University, 33098 Paderborn, Germany

Published online: 17 February 2024

1 Introduction

In many applications in industry, economics, medicine or transport, optimizing several criteria is of interest. In the latter, one wants to reach a destination as fast as possible with minimal power consumption. Drug development aims for maximizing efficacy while minimizing side effects. Even these elementary examples share an inherent feature. The different criteria one seeks to optimize are in general contradictory. There is no design choice that is best for all criteria simultaneously. This insight shifts the focus from finding a single optimal solution to a set of optimal compromises—the Pareto set. Given the Pareto set, a decision maker can select an optimal compromise according to their preferences. In this paper, we derive efficient gradient-based algorithms to compute elements of the Pareto set. Formally, a problem involving multiple criteria can be described as an unconstrained multiobjective optimization problem

$$\min_{x \in \mathcal{H}} \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, \tag{MOP}$$

where $f_i : \mathcal{H} \to \mathbb{R}$ for $i = 1, \ldots, m$ are the objective functions describing the different criteria. Popular approaches to tackle this problem in the differentiable case are first-order methods which exploit the smooth structure of the problem while not being computationally demanding compared to higher-order methods involving exact or approximated Hessians.

While in single-objective optimization, accelerated first-order methods are very popular, these methods are not studied sufficiently from a theoretical point of view in the multiobjective setting. A fruitful approach to analyze accelerated gradient methods is to interpret them as discretizations of suitable gradient-like dynamical systems [31]. The analysis of the continuous dynamics is often easier and can later on be transferred to the discrete setting. So far, this perspective is not fully taken advantage of in the area of multiobjective optimization. In this paper, we utilize this approach to derive accelerated gradient methods for multiobjective optimization. To this end, we define and analyze the following novel dynamical gradient-like system

$$\ddot{x}(t) + \alpha \dot{x}(t) + \operatorname*{proj}_{C(x(t))} (-\ddot{x}(t)) = 0, \tag{IMOG'}$$

with $\alpha > 0$, $C(x) := \operatorname{conv}(\{\nabla f_i(x) : i = 1, \ldots, m\})$, where $\operatorname{conv}(\cdot)$ denotes the convex hull and $\operatorname{proj}_{C(x(t))}(-\ddot{x}(t))$ is the projection of $-\ddot{x}(t)$ onto the convex set $C(x(t))$. The system (IMOG') is an *inertial multiobjective gradient-like system*. We choose the designation (IMOG') to emphasize its relation to the system

$$\mu \ddot{x}(t) + \gamma \dot{x}(t) + \operatorname*{proj}_{C(x(t))} (0) = 0, \tag{IMOG}$$

with $\mu, \gamma > 0$, which was discussed in [7]. In the single-objective setting ($m = 1$), both (IMOG') and (IMOG) reduce to the *heavy ball with friction system*

$$\mu\ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) = 0, \tag{HBF}$$

which is well studied for different types of objective functions f, see, e.g., [1, 10, 27].

We discretize (IMOG') to obtain an iterative scheme of the form

$$x^{k+1} = x^k + a(x^k - x^{k-1}) - b\sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k),$$

with appropriately chosen coefficients $a, b > 0$ and $\theta^k \in \mathbb{R}^m$. This scheme can be interpreted as an inertial gradient method for (MOP). We show that it shares many properties with its continuous counterpart and that iterates defined by this algorithm converge weakly to Pareto critical points. To the best of our knowledge this is the first multiobjective method involving a constant momentum term with guaranteed convergence to Pareto critical solutions.

In a further step, we introduce time-dependent friction and informally define the following *multiobjective gradient-like system with asymptotically vanishing damping*

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))}(-\ddot{x}(t)) = 0. \tag{MAVD}$$

A discussion of the system (MAVD) can be found in [30], where it is shown that trajectories of (MAVD) converge weakly to weakly Pareto optimal solutions with fast convergence of the objective values to an optimal value along the trajectories. In the single-objective setting, this system simplifies to the following inertial system with *asymptotically vanishing damping*

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0. \tag{AVD}$$

It is well-known that (AVD) is naturally linked with Nesterov's accelerated gradient method [4, 5, 11, 31]. Discretizing the dynamical system (MAVD), and using our knowledge about (IMOG'), we derive an accelerated gradient method for multiobjective optimization that takes the form

$$x^{k+1} = x^k + \frac{k-1}{k+2}(x^k - x^{k-1}) - b\sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k),$$

with appropriately chosen coefficients $b > 0$ and $\theta^k \in \mathbb{R}^m$. Tanabe, Fukuda and Yamashita derive an accelerated proximal gradient method for multiobjective optimization using the concept of merit functions [34]. We show that the method we derive from the differential equation (MAVD) achieves the same convergence rate of order $\mathcal{O}(k^{-2})$ for the function values, measured with a merit function.

The remainder of the paper is organized as follows. After introducing some basic definitions and notations in Sect. 2, we prove that solutions to the system (IMOG') exist in finite-dimensional Hilbert spaces in Sect. 3, and show that they converge to

Pareto critical points in Sect. 4. Based on that, we derive a discrete optimization algorithm from an explicit discretization of (IMOG') and show that the iterates defined by this method converge weakly to Pareto critical points, in Sect. 5. Then, we introduce Nesterov acceleration and prove an improved convergence result in Sect. 6. The numerical efficiency of the new methods is discussed in Sect. 7. The two central algorithms are summarized in Algorithms 1 and 2 in the respective sections. We compare the methods on convex and nonconvex example problems in Sect. 8 and conclude our findings and list future research directions in Sect. 9.

2 Background

2.1 Notation

Throughout this paper, $\mathcal{H}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. We denote the open ball with radius $\delta > 0$ and center x by $B_\delta(x) := \{y \in \mathcal{H} \ : \ \|y - x\| < \delta\}$. The closed ball with radius $\delta > 0$ and center x is denoted by $\overline{B_\delta(x)}$. The set $\Delta^m := \left\{\alpha \in \mathbb{R}^m \ : \ \alpha \geq 0, \text{ and } \sum_{i=1}^m \alpha_i = 1\right\}$ is the positive unit simplex. For a set of vectors $\{\xi_1, \ldots, \xi_m\} \subseteq \mathcal{H}$ we denote the convex hull of these vectors by $\operatorname{conv}(\{\xi_1, \ldots, \xi_m\}) := \left\{\sum_{i=1}^m \alpha_i \xi_i \ : \ \alpha \in \Delta^m\right\}$. For a closed convex set $C \subseteq \mathcal{H}$ the projection of a vector $x \in \mathcal{H}$ onto C is $\operatorname{proj}_C(x) := \arg\min_{y \in \mathcal{H}} \|y - x\|^2$. For two vectors $x, y \in \mathbb{R}^m$, we define the partial order $x \leq y :\Leftrightarrow x_i \leq y_i$ for all $i = 1, \ldots, m$. We define $\geq, <, >$ on $\mathbb{R}^m$ analogously. When we treat dynamical systems, $t \in \mathbb{R}$ and $x \in \mathcal{H}$ are the time and state variable, respectively. We denote trajectories in $\mathcal{H}$ with $t \mapsto x(t)$ with first derivative $\dot{x}(t)$ and second derivative $\ddot{x}(t)$.

2.2 Multiobjective Optimization

Consider the unconstrained multiobjective optimization problem

$$\min_{x \in \mathcal{H}} \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, \tag{MOP}$$

with continuously differentiable objective functions $f_i : \mathcal{H} \to \mathbb{R}$ for $i = 1, \ldots, m$. The definitions in this subsection are aligned with [23].

Definition 2.1 Consider the multiobjective optimization problem (MOP).

i) A point $x^* \in \mathcal{H}$ is Pareto optimal if there does not exist another point $x \in \mathcal{H}$ such that $f_i(x) \leq f_i(x^*)$ for all $i = 1, \ldots, m$, and $f_j(x) < f_j(x^*)$ for at least one index j. The set of all Pareto optimal points is the Pareto set, which we denote by P.
ii) A point $x^* \in \mathcal{H}$ is locally Pareto optimal if there exists $\delta > 0$ such that x^* is Pareto optimal in $B_\delta(x^*)$.

iii) A point $x^* \in \mathcal{H}$ is weakly Pareto optimal if there does not exist another vector $x \in \mathcal{H}$ such that $f_i(x) < f_i(x^*)$ for all $i = 1, \ldots, m$.
iv) A point $x^* \in \mathcal{H}$ is locally weakly Pareto optimal if there exists $\delta > 0$ such that x^* is weakly Pareto optimal in $B_\delta(x^*)$.

In this paper, we treat convex MOPs, i.e., the objective functions f_i are convex for all $i = 1, \ldots, m$. In this setting, every locally (weakly) Pareto optimal point is also (weakly) Pareto optimal. For unconstrained MOPs, the so-called Karush–Kuhn–Tucker conditions can be written as follows.

Definition 2.2 A point $x^* \in \mathcal{H}$ satisfies the Karush–Kuhn–Tucker conditions if there exists $\alpha \in \Delta^m$ such that $\sum_{i=1}^m \alpha_i \nabla f_i(x^*) = 0$. If x^* satisfies the Karush–Kuhn–Tucker conditions, we call it Pareto critical.

The condition $0 \in \operatorname{conv}(\{\nabla f_i(x^*) : i = 1, \ldots, m\})$ is equivalent to the Karush–Kuhn–Tucker conditions. Analogously to the single-objective setting, criticality of a point is a necessary condition for optimality. In the convex setting, the KKT conditions are also sufficient conditions for weak Pareto optimality. We denote the Pareto set by P, the weak Pareto set by P_w and the Pareto critical set by P_c. In the setting of smooth and convex multiobjective optimization, we observe the relation

$$P \subset P_w = P_c.$$

2.3 Accelerated Methods for Multiobjective Optimization

Accelerated methods for multiobjective optimization are not sufficiently discussed from a theoretical point of view in the literature yet. In [18] El Moudden and El Moutasim propose an accelerated method for multiobjective optimization which incorporates the multiobjective descent direction by Fliege [19] and the same acceleration scheme as in Nesterov's accelerated method [25]. El Moudden and El Moutasim prove a convergence rate of the function values with rate $\mathcal{O}(k^{-2})$. Their proof relies on the restrictive assumption that the Lagrange multipliers of the quadratic subproblem, that is used to compute the step direction in every iteration, remain fixed from a certain point on. Under this assumption, the method simplifies to Nesterov's method for single-objective optimization problems applied to a weighted sum of the objective functions with fixed weights. Only recently, Tanabe, Fukuda and Yamashita derived an accelerated proximal gradient method for multiobjective optimization problems in [34]. They developed their method using the concept of merit functions (see Sect. 2.5) and show that the function values converge with rate $\mathcal{O}(k^{-2})$ without additional assumptions on the Lagrange multipliers.

2.4 Dynamical Systems Linked to Multiobjective Optimization

In [29] Smale presents the idea of treating multiobjective optimization problems with a continuous time perspective that is motivated from an economical point of view using utility functions in a multi-agent framework. The simplest dynamical system for multiobjective optimization problems is the *multiobjective gradient system*

$$\dot{x}(t) + \underset{C(x(t))}{\text{proj}}\,(0) = 0, \tag{MOG}$$

where $C(x(t)) = \text{conv}\,(\{\nabla f_i(x(t)) \,:\, i = 1, \ldots, m\})$. This system is already treated in [20] and in addition by Cornet in [16]. In [9, 24] the system (MOG) gets introduced as a tool for multiobjective optimization. The system (MOG) can also be seen as a continuous version of the multiobjective steepest descent method by Fliege [19]. In the single-objective setting ($m = 1$), the system (MOG) simplifies to the *steepest descent dynamical system* $\dot{x}(t) + \nabla f(x(t)) = 0$. Generalizations of (MOG) are treated in [8, 9]. In [9] Attouch and Goudou discuss a dynamical system for constrained minimization and in [8] Attouch, Garrigos and Goudou present a differential inclusion for constrained nonsmooth optimization.
In [7], Attouch and Garrigos introduce inertia in the system (MOG) and define the following *inertial multiobjective gradient-like dynamical system*

$$\mu\ddot{x}(t) + \gamma\dot{x}(t) + \underset{C(x(t))}{\text{proj}}\,(0) = 0. \tag{IMOG}$$

Trajectories of (IMOG) converge weakly to Pareto optimal solutions given $\gamma^2 > \mu L$, where L is a common Lipschitz constant of the gradients of the objective functions.

2.5 Merit Functions

A merit function associated with an optimization problem is a function that returns zero at an optimal solution and which is strictly positive otherwise. An overview on merit functions used in multiobjective optimization is given in [35]. In our proofs we use the merit function

$$u_0(x) := \sup_{z \in \mathcal{H}} \min_{i=1,\ldots,m} f_i(x) - f_i(z), \tag{1}$$

which satisfies the following statement.

Theorem 2.3 *It holds that $u_0(x) \geq 0$ for all $x \in \mathcal{H}$. Moreover, $x \in \mathcal{H}$ is weakly Pareto optimal for (MOP), if and only if $u_0(x) = 0$.*

Proof A proof of this result can be found in Theorem 3.1 in [35]. □

Additionally, $u_0(x)$ is lower semicontinuous. Therefore, if $(x^k)_{k\geq 0}$ is a sequence with $u_0(x^k) \to 0$, every cluster point of $(x^k)_{k\geq 0}$ is weakly Pareto optimal. This motivates the usage of $u_0(x)$ as a measure of complexity for multiobjective optimization methods. The function $u_0(x)$ is not the only merit function for multiobjective optimization problems, see also [15, 21, 37] and further references in [35].

3 Global Existence in Finite Dimensions

In this section, we show that solutions exist for the Cauchy problem related to (IMOG'), i.e,

$$\left|\begin{array}{l}\ddot{x}(t)+\alpha\dot{x}(t)+\operatorname{proj}_{C(x(t))}(-\ddot{x}(t))=0,\\ \\ x(0)=x_0,\quad \dot{x}(0)=v_0,\end{array}\right. \tag{CP}$$

with initial data $x_0, v_0 \in \mathcal{H}$. To this end, we show that for this system solutions exist if there exists a solution to a first-order differential inclusion

$$(\dot{u}, \dot{v}) \in F(u, v),$$

with a set-valued map $F : \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}$. Then, we use an existence theorem for differential inclusions from [12]. Our argument works only in finite-dimensional Hilbert spaces. Thus, we assume $\dim(\mathcal{H}) < +\infty$ from here on. In our context, the following set-valued map is of interest:

$$F : \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}, \quad (u, v) \mapsto \{v\} \times \left(\left(-\operatorname*{arg\,min}_{z \in C(u)} \langle z, v\rangle \right) - \alpha v \right). \tag{2}$$

As stated above, $C(u) := \operatorname{conv}(\{\nabla f_i(u) \,:\, i = 1, \ldots, m\})$. We can show that (CP) has a solution if the differential inclusion

$$\left|\begin{array}{l}(\dot{u}(t), \dot{v}(t)) \in F(u, v),\\ \\ (u(0), v(0)) = (u_0, v_0),\end{array}\right. \tag{DI}$$

with appropriate initial data u_0, v_0 has a solution.

Remark 3.1 The motivation for the choice of the differential equation (IMOG') opposing to the choice (IMOG) in [7] is the energy estimate in Proposition 4.1. Solutions x to the Cauchy problem (CP) naturally satisfy the following energy estimate, which holds in the single-objective setting for the heavy ball with friction dynamical system.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2 \right] \leq -\alpha\|\dot{x}(t)\|^2, \text{ for all } i = 1, \ldots, m \text{ and } t \geq 0. \tag{3}$$

In fact, we discovered the system (IMOG') by starting from relation (3) and interpreting it as a variational inequality. Inequality (3) does in general not hold for the system (IMOG) considered in [7].

3.1 Existence of Solutions to (DI)

To show that there exist solutions to (DI), we investigate the set-valued map $(u, v) \rightrightarrows F(u, v)$ defined in (2). The basic definitions for set-valued maps used in this subsection can be found in [12].

Proposition 3.2 *For all $(u, v) \in \mathcal{H} \times \mathcal{H}$, $F(u, v) \subset \mathcal{H} \times \mathcal{H}$ is convex and compact.*

Proof The statement follows directly from the definition. □

To use an existence theorem from [12], we need to show that $(u, v) \rightrightarrows F(u, v)$ is upper semicontinuous. Showing this is elementary. We omit the full proof here but sketch a possible way to prove this result.

Lemma 3.3 *Let $C(u) := \operatorname{conv}(\{c_i(u) : i = 1, \ldots, m\})$ with $c_i : \mathcal{H} \to \mathcal{H}$, $u \to c_i(u)$ continuous for $i = 1, \ldots, m$. Let $(u_0, v_0) \in \mathcal{H} \times \mathcal{H}$ be fixed.*
Then, for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $(u, v) \in \mathcal{H} \times \mathcal{H}$ with $\|u - u_0\| < \delta$ and $\|v - v_0\| < \delta$ and for all $z \in \arg\min_{z \in C(u)} \langle z, v \rangle$ there exists $z_0 \in \arg\min_{z_0 \in C(u_0)} \langle z_0, v_0 \rangle$ with $\|z - z_0\| < \varepsilon$.

Proof The proof follows by continuity arguments. □

Proposition 3.4 *The set-valued map $(u, v) \rightrightarrows F(u, v)$ is upper semicontinuous.*

Proof Using Lemma 3.3 we can show in a straightforward manner

$$F((u_0, v_0) + B_\delta(0)) \subset F(u_0, v_0) + B_\varepsilon(0),$$

using only continuity arguments. Then, the statement follows by the fact that $(u, v) \rightrightarrows F(u, v)$ is locally compact. □

Proposition 3.5 *Let $\mathcal{H}$ have finite dimension. Then, the mapping*

$$\phi : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}, \quad (u, v) \mapsto \left(v, \operatorname*{proj}_{F(u,v)}(0) \right),$$

is locally compact.

Proof If $\dim(\mathcal{H}) < +\infty$ the proof follows easily since all images $F(u, v)$ are compact and depend on (u, v) in a well-behaved manner. On the other hand, from ϕ being locally compact, we get that $v \mapsto v$ is locally compact which is equivalent to $\mathcal{H}$ being finite-dimensional. □

The following existence theorem from [12] is applicable in our setting.

Theorem 3.6 *Let $\mathcal{X}$ be a Hilbert space and let $\Omega \subset \mathbb{R} \times \mathcal{X}$ be an open subset containing $(0, x_0)$. Let $G : \Omega \rightrightarrows \mathcal{X}$ be an upper semicontinuous set-valued map such that $G(\omega)$ is nonempty, convex and closed for all $\omega \in \Omega$. We assume that $\omega \mapsto \operatorname{proj}_{G(\omega)}(0)$ is locally compact on Ω. Then, there exists $T > 0$ and an absolutely continuous function $x(\cdot)$ defined on $[0, T]$ which is a solution to the differential inclusion*

$$\dot{x}(t) \in G(t, x(t)), \quad x(0) = x_0.$$

Proof A proof of this theorem can be found in Theorem 3 in [12, p. 98]. □

We are finally in the position to state an existence theorem for (DI).

Theorem 3.7 *Assume $\mathcal{H}$ is finite-dimensional and that the gradients of the objective function ∇f_i are globally Lipschitz continuous. Then, for all $(u_0, v_0) \in \mathcal{H} \times \mathcal{H}$ there exists $T > 0$ and an absolutely continuous function $(u(\cdot), v(\cdot))$ defined on $[0, T]$ which is a solution to the differential inclusion (DI).*

Proof The proof follows immediately from Propositions 3.1 and 3.5 which show that the set-valued map F given in (2) satisfies all conditions required for Theorem 3.6. □

In the following, we show that under additional conditions on the objective functions f_i, there exist solutions defined on $[0, +\infty)$. The extension of solutions is achieved following a standard argument. We show that the solutions to (DI) remain bounded. Then, we use Zorn's Lemma to retrieve a contradiction if there is a maximal solution that is not defined on $[0, +\infty)$.

Theorem 3.8 *Assume $\mathcal{H}$ is finite-dimensional and that the gradients of the objective function ∇f_i are globally Lipschitz continuous. Then, for all $(u_0, v_0) \in \mathcal{H} \times \mathcal{H}$ there exists an absolutely continuous function $(u(\cdot), v(\cdot))$ defined on $[0, +\infty)$ which is a solution to the differential inclusion (DI).*

Proof Theorem 3.7 guarantees the existence of solutions defined on $[0, T)$ for some $T \geq 0$. Using the domain of definition, we can define a partial order on the set of solutions to the problem (DI). Assuming there is no solution defined on $[0, +\infty)$, Zorn's Lemma guarantees the existence of a solution $(u(\cdot), v(\cdot)) : [0, T) \to \mathcal{H} \times \mathcal{H}$ with $T < +\infty$ which cannot be extended. We will show that $(u(t), v(t))$ does not blow up in finite time and therefore can be extended which contradicts the claimed maximality.
Define

$$h(t) := \|(u(t), v(t)) - (u(0), v(0))\|_{\mathcal{H}\times\mathcal{H}},$$

where $\|(x, y)\|_{\mathcal{H}\times\mathcal{H}} = \sqrt{\|x\|^2 + \|y\|^2}$. We show that $h(t)$ can be bounded by a real-valued function. Using the Cauchy–Schwarz inequality, we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}h^2(t) &= \langle(\dot{u}(t), \dot{v}(t)), (u(t), v(t)) - (u(0), v(0))\rangle_{\mathcal{H}\times\mathcal{H}} \\ &\leq \|(\dot{u}(t), \dot{v}(t))\|\, h(t) \\ &\leq \max_{\xi \in F(u(t), v(t))} \|\xi\|_{\mathcal{H}\times\mathcal{H}} h(t). \end{aligned} \tag{4}$$

We next derive a bound on $\max_{\xi \in F(u(t),v(t))} \|\xi\|_{\mathcal{H}\times\mathcal{H}}$. The basic inequalities between the ℓ_1 and ℓ_2 norm applied to $(\|x\|, \|y\|) \in \mathbb{R}^2$ yield

$$\|(x, y)\|_{\mathcal{H}\times\mathcal{H}} \leq \|x\| + \|y\| \leq \sqrt{2}\|(x, y)\|_{\mathcal{H}\times\mathcal{H}}.$$

Let $(u, v) \in \mathcal{H}\times\mathcal{H}$. Using $C(u) = \operatorname{conv}(\{\nabla f_i(u) : i = 1, \ldots, m\})$ and the definition of $F(u, v)$ from (2), we have

$$
\begin{aligned}
\max_{\xi \in F(u,v)} \|\xi\|_{\mathcal{H}\times\mathcal{H}} &\leq \|v\| + \max_{z \in C(u)} \|z - \alpha v\| \\
&\leq (1+\alpha)\|v\| + \max_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i \nabla f_i(u) \right\| \\
&\leq (1+\alpha)\|v\| + \max_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i \left(\nabla f_i(u) - \nabla f_i(0)\right) \right\| + \max_{\theta \in \Delta^m} \left\| \sum_{i=1}^m \theta_i \nabla f_i(0) \right\| \\
&\leq (1+\alpha)\|v\| + L\|u\| + \max_{i=1,\ldots,m} \|\nabla f_i(0)\| \\
&\leq c(1 + \|(u,v)\|_{\mathcal{H}\times\mathcal{H}}),
\end{aligned}
\tag{5}
$$

where we chose $c = \sqrt{2}\max\left\{1+\alpha, L, \max_{i=1,\ldots,m}\|\nabla f_i(0)\|\right\}$. Combining inequalities (4) and (5), we can show

$$
\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}h^2(t) \leq \tilde{c}(1+h(t))h(t), \quad \text{for all } t \in [0,T),
$$

with $\tilde{c} \geq 0$. Using a Gronwall-type argument (see Lemma A.4 and Lemma A.5 in [14]) just as in Theorem 3.5 in [7], we know that there exists $C \geq 0$ such that for an arbitrary $\varepsilon > 0$

$$
h(t) \leq CT\exp(CT), \quad \text{for all } \ t \in [0, T-\varepsilon].
$$

Since this upper bound is independent of t and ε, it follows that $h \in L^\infty([0,T],\mathbb{R})$. Therefore, solutions to (DI) do not blow up in finite time and can be extended. This is a contradiction to the maximality of the solution $(u(t), v(t))$. □

3.2 Existence of Solutions to (CP)

Using the findings of the previous subsection, we can proceed with the discussion of the Cauchy problem (CP). In this subsection, we show that solutions to the differential inclusion (DI) immediately give solutions to the Cauchy problem (CP).

Theorem 3.9 *Let $x_0, v_0 \in \mathcal{H}$. Assume that $(u(t), v(t))$ for $t \in [0,+\infty)$ is a solution to (DI) with $(u(0), v(0)) = (x_0, v_0)$. Then, it follows that $x(t){:=}u(t)$ satisfies the differential equation*

$$
\ddot{x}(t) + \alpha\dot{x}(t) + \operatorname*{proj}_{C(x(t))}(-\ddot{x}(t)) = 0, \ \textit{for almost all } t \in (0,+\infty),
$$

and $x(0) = x_0$, $\dot{x}(0) = v_0$, where $C(x) = \operatorname{conv}(\{\nabla f_i(x) \ : \ i = 1,\ldots,m\})$.

Proof Since $(u(t), v(t))$ is a solution to (DI), it follows from the definition of set-valued map F given in (2) that for almost all $t \in (0,+\infty)$

$$\dot{u}(t) = v(t),$$
$$\dot{v}(t) \in -\operatorname*{arg\,min}_{z \in C(u(t))} \langle z, v(t)\rangle - \alpha v(t),$$

holds. Using Lemma A.1, the second line gives $-\alpha v(t) = \operatorname{proj}_{C(u(t))+\dot{v}(t)}(0)$, which is equivalent to

$$\dot{v}(t) + \alpha v(t) + \operatorname*{proj}_{C(u(t))}(-\dot{v}(t)) = 0.$$

Rewriting this system using $x(t) = u(t)$, $\dot{x}(t) = \dot{u}(t) = v(t)$ and $\ddot{x}(t) = \dot{v}(t)$ and verifying the initial conditions $x(0) = u(0) = x_0$ and $\dot{x}(0) = v(0) = v_0$ yields the desired result. □

Finally, we can state the full existence theorem for the Cauchy problem (CP).

Theorem 3.10 *Assume $\mathcal{H}$ is finite-dimensional and that the gradients of the objective function ∇f_i are globally Lipschitz continuous. Then, for all $x_0, v_0 \in \mathcal{H}$, there exists a continuously differentiable function x defined on $[0, +\infty)$ which is absolutely continuous with absolutely continuous first derivative $\dot{x}$, and which is a solution to the Cauchy problem (CP) with initial values (x_0, v_0).*

Proof The proof follows immediately combining Theorem 3.8 and Theorem 3.9. □

Remark 3.11 Throughout this section, we have assumed that the gradients ∇f_i of the objective functions are globally Lipschitz continuous. One can relax this condition and only require the gradients to be Lipschitz continuous on bounded sets, if we can guarantee that the solutions remain bounded. This holds for example if one of the objective functions f_i has bounded level sets.

Remark 3.12 The uniqueness of solutions to the differential inclusion (DI) and the Cauchy problem (CP) remain an open problem even in finite dimensions. There are two main problems which can be seen best by considering the implicit differential equation in (CP). Firstly, the steepest descent vector field $s : \mathcal{H} \to \mathcal{H}$, $x \mapsto \operatorname{proj}_{C(x)}(0)$ is in general neither monotone nor Lipschitz continuous but merely $\frac{1}{2}$-Hölder continuous (see [32]). There is a remedy for this problem requiring an extra assumption. If the set $\{\nabla f_i(\overline{u}) : i = 1, \ldots, m\}$ of gradients in $\overline{u}$ is affinely independent, then there exists a neighborhood $B_\rho(\overline{u})$ of $\overline{u}$ with $\rho > 0$ such that the steepest descent vector field is Lipschitz continuous on $B_\rho(\overline{u})$ (see [8, Proposition 3.4]). With this result the (local) uniqueness of solutions to the multiobjective steepest descent dynamical system $\dot{x}(t) + \operatorname{proj}_{C(x(t))}(0) = 0$ with $x_0 = \overline{u}$ can be shown.
The implicit structure of the differential equation in (CP) is the second problem. We are not only dealing with the steepest descent vector field but with the equation $\ddot{x}(t) + \alpha\dot{x}(t) + \operatorname{proj}_{C(x(t))}(-\ddot{x}(t)) = 0$, where the second derivative with respect to time intervenes with the projection. Uniqueness could still be guaranteed under a one-sided Lipschitz condition, i.e.,

$$\langle \omega_1 - \omega_2, F_1 - F_2\rangle_{\mathcal{H}\times\mathcal{H}} \le C\|\omega_1 - \omega_2\|^2_{\mathcal{H}\times\mathcal{H}},$$
$$\text{for all } \omega_1, \omega_2 \in \mathcal{H} \times \mathcal{H}, \text{ and } F_1 \in F(\omega_1),\ F_2 \in F(\omega_2),$$

with $C > 0$ (see, e.g., [17, Theorem 10.4]). Due to the implicit structure it is hard to see whether this inequality can be derived. For this reason the uniqueness of solutions to (CP) remains an open problem for now.

4 Asymptotic Analysis of Trajectories of (IMOG')

In this section, we omit the assumption $\dim(\mathcal{H}) < +\infty$. We show that trajectories of the differential equation (IMOG') converge weakly to Pareto critical points of the optimization problem (MOP). This follows from a dissipative property of the system and an argument that relies on Opial's Lemma. We first define an energy function for the system (IMOG') that has Lyapunov-type properties.

Proposition 4.1 *Let $x : [0, +\infty) \to \mathcal{H}$ be a solution to (CP). For $i = 1, \ldots, m$ define the global energy*

$$\mathcal{E}_i : [0, T) \to \mathbb{R}, \quad t \mapsto f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2.$$

Then, for all $t \in (0, +\infty)$ it holds that $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_i(t) \leq -\alpha\|\dot{x}(t)\|^2$.

Proof From the definition of the differential equation (IMOG'), it follows that $-\alpha\dot{x}(t) = \mathrm{proj}_{C(x(t))+\ddot{x}(t)}(0)$, where $C(x) = \mathrm{conv}\,(\{\nabla f_i(x) \ : \ i = 1, \ldots, m\})$, and the addition $C(x(t)) + \ddot{x}(t)$ has to be understood elementwise. By the variational characterization of the convex projection, we get for all $i = 1, \ldots, m$

$$\langle \alpha\dot{x}(t) + \nabla f_i(x(t)) + \ddot{x}(t), \alpha\dot{x}(t)\rangle \leq 0,$$

which immediately gives

$$\langle \nabla f_i(x(t)), \dot{x}(t)\rangle + \langle \dot{x}(t), \ddot{x}(t)\rangle \leq -\alpha\|\dot{x}(t)\|^2.$$

Applying the chain rule to $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_i(t)$ yields the desired result. □

Proposition 4.2 *Let $x : [0, +\infty) \to \mathcal{H}$ be a bounded solution of (CP) and let further ∇f_i be Lipschitz continuous on bounded sets. Then, for all $i = 1, \ldots, m$ it holds that*

i) $\lim_{t\to+\infty} \mathcal{E}_i(t) = \mathcal{E}_i^\infty > -\infty$.
ii) $\dot{x} \in L^2([0, +\infty)) \cap L^\infty([0, +\infty))$.
iii) $\ddot{x} \in L^\infty([0, +\infty))$, $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$ *and* $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{E}_i^\infty$.
iv) There exists a monotonically increasing unbounded sequence $(t_k)_{k\geq 0}$ *with* $\mathrm{proj}_{C(x(t_k))}(0) \to 0$ *for* $k \to +\infty$.

Proof *i)* From Proposition 4.1, we immediately get that $\mathcal{E}_i$ is monotonically decreasing and therefore $\mathcal{E}_i(t) \to \mathcal{E}_i^\infty$ as $t \to +\infty$. We have to show that in fact $\mathcal{E}_i^\infty > -\infty$. Since ∇f_i is bounded on bounded sets, we can conclude by the mean value theorem

that f_i is bounded on bounded sets. Since $x(t)$ remains bounded by assumption, we conclude that $f_i(x(t))$ is bounded from below, and hence

$$\mathcal{E}_i^\infty \geq \inf_{t\geq 0} f_i(x(t)) > -\infty.$$

ii) We know that $f_i(x(t))$ is bounded. Then, by the definition of $\mathcal{E}_i$ and the fact that $\mathcal{E}_i$ is monotonically decreasing, we immediately get that $\dot{x}$ is bounded for all $t \geq 0$. Since $\dot{x}$ is continuous, it follows that $\dot{x} \in L^\infty([0, +\infty))$. Using Proposition 4.1 it follows that

$$\begin{aligned}\alpha \int_0^{+\infty} \|\dot{x}(t)\|^2 \,\mathrm{d}t &\leq -\int_0^{+\infty} \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_i(s)\,\mathrm{d}s \\ &= \mathcal{E}_i(0) - \mathcal{E}_i^\infty < +\infty,\end{aligned}$$

and therefore $\dot{x} \in L^2([0, +\infty))$.
iii) Since $\dot{x}(t)$ and $\nabla f_i(x(t))$ remain bounded for all $t \geq 0$ it follows that $\ddot{x}(t) = -\alpha\dot{x}(t) - \mathrm{proj}_{C(x(t))}(-\ddot{x}(t))$ remains bounded for all $t \geq 0$. By the fact that $\dot{x}$ is absolutely continuous (on bounded intervals), it follows that $\ddot{x}$ is measurable and hence $\ddot{x} \in L^\infty([0, +\infty))$. Then, from $\dot{x} \in L^2([0, +\infty))$ together with the absolute continuity of $\dot{x}$ and $\ddot{x} \in L^\infty([0, +\infty))$ it follows that $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$. From $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$ and part i) we can immediately conclude $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{E}_i^\infty$.
iv) Assume that the negation of statement *iv)* holds, namely that there exists $M > 0$ and $T > 0$ such that

$$\left\| \underset{C(x(t))}{\mathrm{proj}}\ (0) \right\| \geq 2M, \text{ for all } t \geq T. \tag{6}$$

Fix an arbitrary $\delta > 0$ independent of M and T. Since $\dot{x}(t) \to 0$ and ∇f_i is Lipschitz continuous on a set containing $\{x(t) : t \geq 0\}$ it follows that there exists $T_\delta > T$ such that for all $t > T_\delta$ it holds that

$$\|\nabla f_i(x(s)) - \nabla f_i(x(t))\| < \frac{M}{2} \text{ and } \|\alpha\dot{x}(s)\| < \frac{M}{2} \text{ for all } s \in [t, t+\delta]. \tag{7}$$

Fix an arbitrary $t > T_\delta$. Define $v{:=}\mathrm{proj}_{C(x(t))}/\|\mathrm{proj}_{C(x(t))}\|$. From (6) it follows that

$$\langle \xi, v\rangle \geq 2M \text{ for all } \xi \in C(x(t)).$$

Combining the last statement with (7) and using the Cauchy–Schwarz inequality, we get

$$\langle \xi + \alpha\dot{x}(s), v\rangle \geq M \text{ for all } s \in [t, t+\delta] \text{ and all } \xi \in C(x(s)).$$

And hence

$$\langle -\ddot{x}(s), v\rangle \geq M \text{ for almost all } s \in [t, t+\delta).$$

Using the Cauchy–Schwarz inequality again, we get

$$\begin{aligned}\|\dot{x}(t) - \dot{x}(t+\delta)\| &\geq \langle \dot{x}(t) - \dot{x}(t+\delta), v\rangle = \\ &= \int_t^{t+\delta} \langle -\ddot{x}(s), v\rangle \,\mathrm{d}s \geq \int_t^{t+\delta} M \,\mathrm{d}s = M\delta.\end{aligned}$$

Since we can choose an arbitrary large δ independently from M, this contradicts $\dot{x}(t) \to 0$. □

We will use part *iv)* of Proposition 4.2 to show that a weak limit point of the trajectory $x(t)$ is Pareto critical. To this end, we introduce the following lemma that states a demiclosedness property of the set-valued map

$$C : \mathcal{H} \rightrightarrows \mathcal{H},\ x \mapsto C(x) := \operatorname{conv}\left(\{\nabla f_i(x) \,:\, i = 1, \ldots, m\}\right).$$

Lemma 4.3 *Assume that the objective functions f_i are continuously differentiable. Let $(x^k)_{k\geq 0}$ be a sequence in $\mathcal{H}$ that converges weakly to x^∞, and assume there exists a sequence $(g^k)_{k\geq 0}$ with $g^k \in C(x^k)$ that converges strongly to zero. Then, $0 \in C(x^\infty)$ and hence x^∞ is Pareto critical.*

Proof A proof can be found in Lemma 2.4 in [11] and in Lemma 4.10 in [7]. □

If we can show that the trajectories of (IMOG') converge weakly, Proposition 4.2 together with Lemma 4.3 guarantees that the limit points are Pareto critical. To show that the trajectories are in fact converging, we require Opial's Lemma [26].

Lemma 4.4 (Opial's Lemma) *Let $S \subset \mathcal{H}$ be a nonempty subset of $\mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$. Assume that $x(t)$ satisfies the following conditions.*

i) Every weak sequential cluster point of $x(t)$ belongs to S.
ii) For every $z \in S$, $\lim_{t\to+\infty}\|x(t) - z\|$ exists.

Then, $x(t)$ converges weakly to an element $x^\infty \in S$.

To use Opial's Lemma, we need a suitable nonempty set $S \subset \mathcal{H}$ that we define in the following proposition.

Proposition 4.5 *Let $x(t)$ be a bounded solution to (CP). Then, the set*

$$S := \left\{z \in \mathcal{H} \,:\, f_i(z) \leq \mathcal{E}_i^\infty \text{ for all } i = 1, \ldots, m, \right\}, \tag{8}$$

is nonempty.

Proof Part *iii)* of Proposition 4.2 states that $\lim_{t\to+\infty} f_i(x(t)) = \mathcal{E}_i^\infty$ for all $i = 1, \ldots, m$. Since $x(t)$ is bounded, it possesses at least one weak sequential cluster point x^∞. The objective functions f_i are convex and continuous and therefore weakly lower semicontinuous. From this we conclude $x^\infty \in S$. □

For the set S defined in (8) and a bounded solution $x(t)$ of (CP), the first part of Opial's Lemma is easy to obtain. It follows analogously to the proof of Proposition 4.5 where it is shown that S is nonempty. To show the second part of Opial's Lemma, we verify that $h_z(t):=\frac{1}{2}\|x(t)-z\|^2$ satisfies a differential inequality. Then, the convergence can be deduced from the following lemma.

Lemma 4.6 ([10] Lemma 4.2) *Let $h \in C^1([0,+\infty),\mathbb{R})$ be a positive function satisfying $\alpha\dot{h}(t)+\ddot{h}(t) \leq g(t)$ for all $t \geq 0$, with $g \in L^1([0,+\infty),\mathbb{R})$ and $\alpha > 0$. Then, $\lim_{t\to+\infty} h(t)$ exists.*

With these ingredients, we can formulate the main convergence theorem of this section.

Theorem 4.7 *Assume that the objective functions f_i are convex with gradients ∇f_i that are Lipschitz continuous on bounded sets. Then, every bounded solution $x : [0,+\infty) \to \mathcal{H}$ of (CP) with arbitrary initial conditions $x^0, v^0 \in \mathcal{H}$ converges weakly to a Pareto critical point of (MOP).*

Proof For $z \in S$ define

$$h_z(t):=\frac{1}{2}\|x(t)-z\|^2.$$

Using the chain rule, we compute the first and the second derivative of $h_z(t)$ as

$$\dot{h}_z(t) = \langle x(t)-z, \dot{x}(t)\rangle \text{ and } \ddot{h}_z(t) = \langle x(t)-z, \ddot{x}(t)\rangle + \|\dot{x}(t)\|^2.$$

For a fixed $t \in (0,+\infty)$, write

$$\alpha\dot{h}_z(t)+\ddot{h}_z(t) = \langle \ddot{x}(t)+\alpha\dot{x}(t), x(t)-z\rangle + \|\dot{x}(t)\|^2.$$

Using the definition of (IMOG'), we can write $\ddot{x}(t)+\alpha\dot{x}(t) = -\sum_{i=1}^m \theta_i \nabla f_i(x(t))$ for some weights $\theta \in \Delta^m$. Then, we write

$$\alpha\dot{h}_z(t)+\ddot{h}_z(t) = \sum_{i=1}^m \theta_i \langle \nabla f_i(x(t)), z-x(t)\rangle + \|\dot{x}(t)\|^2. \tag{9}$$

Proposition 4.1 gives for all $i = 1,\ldots,m$

$$\mathcal{E}_i(t) = f_i(x(t)) + \frac{1}{2}\|\dot{x}(t)\|^2 \geq \mathcal{E}_i^\infty \geq f_i(z) \geq f_i(x(t)) + \langle \nabla f_i(x(t)), z-x(t)\rangle,$$

and therefore

$$\sum_{i=1}^m \theta_i \langle \nabla f_i(x(t)), z-x(t)\rangle \leq \frac{1}{2}\|\dot{x}(t)\|^2. \tag{10}$$

Combining inequalities (9) and (10) we get

$$\alpha \dot{h}_z(t) + \ddot{h}_z(t) \leq \frac{3}{2}\|\dot{x}(t)\|^2.$$

By Proposition 4.2, we know $\|\dot{x}(\cdot)\|^2 \in L^1([0, +\infty))$. Then, Lemma 4.6 guarantees that $\lim_{t\to+\infty} h_z(t)$ exists. In addition, we know that every weak sequential cluster point of $x(t)$ belongs to S by the weak lower semicontinuity of the objective functions f_i. Then, we can use Opial's Lemma 5.5 to prove that $x(t)$ converges weakly to an element in S. Let x^∞ be the weak limit of $x(t)$. Then, by Proposition 4.2, there exists a monotonically increasing unbounded sequence $(t_k)_{k\geq 0}$ with $\text{proj}_{C(x(t_k))}(0) \to 0$ (strongly in $\mathcal{H}$) as $k \to +\infty$. Since $x(t_k)$ converges weakly to x^∞, Lemma 4.3 states that x^∞ is Pareto critical. □

5 An Inertial Multiobjective Gradient Algorithm

In this section, we derive an inertial first-order method for multiobjective optimization problems from an explicit discretization of the differential equation (IMOG'). We write the system (IMOG') in the equivalent form

$$\alpha \dot{x}(t) + \underset{C(x(t))+\ddot{x}(t)}{\text{proj}}(0) = 0,$$

with $C(x) = \text{conv}(\{\nabla f_i(x) : i = 1, \ldots, m\})$, and use the following discretization of the differential equation

$$\begin{aligned}
&\alpha \frac{x^{k+1} - x^k}{h} + \underset{C(x^k)+\frac{x^{k+1}-2x^k+x^{k-1}}{h^2}}{\text{proj}}(0) = 0,\\
&\alpha h(x^{k+1} - x^k) + \underset{h^2C(x^k)+x^{k+1}-2x^k+x^{k-1}}{\text{proj}}(0) = 0.
\end{aligned}$$

Lemma A.2 states that x^{k+1} is uniquely defined as

$$\begin{aligned}
x^{k+1} &= -\left(\frac{1}{1+\alpha h} \underset{h^2C(x^k)-2x^k+x^{k-1}}{\text{proj}}(-x^k) - \frac{\alpha h}{1+\alpha h}x^k\right)\\
&= -\left(\frac{1}{1+\alpha h}\left[-x^k + \underset{h^2C(x^k)-x^k+x^{k-1}}{\text{proj}}(0)\right] - \frac{\alpha h}{1+\alpha h}x^k\right)\\
&= x^k - \frac{1}{1+\alpha h} \underset{h^2C(x^k)-x^k+x^{k-1}}{\text{proj}}(0).
\end{aligned} \tag{11}$$

Therefore, x^{k+1} can be written as

$$x^{k+1} = x^k + \frac{1}{1+\alpha h}(x^k - x^{k-1}) - \frac{h^2}{1+\alpha h}\sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k), \tag{12}$$

where $\theta^k \in \Delta^m$ is the solution to the quadratic optimization problem

$$\min_{\theta\in\mathbb{R}^m} \left\| h^2\left(\sum_{i=1}^{m}\theta_i \nabla f_i(x^k)\right) - (x^k - x^{k-1})\right\|^2 \quad \text{s.t. } \theta \geq 0 \text{ and } \sum_{i=1}^{m}\theta_i = 1. \tag{13}$$

The objective function of the problem in (13) can be rewritten into

$$\left\|\sum_{i=1}^{m}\theta_i\left(h^2\nabla f_i(x^k) - (x^k - x^{k-1})\right)\right\|^2.$$

Therefore, solving problem (13) is as difficult as solving the optimization problem required in the classical multiobjective steepest descent method [19]. The problem is a quadratic optimization problem with linear constraints. The dimension m of the problem is usually small since in most application we do not consider many objective functions. In the following subsection, we analyze the asymptotic behavior of the sequence $(x^k)_{k\geq 0}$ that is defined by equations (12) and (13).

5.1 Asymptotic Analysis

The asymptotic analysis of the sequence $(x^k)_{k\geq 0}$ defined by (12) and (13) works surprisingly similar to the asymptotic analysis of the trajectories $x(t)$ of the differential equation (IMOG'). We start by proving that the sequence $(x^k)_{k\geq 0}$ satisfies a dissipative property. To this end, we introduce the following preparatory lemma.

Lemma 5.1 *Let $(x^k)_{k\geq 0}$ be defined by (12) and (13) with $x^0 = x^1 \in \mathcal{H}$ and $\alpha, h > 0$. Then, for all $i = 1, \ldots, m$ it holds that*

$$\langle \nabla f_i(x^k), x^{k+1} - x^k\rangle \leq -\frac{\alpha}{h}\|x^{k+1} - x^k\|^2 + \frac{1}{2h^2}\left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right].$$

Proof Using the variational characterization of the convex projection in the identity (11), we get for all $i = 1, \ldots, m$,

$$\langle \alpha h(x^{k+1} - x^k) + h^2\nabla f_i(x^k) + (x^{k+1} - x^k) - (x^k - x^{k-1}), \alpha h(x^{k+1} - x^k)\rangle \leq 0,$$

which can be rearranged into

$$\begin{aligned}
&\langle \nabla f_i(x^k), x^{k+1} - x^k \rangle \\
&\quad \leq -\left(\frac{1}{h^2} + \frac{\alpha}{h}\right) \|x^{k+1} - x^k\|^2 + \frac{1}{h^2} \langle x^{k+1} - x^k, x^k - x^{k-1} \rangle \\
&\quad \leq -\frac{\alpha}{h} \|x^{k+1} - x^k\|^2 + \frac{1}{2h^2} \left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2 \right].
\end{aligned}$$

□

Using Lemma 5.1, we show that there exists an energy sequence which can be seen as a discretization of the energy function defined in Proposition 4.2.

Proposition 5.2 *Assume that the gradients ∇f_i of the objective functions are globally L-Lipschitz continuous for all $i = 1, \ldots, m$ and further assume $Lh < 2\alpha$. Then*

$$\mathcal{E}_{i,k} := f_i(x^k) + \frac{1}{2h^2} \|x^k - x^{k-1}\|^2,$$

is monotonically decreasing.

Proof We start with investigating the difference

$$\mathcal{E}_{i,k+1} - \mathcal{E}_{i,k} = f_i(x^{k+1}) - f_i(x^k) + \frac{1}{2h^2} \left[\|x^{k+1} - x^k\|^2 - \|x^k - x^{k-1}\|^2 \right].$$

Using that f_i is convex with L-Lipschitz continuous gradient, we estimate the expression above by

$$\leq \langle \nabla f_i(x^k), x^{k+1} - x^k \rangle + \frac{L}{2} \|x^{k+1} - x^k\|^2 + \frac{1}{2h^2} \left[\|x^{k+1} - x^k\|^2 - \|x^k - x^{k-1}\|^2 \right].$$

Using Lemma 5.1, we estimate this term by

$$\leq \left(\frac{L}{2} - \frac{\alpha}{h}\right) \|x^{k+1} - x^k\|^2 - \frac{1}{2h^2} \|x^{k+1} - 2x^k + x^{k-1}\|^2.$$

For $hL < 2\alpha$ it holds that $\left(\frac{L}{2} - \frac{\alpha}{h}\right) < 0$ and we get

$$\mathcal{E}_{i,k+1} - \mathcal{E}_{i,k} \leq \left(\frac{L}{2} - \frac{\alpha}{h}\right) \|x^{k+1} - x^k\|^2 - \frac{1}{2h^2} \|x^{k+1} - 2x^k + x^{k-1}\|^2, \tag{14}$$

which completes the proof. □

The following corollary is an immediate consequence of Proposition 5.2.

Corollary 5.3 *Assume all conditions of Proposition 5.2 are met. Then, for all $i = 1, \ldots, m$ and all $k \geq 1$ it holds that $f_i(x^k) \leq f_i(x^0)$.*

Corollary 5.3 hints at a condition that guarantees that the sequence $(x^k)_{k\geq 0}$ remains bounded. If the level set $\mathcal{L}_i(f_i(x^0)):=\{x \in \mathcal{H} : f_i(x) \leq f_i(x^0)\}$ of one objective function f_i is bounded, the sequence $(x^k)_{k\geq 0}$ remains bounded. In the following proposition we collect some immediate consequences of Proposition 5.2.

Proposition 5.4 *Assume the gradients ∇f_i of the objective functions are L-Lipschitz continuous on a bounded set containing the sequence $(x^k)_{k\geq 0}$, that is defined by equations (12) and (13). Assume $Lh < 2\alpha$, then for all $i = 1, \ldots, m$ the following statements hold.*

i) $\mathcal{E}_{i,k} \to \mathcal{E}_i^\infty > -\infty$ *as* $k \to +\infty$
ii) $\sum_{k=0}^\infty \|x^{k+1} - x^k\|^2 < +\infty$
iii) $f_i(x^k) \to \mathcal{E}_i^\infty$ *as* $k \to +\infty$

Proof *i)* Proposition 5.2 states that $\mathcal{E}_{i,k}$ is monotonically decreasing. Therefore, $\mathcal{E}_{i,k} \to \mathcal{E}_i^\infty$ holds. We have to show that $\mathcal{E}_i^\infty > -\infty$. Since the objective functions f_i have Lipschitz continuous gradients on a bounded set containing $(x^k)_{k\geq 0}$, it follows by the mean value theorem that f_i is bounded on this sets and in particular on $(x^k)_{k\geq 0}$. Therefore, we conclude

$$\mathcal{E}_i^\infty = \lim_{k\to+\infty} f_i(x^k) + \frac{1}{2h^2}\|x^k - x^{k-1}\|^2 \geq \liminf_{k\to+\infty} f_i(x^k) > -\infty.$$

ii) From inequality (14) we immediately follow

$$\begin{aligned}\mathcal{E}_{i,K+1} - \mathcal{E}_{i,1} &= \sum_{k=1}^K \left(\mathcal{E}_{i,k+1} - \mathcal{E}_{i,k}\right)\\ &\leq \sum_{k=1}^K \left(\frac{L}{2} - \frac{\alpha}{h}\right)\|x^{k+1} - x^k\|^2 - \frac{1}{h^2}\sum_{k=1}^K \|x^{k+1} - 2x^k + x^{k-1}\|^2.\end{aligned}$$

Since $Lh < 2\alpha$, it holds that $\left(\frac{\alpha}{h} - \frac{L}{2}\right) > 0$ and therefore we get for all $K \geq 1$

$$\left(\frac{\alpha}{h} - \frac{L}{2}\right)\sum_{k=1}^K \|x^{k+1} - x^k\|^2 \leq \mathcal{E}_{i,1} - \mathcal{E}_{i,K+1}.$$

From part *i)*, we know that the right hand side converges which completes the proof of *ii)*.
iii) Since $\mathcal{E}_{i,k} \to \mathcal{E}_i^\infty$ and $\|x^{k+1} - x^k\|^2 \to 0$, it follows that $f_i(x^k) \to \mathcal{E}_i^\infty$. □

We use the following discrete version of Opial's Lemma to prove that $(x^k)_{k\geq 0}$ converges weakly to a Pareto critical point of (MOP).

Lemma 5.5 (Opial's Lemma) *Let $S \subset \mathcal{H}$ be nonempty and let $(x^k)_{k\geq 0}$ be a sequence in $\mathcal{H}$ that satisfies the following conditions.*

i) *For all $z \in S$ $\lim_{k\to+\infty}\|x^k - z\|$ exists.*

ii) Every weak sequential cluster point of $(x^k)_{k\geq 0}$ *belongs to* S.

Then, it follows that $(x^k)_{k\geq 0}$ *converges weakly to an element in* S.

We will use Opial's Lemma on the set

$$S := \left\{ z \in \mathcal{H} \,:\, f_i(z) \leq \mathcal{E}_i^{\infty} \text{ for all } i = 1, \ldots, m, \right\}. \tag{15}$$

Theorem 5.6 *Assume the gradients* ∇f_i *of the objective functions are* L*-Lipschitz continuous on a bounded set containing the sequence* $(x^k)_{k\geq 0}$*, defined by (12) and (13) and further assume* $Lh < 2\alpha$*. Then,* $(x^k)_{k\geq 0}$ *converges weakly to a Pareto critical point of (MOP).*

Proof We show that $(x^k)_{k\geq 0}$ satisfies Opial's Lemma for the set S defined by (15). We start by showing a quasi Fejér property of the sequence $(x^k)_{k\geq 0}$. For a fixed $z \in S$, define the sequence

$$h_k := \frac{1}{2}\|x^k - z\|^2.$$

It is easy to check that

$$h_{k+1} = h_k + \langle x^{k+1} - x^k, x^k - z\rangle + \frac{1}{2}\|x^{k+1} - x^k\|^2.$$

Proposition 5.4 guarantees the monotonicity of $\mathcal{E}_{i,k}$. Since $z \in S$, from the convexity of f_i we can deduce for all $i = 1, \ldots, m$ that

$$\mathcal{E}_{i,k} = f_i(x^k) + \frac{1}{2h^2}\|x^k - x^{k-1}\|^2 \geq \mathcal{E}_i^{\infty} \geq f_i(z) \geq f_i(x^k) + \langle \nabla f_i(x^k), z - x^k\rangle,$$

and therefore

$$\left\langle \sum_{i=1}^{m} \theta_i \nabla f_i(x^k), z - x^k \right\rangle \leq \frac{1}{2h^2}\|x^k - x^{k-1}\|^2.$$

Using this inequality we can show

$$\begin{aligned}
\frac{h^2}{1+\alpha h}\left\langle \sum_{i=1}^{m} \theta_i^k \nabla f_i(x^k), z - x^k \right\rangle &= \left\langle x^k - x^{k+1} - \frac{1}{1+\alpha h}(x^k - x^{k-1}), z - x^k \right\rangle \\
&= \langle x^{k+1} - x^k, x^k - z\rangle - \frac{1}{1+\alpha h}\langle x^{k-1} - x^k, x^k - z\rangle \leq \frac{1}{2(1+\alpha h)}\|x^k - x^{k-1}\|^2,
\end{aligned}$$

which leads to the inequality

$$\langle x^{k+1} - x^k, x^k - z\rangle \leq \frac{1}{1+\alpha h}\langle x^{k-1} - x^k, x^k - z\rangle + \frac{1}{2(1+\alpha h)}\|x^k - x^{k-1}\|^2.$$

We use this inequality to show

$$\begin{aligned}
h_{k+1} - h_k &= \langle x^{k+1} - x^k, x^k - z\rangle + \frac{1}{2}\|x^{k+1} - x^k\|^2 \\
&\leq \frac{1}{1+\alpha h}\langle x^{k-1} - x^k, x^k - z\rangle + \frac{1}{2}\|x^{k+1} - x^k\|^2 + \frac{1}{2(1+\alpha h)}\|x^k - x^{k-1}\|^2 \\
&= \frac{1}{1+\alpha h}\left[h_k - h_{k-1} + \frac{1}{2}\|x^k - x^{k-1}\|^2\right] + \frac{1}{2}\|x^{k+1} - x^k\|^2 \\
&\quad + \frac{1}{2(1+\alpha h)}\|x^k - x^{k-1}\|^2 \\
&\leq \frac{1}{1+\alpha h}(h_k - h_{k-1}) + \frac{1}{2}\|x^{k+1} - x^k\|^2 + \frac{1}{1+\alpha h}\|x^k - x^{k-1}\|^2.
\end{aligned}$$

Defining $\theta_k := h_{k+1} - h_k$, $\delta_k := \frac{1}{1+\alpha h}\|x^k - x^{k-1}\|^2 + \frac{1}{2}\|x^{k+1} - x^k\|^2$ and $a := \frac{1}{1+\alpha h}$, we can therefore conclude

$$\theta_{k+1} \leq a\theta_k + \delta_k.$$

Proposition 5.4 states that $\sum_{k=1}^{\infty} \delta_k < +\infty$. Therefore, we can use Theorem 2.1 in [2] or Theorem 3.1 in [1] to show that h_k converges. To use Opial's Lemma, we also have to show that all weak sequential cluster points of $(x^k)_{k\geq 0}$ belong to S. Since the sequence $(x^k)_{k\geq 0}$ is bounded, it possesses at least one sequential cluster point that we denote by x^∞ and a subsequence $(x_{k_l})_{l\geq 0}$ that converges weakly to x^∞. Since f_i is convex and continuous, it is also weakly lower semicontinuous and it follows that for all $i = 1, \ldots, m$

$$f_i(x^\infty) \leq \liminf_{l\to+\infty} f_i(x^{k_l}) = \lim_{k\to+\infty} f_i(x^k) = \mathcal{E}_i^\infty,$$

where the equality follows from the fact that the limit exists. Therefore, $x^\infty \in S$ and hence S is nonempty. Then, Opial's Lemma 5.5 states that $(x^k)_{k\geq 0}$ converges weakly to an element in S that we denote by x^∞. We will show that each weak sequential cluster point of $(x^k)_{k\geq 0}$ is Pareto critical. By the definition of the sequence $(x^k)_{k\geq 0}$ in (12), it holds that

$$\sum_{k=1}^{\infty}\left\|\sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k)\right\|^2 = \sum_{k=1}^{\infty}\left\|\frac{1+\alpha h}{h^2}(x^{k+1} - x^k) + \frac{1}{h^2}(x^k - x^{k-1})\right\|^2.$$

This sum is finite by part ii) of Proposition 5.4. Thus, we know that the sequence $g^k := \sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k) \in \operatorname{conv}(\nabla f_i(x^k))$ converges strongly to zero. Since x^k converges weakly to x^∞, Lemma 4.3 states that $0 \in C(x^\infty)$ and hence x^∞ is Pareto critical. □

6 An Accelerated Multiobjective Gradient Method

In this section, we define a multiobjective gradient method with Nesterov acceleration based on the inertial method we discussed in the previous subsection.

6.1 The Single-objective Case

In this subsection we present Nesterov's method in the single-objective setting and point out its relation to an inertial gradient-like dynamical system with asymptotically vanishing damping. Consider the problem

$$\min_{x\in\mathcal{H}} f(x),$$

where $f:\mathcal{H}\to\mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient $\nabla f(x)$. For $\alpha\geq 3$, $0<s\leq\frac{1}{L}$ and $x^0, x^1\in\mathcal{H}$, define the sequence $(x^k)_{k\geq 0}$ by

$$\left.\begin{aligned} y^k &= x^k+\tfrac{k-1}{k+\alpha-1}(x^k-x^{k-1}),\\ x^{k+1} &= y^k - s\nabla f(y^k)\end{aligned}\right\}\ \text{for } k\geq 1. \tag{16}$$

If $\arg\min f\neq\emptyset$, it can be shown that $f(x^k)-\min_{x\in\mathcal{H}} f(x)=\mathcal{O}(k^{-2})$ and that $\|x^{k+1}-x^k\|=\mathcal{O}(k^{-1})$. For $\alpha>3$, it holds that $f(x^k)-\min_{x\in\mathcal{H}} f(x)=o(k^{-2})$, $\|x^{k+1}-x^k\|=o(k^{-1})$ and that $(x^k)_{k\geq 0}$ converges weakly to an element in $\arg\min f$ [11]. Nesterov's method is related to the following gradient system with asymptotically vanishing damping

$$\ddot{x}(t)+\frac{\alpha}{t}\dot{x}(t)+\nabla f(x(t))=0. \tag{17}$$

The algorithm (16) can be derived as a discretization of (17). This relation is further investigated in [6, 31].

6.2 Introducing Nesterov Acceleration in (IMOG')

We formally define the following gradient-like system with asymptotically vanishing damping for multiobjective optimization.

$$\ddot{x}(t)+\frac{\alpha}{t}\dot{x}(t)+\operatorname*{proj}_{C(x(t))}(-\ddot{x}(t))=0, \tag{18}$$

with $\alpha\geq 3$ and $C(x)=\operatorname{conv}(\{\nabla f_i(x) : i=1,\dots,m\})$. We give a full discussion of the system (18) in [30]. It can be shown that for $\alpha\geq 3$ the function values converge with rate $\mathcal{O}(t^{-2})$ to an optimal value measured with the merit function (1). For $\alpha>3$ the trajectories converge weakly to weakly Pareto optimal solutions. This is in line with the results for the single-objective system (17). We restrict the analysis of the discrete method in this paper to the case $\alpha=3$. We show that an implicit discretization of

this system leads to an accelerated multiobjective gradient method with an improved convergence rate of the function values. We equivalently write (18) as

$$\frac{3}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0.$$

Using the same Ansatz as in Section 2 of [31], we show that we can derive the differential equation (18) from the scheme

$$\frac{3}{k}(x^{k+1} - x^k) + \operatorname*{proj}_{sC(y^k)+(x^{k+1}-2x^k+x^{k-1})}(0) = 0, \tag{19}$$

with $y^k = x^k + \frac{k-1}{k+2}(x^k - x^{k-1})$. We divide (19) by $\sqrt{s}$ to get

$$\frac{3}{k}\frac{x^{k+1} - x^k}{\sqrt{s}} + \operatorname*{proj}_{\sqrt{s}C(y^k)+\frac{x^{k+1}-2x^k+x^{k-1}}{\sqrt{s}}}(0) = 0. \tag{20}$$

We use the Ansatz $x^k \approx x(k\sqrt{s})$ for some smooth curve $x(t)$ defined for all $t \geq 0$. Write $k = \frac{t}{\sqrt{s}}$. When the step size s goes to zero $X(t) \approx x_{\frac{t}{\sqrt{s}}} = x^k$ and $X(t) \approx x_{\frac{t+\sqrt{s}}{\sqrt{s}}} = x_{k+1}$. Then, Taylor expansion gives

$$\frac{x^{k+1} - x^k}{\sqrt{s}} = \dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s}), \quad \frac{x^k - x^{k-1}}{\sqrt{s}} = \dot{x}(t) - \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s}), \tag{21}$$

and hence

$$\frac{x^{k+1} - 2x^k + x^{k-1}}{\sqrt{s}} = \ddot{x}(t)\sqrt{s} + o(\sqrt{s}). \tag{22}$$

For all $i = 1, \ldots, m$, we have $\sqrt{s}\nabla f_i(y^k) = \sqrt{s}\nabla f_i(x(t)) + o(\sqrt{s})$. Since the convex projection depends in a well-behaved manner on the convex set we project onto, we get

$$\operatorname*{proj}_{\sqrt{s}C(y^k)+\frac{x^{k+1}-2x^k+x^{k-1}}{\sqrt{s}}}(0) = \sqrt{s}\operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) + o(\sqrt{s}). \tag{23}$$

Combining (21), (22) and (23), we get from (20)

$$\frac{3\sqrt{s}}{t}\left(\dot{x}(t) + \frac{1}{2}\ddot{x}(t)\sqrt{s} + o(\sqrt{s})\right) + \sqrt{s}\operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) + o(\sqrt{s}) = 0.$$

Comparing the coefficients of $\sqrt{s}$, we obtain

$$\frac{3}{t}\dot{x}(t) + \underset{C(x(t))+\ddot{x}(t)}{\text{proj}}(0) = 0.$$

We have shown that the differential equation (18) can be derived from the scheme (19). Using Lemma A.1 on (19), we get that x^{k+1} is uniquely defined as

$$\begin{aligned} x^{k+1} &= -\left(\frac{k}{k+3}\underset{sC(y^k)-2x^k+x^{k-1}}{\text{proj}}(-x^k) - \frac{3}{k+3}x^k\right) \\ &= x^k - \frac{k}{k+3}\underset{sC(y^k)-(x^k-x^{k-1})}{\text{proj}}(0). \end{aligned}$$

The last term can be written as

$$x^k + \frac{k}{k+3}(x^k - x^{k-1}) - \frac{sk}{k+3}\sum_{i=1}^{m}\theta_i^k \nabla f_i(y^k),$$

where $\theta^k \in \mathbb{R}^m$ is a solution to the quadratic optimization problem

$$\min_{\theta\in\mathbb{R}^m}\left\| s\left(\sum_{i=1}^{m}\theta_i\nabla f_i(y^k)\right) - (x^k - x^{k-1})\right\|^2 \quad \text{s.t. } \theta \geq 0 \text{ and } \sum_{i=1}^{m}\theta_i = 1. \tag{24}$$

We want to drop the factor $\frac{k}{k+3}$ in front of the term $\sum_{i=1}^{m}\theta_i^k\nabla f_i(y^k)$ to get a method that more closely resembles (16). In addition, we perform a shift of the index k to transform $\frac{k}{k+3}$ into $\frac{k-1}{k+2}$. The final method we define in this subsection is given by the following scheme. Let $x^0 = x^1 \in \mathcal{H}$ and $s > 0$. Define the scheme

$$\left.\begin{aligned} y^k &= x^k + \tfrac{k-1}{k+2}(x^k - x^{k-1}), \\ x^{k+1} &= y^k - s\textstyle\sum_{i=1}^{m}\theta_i^k\nabla f_i(y^k) \end{aligned}\right\} \quad \text{for } k \geq 1, \tag{25}$$

where in each step $\theta^k \in \mathbb{R}^m$ is a solution to the quadratic optimization problem

$$\min_{\theta\in\mathbb{R}^m}\left\| s\left(\sum_{i=1}^{m}\theta_i\nabla f_i(y^k)\right) - \frac{k-1}{k+2}(x^k - x^{k-1})\right\|^2 \quad \text{s.t. } \theta \geq 0 \text{ and } \sum_{i=1}^{m}\theta_i = 1. \tag{26}$$

Similar to problem (13), the objective function of problem (26) can be rewritten into

$$\left\|\sum_{i=1}^{m}\theta_i\left(s\nabla f_i(x^k) - \frac{k-1}{k+2}(x^k - x^{k-1})\right)\right\|^2.$$

Hence, computing the step direction using (26) is as difficult as computing the step direction in the classical multiobjective steepest descent method [19]. In both cases we have to solve a low-dimensional quadratic optimization problem with linear constraints. The fact that we have to transform the quadratic optimization problem from (24) into (26) is an observation from the proof of Proposition 6.1. The presented method is still asymptotically equivalent to the scheme defined by (19). We summarize the defined method in Algorithm 1 for later references.

Algorithm 1 Accelerated multiobjective gradient method

Require: Choose $x^0 = x^1 \in \mathcal{H}$, $0 < s \leq \frac{1}{L}$ and set $k = 1$.
1: Set $y^k = x^k + \frac{k-1}{k+2}(x^k - x^{k-1})$.
2: Compute $\theta^k \in \mathbb{R}^m$ by solving

$$\min_{\theta \in \mathbb{R}^m} \left\| s\left(\sum_{i=1}^m \theta_i \nabla f_i(y^k)\right) - \frac{k-1}{k+2}(x^k - x^{k-1}) \right\|^2 \quad \text{s.t. } \theta \geq 0 \text{ and } \sum_{i=1}^m \theta_i = 1.$$

3: Set $x^{k+1} = y^k - s\sum_{i=1}^m \theta_i^k \nabla f_i(y^k)$
4: **if** stopping condition is true **then**
5: Stop.
6: **else**
7: Update $k \leftarrow k+1$ and go to step 1.
8: **end if**

6.3 A Dissipative Property

We start our investigations of Algorithm 1 with an energy estimate analogous to Proposition 5.2 for the inertial method.

Proposition 6.1 *Assume that the gradients ∇f_i of the objective functions are globally L-Lipschitz continuous for all $i = 1, \ldots, m$ and further assume $sL \leq 1$. Define for all $k \geq 1$ the energy sequence*

$$\mathcal{E}_{i,k} := f_i(x^k) + \frac{1}{2s}\|x^k - x^{k-1}\|^2.$$

For all $k \geq 1$, it holds that

$$\mathcal{E}_{i,k+1} - \mathcal{E}_{i,k} \leq -\frac{1}{2s}\frac{3}{k+2}\|x^k - x^{k-1}\|^2.$$

Proof From the definition of x^k and y^k in (25) we get

$$x^{k+1} - x^k + \operatorname*{proj}_{sC(y^k) - \frac{k-1}{k+2}(x^k - x^{k-1})}(0) = 0.$$

Hence, for all $i = 1, \ldots, m$ it holds that

$$\left\langle x^{k+1} - x^k + s\nabla f_i(y^k) - \frac{k-1}{k+2}(x^k - x^{k-1}), x^{k+1} - x^k \right\rangle \leq 0,$$

from which we follow

$$\begin{aligned} s\langle \nabla f_i(y^k), x^{k+1} - x^k\rangle &\leq -\|x^{k+1} - x^k\|^2 + \frac{k-1}{k+2}\langle x^{k+1} - x^k, x^k - x^{k-1}\rangle \\ &= -\frac{3}{k+2}\|x^{k+1} - x^k\| - \frac{1}{2}\frac{k-1}{k+2}\|x^{k+1} - 2x^k + x^{k-1}\|^2 \\ &\quad + \frac{1}{2}\frac{k-1}{k+2}\left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right]. \end{aligned}$$

Writing out the definition of y^k, one can easily verify that

$$\|x^{k+1} - y^k\|^2 \leq \frac{k-1}{k+2}\|x^{k+1} - 2x^k + x^{k-1}\|^2 + \frac{3}{k+2}\|x^{k+1} - x^k\|^2.$$

Combining the inequalities above and using $sL \leq 1$ we get

$$\begin{aligned} s(f_i(x^{k+1}) - f_i(x^k)) &\leq s\langle \nabla f_i(y^k), x^{k+1} - x^k\rangle + \frac{1}{2}\|x^{k+1} - y^k\|^2 \\ &\leq -\frac{1}{2}\frac{3}{k+2}\|x^{k+1} - x^k\|^2 + \frac{1}{2}\frac{k-1}{k+2}\left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right] \\ &= \frac{1}{2}\left[\|x^k - x^{k-1}\|^2 - \|x^{k+1} - x^k\|^2\right] - \frac{1}{2}\frac{3}{k+2}\|x^k - x^{k-1}\|^2, \end{aligned}$$

which completes the proof. □

Corollary 6.2 *Let $(x^k)_{k\geq 0}$ be a sequence defined by (25). Then, it holds that for all $k \geq 0$ and all $i = 1, \ldots, m$*

$$f_i(x^k) \leq f_i(x^0).$$

6.4 Convergence of Function Values with Rate $\mathcal{O}(k^{-2})$

The proof in this section relies on the proof by Fukuda, Tanabe and Yamashita [34] for their accelerated gradient method and the proof of Attouch and Peypouquet [11] for the single-objective case. The following definition is aligned with [35] and the concept of merit functions that gets introduced in [35] and further utilized in [33, 34]. For $z \in \mathcal{H}$ define

$$\sigma_k(z) := \min_{i=1,\ldots,m} f_i(x^k) - f_i(z).$$

Lemma 6.3 *It holds that*

$$\sigma_{k+1}(z) \leq -\frac{1}{s}\langle x^{k+1} - y^k, y^k - z\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2.$$

Proof The objective functions f_i are convex with L-Lipschitz continuous gradients. Therefore, for all $i = 1, \ldots, m$ it holds that

$$\begin{aligned} f_i(x^{k+1}) - f_i(z) &\leq f_i(x^{k+1}) - f_i(y^k) + f_i(y^k) - f_i(z) \\ &\leq \langle \nabla f_i(y^k), x^{k+1} - y^k\rangle + \frac{L}{2}\|x^{k+1} - y^k\|^2 + \langle \nabla f_i(y^k), y^k - z\rangle. \end{aligned} \tag{27}$$

The definition of $\sigma_k(z)$ gives

$$\sigma_{k+1}(z) = \min_{i=1,\ldots,m} f_i(x^{k+1}) - f_i(z) \leq \sum_{i=1}^{m} \theta_i^k \left(f_i(x^{k+1}) - f_i(z)\right). \tag{28}$$

Combining (27) and (28) and using $\sum_{i=1}^m \theta_i^k \nabla f_i(y^k) = \frac{1}{s}(y^k - x^{k+1})$ we get the desired inequality. □

We want to find a similar inequality for the expression $f_i(x^{k+1}) - f_i(x^k)$. To this end, we introduce the following lemma.

Lemma 6.4 *Define the optimization problem*

$$\begin{aligned} \min_{(v,\alpha)\in\mathcal{H}\times\mathbb{R}} \; & \Phi(v,\alpha) := \frac{1}{2}\|sv + (y^k - x^k)\|^2 + \alpha \\ \text{s.t. } & g_i(v,\alpha) := \langle s\nabla f_i(y^k) - (y^k - x^k), sv + (y^k - x^k)\rangle - \alpha \leq 0. \end{aligned} \tag{29}$$

Then, it holds that the dual problem to this problem is the quadratic problem (26). An optimal solution θ^* to (26) satisfies

$$\left\langle s\sum_{i=1}^{m} \theta_i^* \nabla f_i(y^k), x^{k+1} - x^k \right\rangle = \max_{i=1,\ldots,m} \langle s\nabla f_i(y^k), x^{k+1} - x^k\rangle.$$

Proof Since $\mathcal{H}$ is potentially infinite-dimensional, we need duality statements for infinite-dimensional constrained optimization problems. The statements we use in this proof can be found in Sections 8.3 to 8.6 of [22]. Since the optimization problem (29) has a fairly simple structure, we will not write out every result we use. The duality between (29) and (26) follows from a straightforward computation. Since the objective function $\Phi(v,\alpha)$ of (29) is convex and all constraints $g_i(v,\alpha)$ are linear, strong duality holds. Hence a KKT point $((v^*,\alpha^*),\theta^*) \in (\mathcal{H}\times\mathbb{R})\times\mathbb{R}^m$ of problem (29) yields a solution to (26). From the KKT conditions for (29) we derive

$$v^* = -s\sum_{i=1}^{m} \theta_i^* \nabla f_i(y^k).$$

For all $i = 1, \ldots, m$ it holds that $g_i(v, \alpha) \leq 0$ and hence

$$\langle s\nabla f_i(y^k) - (y^k - x^k), sv + (y^k - x^k)\rangle \leq \alpha.$$

By the complementarity of θ_i^* and $g_i(v^*, \alpha^*)$ we get

$$\left\langle s\sum_{i=1}^{m}\theta_i^*\nabla f_i(y^k) - (y^k - x^k), sv^* + (y^k - x^k)\right\rangle = \alpha^*$$
$$= \max_{i=1,\ldots,m}\langle s\nabla f_i(y^k) - (y^k - x^k), sv^* + (y^k - x^k)\rangle.$$

The second equality above follows from the fact that $\theta_j^* > 0$ holds for at least one $j \in \{1, \ldots, m\}$ as a consequence of the dual feasibility.
Using $v^* = -\sum_{i=1}^{m}\theta_i^*\nabla f_i(y^k)$, we get $sv^* = x^{k+1} - y^k$ and therefore

$$\left\langle s\sum_{i=1}^{m}\theta_i^*\nabla f_i(y^k) - (y^k - x^k), x^{k+1} - x^k\right\rangle$$
$$= \max_{i=1,\ldots,m}\langle s\nabla f_i(y^k) - (y^k - x^k), x^{k+1} - x^k)\rangle.$$

□

Lemma 6.5

$$\sigma_{k+1}(z) - \sigma_k(z) \leq -\frac{1}{s}\langle x^{k+1} - y^k, y^k - x^k\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2.$$

Proof For all $a, b \in \mathbb{R}^m$ it holds that

$$\left(\min_{i=1,\ldots,m} a_i\right) - \left(\min_{i=1,\ldots,m} b_i\right) \leq \max_{i=1,\ldots,m}(a_i - b_i)$$

and therefore for all $z \in \mathcal{H}$

$$\sigma_{k+1}(z) - \sigma_k(z) \leq \max_{i=1,\ldots,m}\left(f_i(x^{k+1}) - f_i(x^k)\right).$$

Using that the objective functions f_i are convex with L-Lipschitz continuous gradients and the fact that $sL \leq 1$, we can bound this expression by

$$\leq \max_{i=1,\ldots,m}\left(\langle\nabla f_i(y^k), x^{k+1} - x^k\rangle + \frac{1}{2s}\|x^{k+1} - y^k\|^2\right).$$

Now we use Lemma 6.4 and get the equality

$$= \sum_{i=1}^{m}\theta_i^k\langle\nabla f_i(y^k), x^{k+1} - x^k\rangle + \frac{1}{2s}\|x^{k+1} - y^k\|^2.$$

From here, we continue by using the definitions of x^k and y^k from (25) to get

$$= \frac{1}{s}\langle y^k - x^{k+1}, x^{k+1} - x^k\rangle + \frac{1}{2s}\|x^{k+1} - y^k\|^2$$
$$= -\frac{1}{s}\langle x^{k+1} - y^k, y^k - x^k\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2.$$

□

Theorem 6.6 *The sequence $(x^k)_{k\geq 0}$ defined by (25) satisfies*

$$\sigma_k(z) \leq \frac{2\left(\|x^1 - z\|^2 + \|x^2 - z\|^2\right)}{s(k+1)^2}.$$

Proof Lemma 6.3 and Lemma 6.5 state

$$\sigma_{k+1}(z) \leq -\frac{1}{s}\langle x^{k+1} - y^k, y^k - z\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2 \quad \text{and}$$
$$\sigma_{k+1}(z) - \sigma_k(z) \leq -\frac{1}{s}\langle x^{k+1} - y^k, y^k - x^k\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2.$$

Taking a convex combination of the last inequalities with weights $\frac{2}{k+2}$ and $\frac{k}{k+2}$ yields

$$\begin{aligned}
&\sigma_{k+1}(z) - \frac{k}{k+2}\sigma_k(z) \\
&\quad \leq -\frac{1}{s}\left\langle x^{k+1} - y^k, y^k - \frac{k}{k+2}x^k - \frac{2}{k+2}z\right\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2 \\
&\quad = \frac{1}{s}\left\langle x^{k+1} - y^k, \frac{k}{k+2}(x^k - y^k) + \frac{2}{k+2}(z - y^k)\right\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2.
\end{aligned} \tag{30}$$

Define

$$z^k := \frac{k+2}{2}y^k - \frac{k}{2}x^k = x^k + \frac{k-1}{2}(x^k - x^{k-1}), \tag{31}$$

and notice that

$$\frac{k}{k+2}(y^k - x^k) + \frac{2}{k+2}(y^k - z) = \frac{2}{k+2}(z^k - z). \tag{32}$$

Using the identity (32) in (30) we get

$$\sigma_{k+1}(z) \leq \frac{k}{k+2}\sigma_k(z) - \frac{2}{s(k+2)}\langle x^{k+1} - y^k, z^k - z\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2. \tag{33}$$

From the definition of z^k in (31), one can see that

$$z^{k+1} = z^k + \frac{k+2}{2}(x^{k+1} - y^k).$$

Using this identity, we can simply compute the squared norm of $\|z^{k+1} - z\|^2$ as

$$\|z^{k+1} - z\|^2 = \|z^k - z\|^2 + (k+2)\langle z^k - z, x^{k+1} - y^k\rangle + \left(\frac{k+2}{2}\right)^2 \|x^{k+1} - y^k\|^2.$$

Rearranging this identity and multiplying with $\frac{2}{s(k+2)^2}$ yields

$$\begin{aligned} &\frac{2}{s(k+2)^2}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right) \\ &\quad = -\frac{2}{s(k+2)}\langle z^k - z, x^{k+1} - y^k\rangle - \frac{1}{2s}\|x^{k+1} - y^k\|^2. \end{aligned} \tag{34}$$

Combining (33) and (34), in total we get

$$\sigma_{k+1}(z) \leq \frac{k}{k+2}\sigma_k(z) + \frac{4}{2s(k+2)^2}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right).$$

Multiplying both sides with $(k+2)^2$ then yields

$$(k+2)^2\sigma_{k+1}(z) \leq k(k+2)\sigma_k(z) + \frac{2}{s}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right).$$

Using $k(k+2) \leq (k+1)^2$ we get

$$(k+2)^2\sigma_{k+1}(z) - (k+1)^2\sigma_k(z) \leq \frac{2}{s}\left(\|z^k - z\|^2 - \|z^{k+1} - z\|^2\right).$$

Summing this inequality from $k = 1, \ldots, K$, we get for all $z \in \mathcal{H}$

$$(K+2)^2\sigma_{K+1}(z) \leq \frac{2}{s}\|x^1 - z\|^2 + 4\sigma_1(z).$$

Similar computations to Lemma 6.3 yield

$$\sigma_1(z) \leq \frac{1}{2s}\|x^2 - z\|^2 - \frac{1}{2s}\|x^2 - x^1\|^2.$$

Then, for all $k \geq 1$, we obtain

$$\sigma_k(z) \leq \frac{2\left(\|x^1 - z\|^2 + \|x^2 - z\|^2\right)}{s(k+1)^2}.$$

□

The theorem above is not straightforward to interpret since we only get convergence of order $\mathcal{O}(k^{-2})$ for $\min_{i=1,\dots,m} f_i(x^k) - f_i(z)$. This on its own does not state that the vector $f(x^k) = \big(f_1(x^k), \dots, f_m(x^k)\big)$ converges to an element of the Pareto front. However we can refine the statement of Theorem 6.6 in the following way to get a stronger convergence statement under a weak additional assumption.

Theorem 6.7 *Assume in addition to the assumption in Theorem 6.6 that for all $x \in \mathcal{L}(f(x_0))$ there exists an $x^* \in \mathcal{L}^*{:=}P_w \cap \mathcal{L}(f(x_0))$ with $f(x^*) \leq f(x)$ and*

$$\sup_{f^* \in f(\mathcal{L}^*)} \inf_{x \in f^{-1}(\{f^*\})} \|x - x^0\| < +\infty. \tag{35}$$

Then, there exists $R \geq 0$ such that

$$\sup_{z \in \mathcal{H}} \sigma_k(z) \leq \frac{4R}{(k+1)^2}, \quad \textit{for all } k \geq 0.$$

Proof Theorem 6.6 gives for all $z \in \mathcal{H}$

$$\sigma_k(z) \leq \frac{2\left(\|x^1 - z\|^2 + \|x^2 - z\|^2\right)}{s(k+1)^2}.$$

Taking a supremum over this inequality, we get

$$\sup_{f^* \in f(\mathcal{L}^*)} \inf_{z \in f^{-1}(f^*)} \sigma_k(z) \leq \sup_{f^* \in f(\mathcal{L}^*)} \inf_{z \in f^{-1}(f^*)} \frac{2\left(\|x^1 - z\|^2 + \|x^2 - z\|^2\right)}{s(k+1)^2}.$$

Since $x^1, x^2 \in \mathcal{L}(f(x^0))$ assumption (35) yields

$$\sup_{f^* \in f(\mathcal{L}^*)} \inf_{z \in f^{-1}(f^*)} \frac{2\left(\|x^1 - z\|^2 + \|x^2 - z\|^2\right)}{s(k+1)^2} \leq \frac{4R}{s(k+1)^2},$$

with

$$R = \max_{j=1,2} \left\{ \sup_{f^* \in f(\mathcal{L}^*)} \inf_{z \in f^{-1}(f^*)} \|x^j - z\|^2 \right\}.$$

It remains to show that

$$\sup_{z \in \mathcal{H}} \sigma_k(z) = \sup_{f^* \in f(\mathcal{L}^*)} \inf_{z \in f^{-1}(f^*)} \sigma_k(z).$$

Writing out the definition of $\sigma_k(z)$, we get

$$\begin{aligned}\sup_{f^*\in f(\mathcal{L}^*)}\inf_{z\in f^{-1}(f^*)}\sigma_k(z) &= \sup_{f^*\in f(\mathcal{L}^*)}\inf_{z\in f^{-1}(f^*)}\min_{i=1,\dots,m}\left(f_i(x^k)-f_i(z)\right)\\ &= \sup_{f^*\in f(\mathcal{L}^*)}\min_{i=1,\dots,m}\left(f_i(x^k)-f_i^*\right) = \sup_{z\in\mathcal{L}^*}\min_{i=1,\dots,m}\left(f_i(x^k)-f_i(z)\right)\\ &= \sup_{z\in\mathcal{H}}\min_{i=1,\dots,m}\left(f_i(x^k)-f_i(z)\right).\end{aligned}$$

□

The function $u_0(x) = \sup_{z\in\mathcal{H}}\min_{i=1,\dots,m} f_i(x) - f_i(z)$ attains the value zero if and only if x is weakly Pareto optimal. Theorem 6.7 shows that $u_0(x^k) = \mathcal{O}(k^{-2})$.

6.5 Relation to Tanabe's Accelerated Multiobjective Gradient Method

In the recent preprint [34], Tanabe, Fukuda and Yamashita define an accelerated proximal gradient method for MOPs with objective functions that have a separable structure of the form $f_i = g_i + h_i$, where $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex, continuously differentiable with L-Lipschitz continuous gradient and $h_i : \mathbb{R}^n \to \mathbb{R}$ is convex, lower semicontinuous and proper for all $i = 1, \dots, m$. Since we only treat the case of smooth objective functions f_i, we set from here on $h_i \equiv 0$. Tanabe et al. discovered their method using techniques different from the ones used throughout this paper, using the concept of merit functions. We will not recite their method here but refer the reader to [34]. To understand the similarity between their method and Algorithm 1, we investigate the quadratic optimization problems that have to be solved in each iteration of the methods, respectively. In the method from [34], the step direction is computed by solving a quadratic optimization problem with the following objective function $\Psi : \mathbb{R}^m \to \mathbb{R}$,

$$\Psi(\theta) := \frac{s}{2}\left\|\sum_{i=1}^m \theta_i \nabla f_i(y^k)\right\|^2 + \sum_{i=1}^m \theta_i\left(f_i(x^k) - f_i(y^k)\right).$$

Using the first-order approximation $f_i(y^k) - f_i(x^k) \approx \langle \nabla f_i(y^k), y^k - x^k\rangle$, we get

$$\Psi(\theta) \approx \frac{s}{2}\left\|\sum_{i=1}^m \theta_i \nabla f_i(y^k)\right\|^2 + \left\langle \sum_{i=1}^m \theta_i \nabla f_i(y^k), x^k - y^k\right\rangle.$$

Minimizing $\Psi(\theta)$ is equivalent to minimizing the function $\Phi : \mathbb{R}^m \to \mathbb{R}$,

$$\begin{aligned}\Phi(\theta) &:= \frac{s^2}{2}\left\|\sum_{i=1}^m \theta_i \nabla f_i(y^k)\right\|^2 + \left\langle s\sum_{i=1}^m \theta_i \nabla f_i(y^k), x^k - y^k\right\rangle + \frac{1}{2}\|x^k - y^k\|^2\\ &= \frac{1}{2}\left\|s\sum_{i=1}^m \theta_i \nabla f_i(y^k) + (x^k - y^k)\right\|^2.\end{aligned}$$

Using $x^k - y^k = -\frac{k-1}{k+2}(x^k - x^{k-1})$ we note that $\Phi(\theta)$ is in fact the objective function of the quadratic optimization problem (26). After this observation, it is not surprising that the method from [34] shows convergence behavior similar to Algorithm 1.

7 Improving the Numerical Efficiency

First-order methods for multiobjective optimization that are based on the steepest descent method by Fliege and Svaiter [19] require the solution of a quadratic subproblem in each iteration. Computing the solutions of these problems is computational demanding. In the following subsection, we present a possible approach to overcome this problem.

7.1 A Multiobjective Gradient Method Without Quadratic Subproblems

In this subsection, we define a method based on Algorithm 1 which does not require the solution of a quadratic subproblem in each iteration. In Sect. 6.2, we derived Algorithm 1 from the scheme

$$\frac{3}{k}(x^{k+1} - x^k) + \operatorname*{proj}_{sC(y^k)+(x^{k+1}-2x^k+x^{k-1})}(0) = 0,$$

which can be interpreted as a discretization of the differential equation

$$\frac{3}{t}\dot{x}(t) + \operatorname*{proj}_{C(x(t))+\ddot{x}(t)}(0) = 0.$$

If, instead, we use the discretization

$$\frac{3}{k}(x^k - x^{k-1}) + \operatorname*{proj}_{sC(y^k)+(x^{k+1}-2x^k+x^{k-1})}(0) = 0,$$

we obtain a different method. Lemma A.1 gives a formula to compute x^{k+1}

$$\begin{aligned} x^{k+1} &= -\frac{3}{k}(x^k - x^{k-1}) - s\sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k) + 2x^k - x^{k-1}, \\ &= x^k + \frac{k-3}{k}(x^k - x^{k-1}) - s\sum_{i=1}^{m}\theta_i^k \nabla f_i(x^k), \end{aligned} \tag{36}$$

where $\theta^k \in \mathbb{R}^m$ is a solution to the problem

$$\min -\sum_{i=1}^{m}\theta_i \langle \nabla f_i(x^k), x^k - x^{k-1}\rangle \ \text{ s.t. } \ \theta \geq 0 \text{ and } \sum_{i=1}^{m}\theta_i = 1.$$

This can be solved efficiently by computing m inner products. After changing $\frac{k-3}{k}$ into $\frac{k-1}{k+2}$ in (36), we define Algorithm 2. There is no proof of convergence for the method defined by Algorithm 2 but we discuss its numerical behavior in Sect. 8. Also, convergence can be guaranteed by switching from the significantly faster Algorithm 2 to Algorithm 1 as soon as some heuristic criterion is met.

Algorithm 2 Accelerated multiobjective gradient method without quadratic subproblems

Require: Choose $x^0 = x^1 \in \mathcal{H}$, $s > 0$ and set $k = 1$.
1: Set $y^k = x^k + \frac{k-1}{k+2}(x^k - x^{k-1})$.
2: Compute $j = \arg\max_{i=1,\dots,m} \langle \nabla f_i(y^k), x^k - x^{k-1} \rangle$.
3: Set $x^{k+1} = y^k - s\nabla f_j(y^k)$
4: **if** stopping condition is true **then**
5: Stop.
6: **else**
7: Update $k \leftarrow k+1$ and go to step 1.
8: **end if**

7.2 Backtracking for Unknown Lipschitz Constants

In all presented algorithms, we can include backtracking if the Lipschitz constants of the gradients ∇f_i of the objective functions are unknown. We can do this as stated in [13, 34]. To include backtracking, we choose an initial step size $s_0 > 0$ and a parameter $\sigma \in (0, 1)$. In all discussed algorithms there is a step $x^{k+1} = w^k - sd^k$, with $d^k \in \mathcal{H}$ and $w^k = x^k$ or $w^k = y^k$. One can replace this step with $x^{k+1} = w^k - s_k d^k$, with a step size s_k that is determined using backtracking. We choose in every step $s_k = \sigma^{l_k} s_{k-1}$ where $l_k \geq 0$ is the smallest nonnegative integer satisfying for all $i = 1, \dots, m$

$$f_i(w^k - \sigma^{l_k} s_{k-1} d^k) \leq f_i(w^k) - \sigma^{l_k} s_{k-1} \langle \nabla f_i(w^k), d^k \rangle + \frac{\sigma^{l_k} s_{k-1}}{2} \|d^k\|^2.$$

The sequence $(s_k)_{k\geq 0}$ is monotonically decreasing by definition. Under the condition that the objective functions posses L-Lipschitz continuous gradients, it is guaranteed that the sequence $(s_k)_{k\geq 0}$ is constant from same k on. This is true since s_k can only decrease as long as $s_k > \frac{1}{L}$. Therefore, s_k can only decrease finitely many times until it reaches a point where $s_k \leq \frac{1}{L}$. Using this observation, we can include backtracking in Algorithm 1 and still use the proofs of Theorem 6.6 and Theorem 6.7 to show that the same convergence results can be achieved.

8 Numerical Examples

In this section, we present the typical behavior of our algorithms on two test problems. We compare Algorithms 1 and 2 with the steepest descent method by Fliege and Svaiter

with constant step sizes [19]. Throughout this section we denote the steepest descent method by SD, Algorithm 1 by AccG (accelerated gradient method) and Algorithm 2 by AccG w\o Q (accelerated gradient method without quadratic subproblems). We implemented all codes using MATLAB R2021b and executed the algorithms on a machine with a 2.80 GHz Intel Core i7 processor and 48 GB memory. We solved the quadratic subproblems for SD and AccG using the built-in MATLAB function quadprog.

8.1 Example 1: A Convex MOP with Three Objective Functions

In our first example, we choose a problem with input dimension $n = 20$ and three objective functions ($m = 3$). We define the objective functions using the following parameters. For $p = 50$ and $i = 1, 2, 3$ we generate matrices $A^i = \left(a_1^i, \ldots, a_p^i\right)^\top \in \mathbb{R}^{p\times n}$ with $a_j^i \in \mathbb{R}^n$ for $j = 1, \ldots, p$ and vectors $b^i \in \mathbb{R}^p$. Then, for $i = 1, 2, 3$, we define the objective functions

$$f_i : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto \ln\left(\sum_{j=1}^{p} \exp\left((a_j^i)^\top x - b_j^i\right)\right).$$

For the first experiment we randomly generate matrices $A^i \in \mathbb{R}^{p\times n}$ and vectors $b_i \in \mathbb{R}^p$ with entries uniformly sampled in $[-1, 1]$ for $i = 1, 2, 3$. The starting vector x_0 is uniformly randomly drawn from $[-15, 15]^n$. We use the step size $s = 5\text{e}{-2}$ and execute maximally $k_{\max} = 1000$ iterations. Figure 1 contains plots of the sequences $(x^k)_{k\geq 0}$ for the different algorithms. In Fig. 1a, it can be seen that the sequences generated with AccG and AccG w\o Q advance much faster in the beginning, while the velocity of the sequence generated with SD remains constant. The sequences generated by AccG and AccG w\o Q give very similar trajectories in the beginning. This result is intuitive given that the schemes in the algorithms are derived from different discretizations of the same differential equation. However, this result is still surprising keeping in mind that in Algorithm 2 we do not solve a quadratic subproblem in each iteration. Only in Fig. 1b, we see that the sequences differ more substantially in the long run. It is also noteworthy that the sequence generated by AccG is smoother compared to the trajectory generated by AccG w\o Q. This is due to the fact that in AccG w\o Q we choose one of the gradients of the objective functions for the gradient component of the step direction while in AccG we choose an element of the convex hull of the gradients. AccG and AccG w\o Q are superior to SD in terms of convergence of the function values for all objective functions, as shown in Fig. 2. AccG and AccG w\o Q experience fast convergence within the first 200 iterations. Comparing the different objective functions in Fig. 2a–c, we see that AccG and AccG w\o Q yield outputs with similar function values for all objective functions.

In a second experiment, we execute all algorithms for 50 starting values uniformly sampled in $[-5, 5]^n$ with step size $s = 5\text{e}{-2}$. We use the stopping criterion $\|f(x^k) - f(x^{k-1})\|_\infty < 1\text{e}{-4}$ to stop the algorithms if the function values do not change

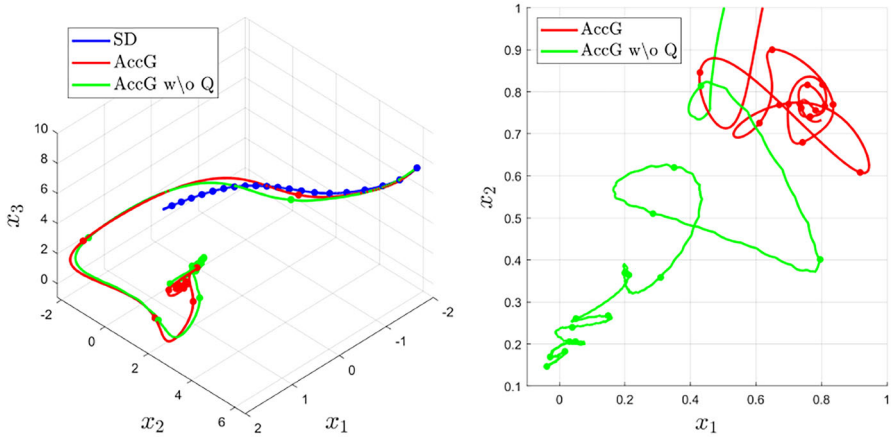
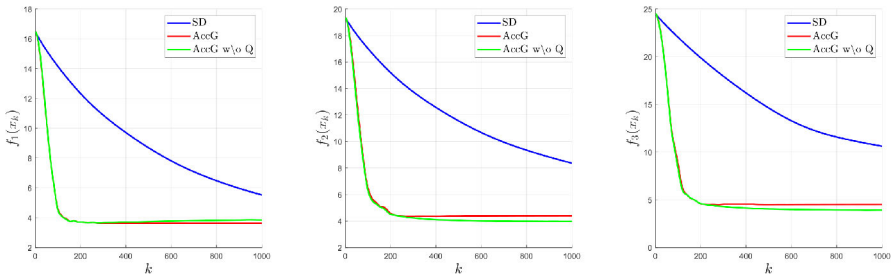


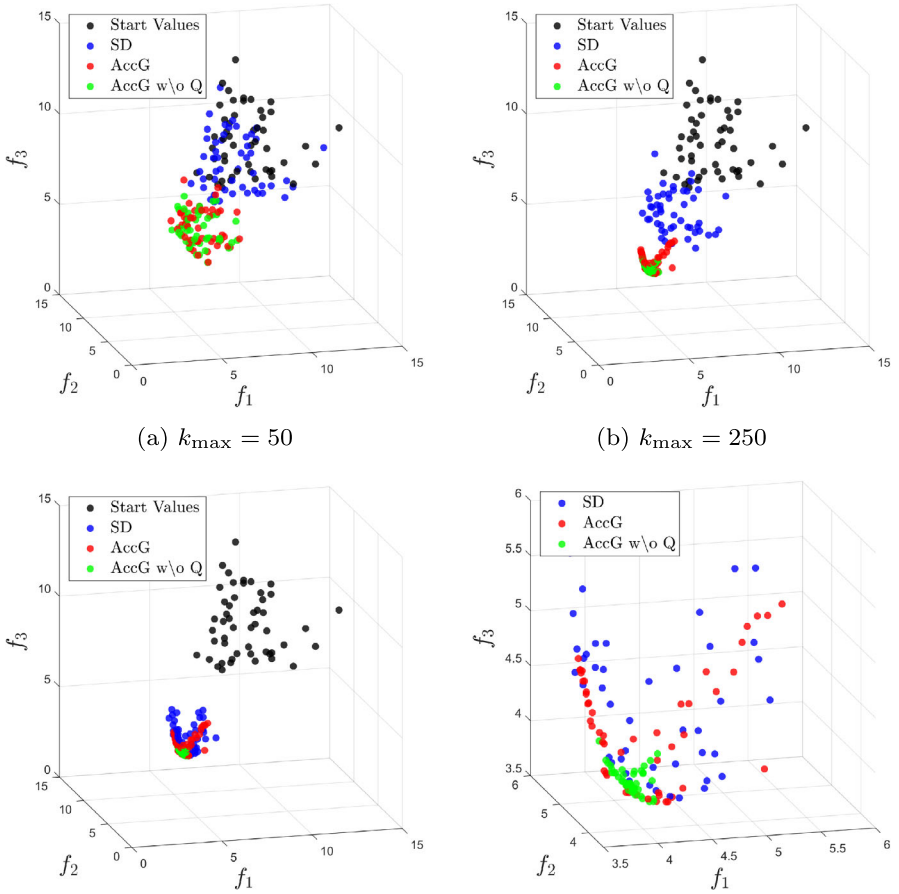
Fig. 1 Coordinates (x_1, x_2, x_3) of the sequences $(x^k)_{k\geq 0}$ for SD, AccG and AccG w\o Q. Line plot for 1000 iterations with a filled circle every 50 iterations to compare the velocities



(a) Function values $f_1(x^k)$ (b) Function values $f_2(x^k)$ (c) Function values $f_3(x^k)$

Fig. 2 Function values $(f_i(x^k))_{k\geq 0}$ of the iterates for the objective functions $i = 1, 2, 3$ for the different algorithms

significantly. In Fig. 3a, we perform up to $k_{\max} = 50$, in Fig. 3b up to $k_{\max} = 250$ and in Fig. 3c, d up to $k_{\max} = 1000$ iterations. Similar to the results observed in Figs. 1 and 2, AccG and AccG w\o Q advance much faster in the beginning compared to SD. Comparing Fig. 2b, c, we see that after 250 iterations the function values for the accelerated methods are converging or the stopping conditions were met. The different behavior of the accelerated methods can be observed in Fig. 3d. While the solutions of AccG are farther spread, it looks like the solutions of AccG w\o Q are drawn toward the center of the Pareto front. Altogether, the accelerated methods perform better for this problem in terms of convergence speed of the function values. In Table 1 the total number of iterations and computation times for the experiments are listed. The accelerated methods require fewer iterations. Compared to SD, AccG requires only approximately 25 % and AccG w\o Q only approximately 50 % iterations. For the computation times the results are different. SD and AccG behave similar, with AccG requiring approximately 25 % of the computation time that is required for SD.



(a) $k_{\max} = 50$ (b) $k_{\max} = 250$

(c) $k_{\max} = 1000$ (d) $k_{\max} = 1000$

Fig. 3 Function values of the objective functions in the image space for the different algorithms and a different maximum number of iterations $k_{\max} = 50, 250, 1000$

Table 1 Total iterations and computation times for algorithm executions using parameters $s = 5\text{e}{-2}$, $k_{\max} = 1000$ and stopping condition $\|f(x^k) - f(x^{k-1})\|_\infty < 1\text{e}{-4}$ for 50 start values uniformly sampled in $[-15, 15]^n$

	SD	AccG	AccG w\o Q
Total iterations	49924	12230	25906
Total time	436.54 s	100.94 s	1.61 s

However, AccG w\o Q needs less the 2 % of the time which is consumed by AccG. This improvement stems from the quadratic optimization problems that are not required in AccG w\o Q.

8.2 Example 2: A Nonconvex MOP with Two Objective Functions

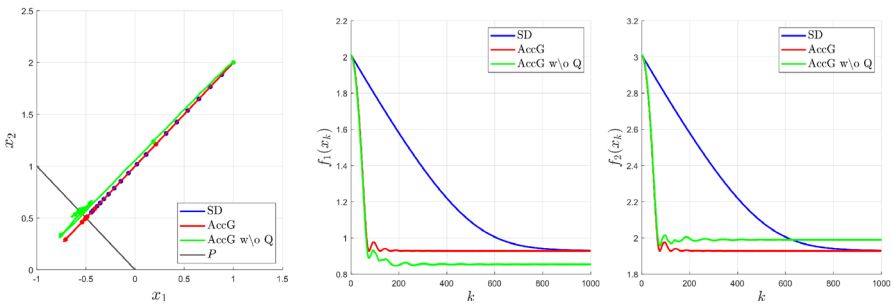
For our second test problem, we choose an example from [36] with input dimension $n = 2$ and the two objective functions

$$f_1(x) = \frac{1}{2}\left(\sqrt{1+(x_1+x_2)^2} + \sqrt{1+(x_1-x_2)^2} + x_1 - x_2\right) + \lambda \exp\left(-(x_1-x_2)^2\right),$$
$$f_2(x) = \frac{1}{2}\left(\sqrt{1+(x_1+x_2)^2} + \sqrt{1+(x_1-x_2)^2} - x_1 + x_2\right) + \lambda \exp\left(-(x_1-x_2)^2\right),$$

with $\lambda = 0.6$. For the multiobjective optimization problem (MOP) with these objective functions it can easily be verified that the Pareto set is

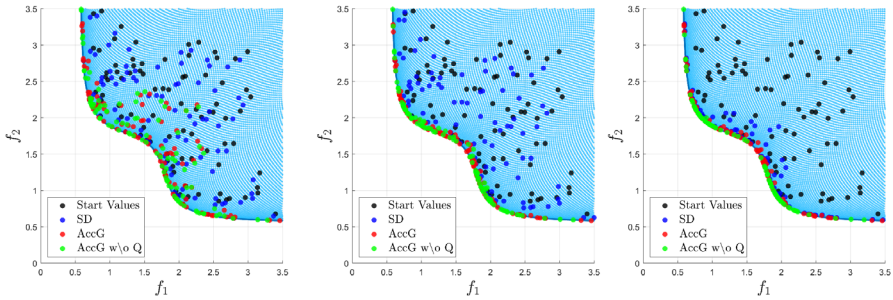
$$P = \left\{x \in \mathbb{R}^2 \ : \ x_1 + x_2 = 0\right\}.$$

In the first experiment, we execute Algorithms SD, AccG and AccG w\o Q with the starting vector $x^0 = (1, 2)^\top$ and perform $k_{\max} = 1000$ iterations. The step size is set to $s = 5\text{e}{-3}$. The sequences and function values of the objective functions are shown in Fig. 4. Similarly to the first experiment in Fig. 4a, the sequences $(x^k)_{k\geq 0}$ of the accelerated methods advance faster in the beginning. While algorithms SD and AccG converge to the same element in the Pareto set the algorithm AccG w\o Q produces a trajectory that deviates from the trajectories of SD and AccG and moves to a different part of the Pareto set. The values of the objective functions in Fig. 4b, c indicate a similar behavior. For the accelerated methods we have faster decrease in the beginning and we note that the function values for SD and AccG converge to similar values. In Fig. 5 we use 100 random starting points uniformly sampled in $[-2, 2]^2$. For the experiments we use different maximal numbers of iterations $k_{\max}$. In addition we stop the algorithm if $\|f(x^k) - f(x^{k-1})\|_\infty < 1\text{e}{-4}$. Comparing Fig. 5a–c, we note that the objective function values of the accelerated methods decrease much faster in the beginning. For $k_{\max} = 100$ algorithms AccG and AccG w\o Q yield solutions that are distributed along the Pareto front. In Table 2 we list the total number of iterations



(a) Sequences $(x^k)_{k\geq 0}$ (b) Function values $f_1(x^k)$ (c) Function values $f_2(x^k)$

Fig. 4 Sequences $(x^k)_{k\geq 0}$ and function values $(f_i(x^k))_{k\geq 0}$ of iterates for $i = 1, 2$. For the sequences we use a line plot for 1000 iterations with a filled circle every 50 iterations to compare velocities

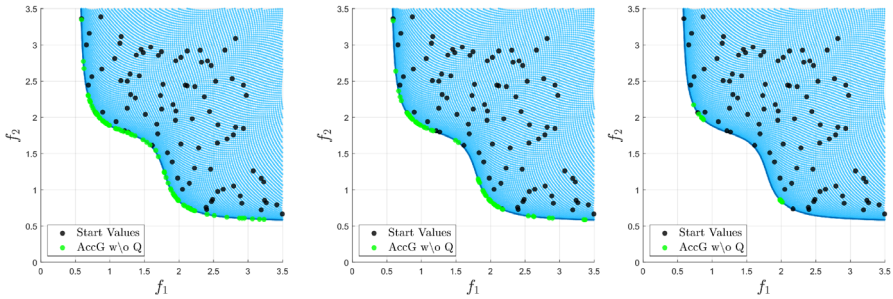


(a) $k_{\max} = 50$ (b) $k_{\max} = 100$ (c) $k_{\max} = 500$

Fig. 5 Values of the objective functions in the image space for different maximum numbers of iterations $k_{\max} = 50, 100, 500$ for the different algorithms SD, AccG and AccG w\o Q

Table 2 Total iterations and computation times for algorithm executions using parameters $h = 5\text{e}{-3}$, $k_{\max} = 1000$ and stopping condition $\|f(x^k) - f(x^{k-1})\|_\infty < 1\text{e}{-4}$ for 100 start values uniformly sampled in $[-2, 2]^n$

	SD	AccG	AccG w\o Q
Total iterations	45543	6632	23034
Total time	431.75 s	62.90 s	0.31 s



(a) $s = 5\text{e}{-3}$ (b) $s = 1\text{e}{-2}$ (c) $s = 5\text{e}{-2}$

Fig. 6 Values of the objective functions in the image space for different step sizes $s = 5\text{e}{-3}, 1\text{e}{-2}, 5\text{e}{-2}$ for the algorithm AccG w\o Q

and total computation times for executions with up to $k_{\max} = 1000$ iterations with stopping condition $\|f(x^k) - f(x^{k-1})\|_\infty < 1\text{e}{-4}$. Compared to SD, AccG needs only approximately 15 % and AccG w\o Q only approximately 51 % iterations.

In another experiment we compare how the choice of the step size s affects the solutions of AccG w\o Q. We use the step sizes $s = 5\text{e}{-3}, 1\text{e}{-2}, 5\text{e}{-2}$. For all executions we perform $k_{\max} = 1000$ iterations, with the stopping criterion $\|f(x^k) - f(x^{k-1})\|_\infty < 1\text{e}{-4}$. Comparing Fig. 6a–c we see that for the smallest step size $s = 5\text{e}{-3}$ solutions are distributed on the whole Pareto front. For the biggest step size $s = 5\text{e}{-2}$ Algorithm AccG w\o Q yields solutions that cluster at two points of the Pareto front, which is not desirable in general. The two points where the solutions

cluster correspond to the knee points in the Pareto front. This is not surprising since these points correspond to solutions where the individual gradients are similar in magnitude, which is why the solution is zig-zagging back and forth between the objectives in these locations. However, this disadvantage is compensated by the fact that we do not need to solve a quadratic subproblem in every step. We need potentially more iterations when choosing smaller step sizes but every iteration is computationally cheaper in comparison to SD and AccG. Moreover, a small adaptation to step 2 of Algorithm 2, where we include a weighting parameter in the max problem, might allow us to diversify solutions, similar to the weighted sum method.

9 Conclusion and Open Questions

We present the novel inertial gradient-like dynamical system (IMOG') for Pareto optimization. We show that trajectories of this system converge weakly to Pareto critical points of (MOP). Based on this, we define a novel inertial gradient method for multiobjective optimization and show weak convergence to Pareto critical points. We derive an accelerated gradient method from the informally introduced inertial gradient-like system (MAVD) which incorporates asymptotically vanishing damping. Using the concept of merit functions, we show that our method possesses an improved convergence rate. Using a different discretization of the system (MAVD), we define an accelerated gradient method which does not require the solution to a quadratic optimization problem in every iteration. A comparison on selected test problems shows that the accelerated methods are in fact superior to the plain multiobjective steepest descent method.

There are a lot of open questions arising from the presented work. The gradient system (IMOG') can be analyzed for different problem classes. In addition, we can adapt our gradient systems and algorithms to treat problems with a separable smooth and nonsmooth structure using proximal methods. Another research direction is the adaption of the presented gradient systems and algorithms by the means of Hessian driven damping (see, e.g., [3]) which attenuates oscillations of the trajectories naturally arising in inertial systems. This way it improves the behavior of inertial gradient methods. It would be interesting to analyze Hessian driven damping in the context of multiobjective optimization. It would also be interesting to investigate the behavior of our algorithms for high-dimensional and nonconvex problems. In addition, one could apply the presented algorithms in the area of machine learning, e.g., for multitask learning problems [28].

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If

material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

A Two Lemmas on Convex Projections

Lemma A.1 *Let $\mathcal{H}$ be a real Hilbert space, $C \subset \mathcal{H}$ a convex and compact set and $\eta \in \mathcal{H}$ a fixed vector. Then, $\xi \in \mathcal{H}$ is a solution to the problem*

$$\text{Find } \xi \in \mathcal{H} \text{ such that : } \eta = \operatorname*{proj}_{C+\xi}(0), \tag{37}$$

if and only if it has the form $\xi = \eta - \mu$, where μ is a solution to the constrained optimization problem $\min_{\mu \in C} \langle \mu, \eta \rangle$.

Proof First, we show that an element of the form $\xi = \eta - \mu$, with μ a solution to $\min_{\mu \in C} \langle \mu, \eta \rangle$ is a solution to problem (37). The set of minimizers of the problem $\min_{\mu \in C} \langle \mu, \eta \rangle$ is nonempty, since C is compact. Fix an arbitrary solution $\mu \in \arg\min_{\mu \in C} \langle \mu, \eta \rangle$. Since C is convex, the first-order optimality condition for this problem gives that for all $x \in C$ it holds that $\langle x - \mu, \eta \rangle \geq 0$ and hence

$$\langle x + \xi - (\mu + \xi), \eta \rangle \geq 0.$$

Since we have chosen $\xi = \eta - \mu$ the equation above reads as

$$\langle x + \xi - \eta, \eta \rangle \geq 0,$$

which is equivalent to $\eta = \operatorname{proj}_{C+\xi}(0)$. The other direction works analogously. If the vector ξ is a solution to problem (37) this guarantees that $\mu = \xi - \eta$ satisfies the first-order optimality condition for problem $\min_{\mu \in C} \langle \mu, \eta \rangle$. Since problem $\min_{\mu \in C} \langle \mu, \eta \rangle$ is convex and defined over a convex set, this is equivalent to μ being an optimal solution to $\min_{\mu \in C} \langle \mu, \eta \rangle$. □

Lemma A.2 *Let $\mathcal{H}$ be a real Hilbert space, $C \subset \mathcal{H}$ a convex and closed set and $a > 0$, $\nu \in \mathcal{H}$ fixed. Then, the problem*

$$\text{Find } \xi \in \mathcal{H} \text{ such that : } -a(\xi + \nu) = \operatorname*{proj}_{C+\xi}(0), \tag{38}$$

has the unique solution $\xi = -\left(\frac{1}{1+a}\operatorname{proj}_C(\nu) + \frac{a}{1+a}\nu\right)$.

Proof First, we show that $\xi = -\left(\frac{1}{1+a}\operatorname{proj}_C(\nu) + \frac{a}{1+a}\nu\right)$ is a solution to (38). It is easy to check that $-a(\xi + \nu) \in C + \xi$. Define the projection $p := \operatorname{proj}_C(\nu)$. For all $x \in C$ it holds that $\langle x - p, p - \nu \rangle \geq 0$ and hence for all $x \in C$ we get

$$\langle x + \xi + a(\xi + \nu), a(\xi + \nu) \rangle \leq 0,$$

which is equivalent to

$$-a(\xi + \nu) = \operatorname*{proj}_{C+\xi}(0).$$

The uniqueness follows the same way. Assume we have a solution $\tilde{\xi}$ to (38). By the same computations as above it holds that for all $x \in C$

$$\langle x + (1+a)\tilde{\xi} + a\nu, \tilde{\xi} + \nu \rangle \leq 0.$$

This is equivalent to

$$-((1+a)\tilde{\xi} + a\nu) = \operatorname*{proj}_{C}(\nu),$$

from which follows that $\xi = \tilde{\xi}$ is the unique solution. □

References

1. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. SIAM J. Control. Optim. **38**(4), 1102–1119 (2000). https://doi.org/10.1137/S0363012998335802
2. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set-Valued Anal. **9**, 3–11 (2001). https://doi.org/10.1023/A:1011253113155
3. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping: application to optimization and mechanics. Journal de Mathématiques Pures et Appliquées **81**(8), 747–779 (2002). https://doi.org/10.1016/S0021-7824(01)01253-3
4. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**(1), 123–175 (2018). https://doi.org/10.1007/s10107-016-0992-8
5. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. ESAIM Control Optim. Calculus Var. **25**, 2 (2019). https://doi.org/10.1051/cocv/2017083
6. Attouch, H., Fadili, J.: From the Ravine method to the Nesterov method and vice versa: a dynamical system perspective. SIAM J. Optim. **32**(3), 2074–2101 (2022). https://doi.org/10.1137/22M1474357
7. Attouch, H., Garrigos, G.: Multiobjective optimization: an inertial dynamical approach to Pareto optima. (2015). arXiv preprint arXiv:1506.02823
8. Attouch, H., Garrigos, G., Goudou, X.: A dynamic gradient approach to Pareto optimization with nonsmooth convex objective functions. J. Math. Anal. Appl. **422**(1), 741–771 (2015). https://doi.org/10.1016/j.jmaa.2014.09.001
9. Attouch, H., Goudou, X.: A continuous gradient-like dynamical approach to Pareto-optimization in Hilbert spaces. Set-Valued Var. Anal. **22**(1), 189–219 (2014). https://doi.org/10.1007/s11228-013-0245-4
10. Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. Commun. Contemp. Math. **2**(01), 1–34 (2000). https://doi.org/10.1142/S0219199700000025

11. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. (2015). https://doi.org/10.1137/15M1046095
12. Aubin, J.P., Cellina, A.: Differential Inclusions: Set-Valued Maps and Viability Theory, vol. 264. Springer, Berlin (2012). https://doi.org/10.1007/978-3-642-69512-4
13. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. **2**(1), 183–202 (2009). https://doi.org/10.1137/080716542
14. Brezis, H.: Operateurs Maximaux Monotones Et Semi-Groupes De Contractions Dans Les Espaces De Hilbert. North Hollad, Amsterdam (1973). https://doi.org/10.1016/S0304-0208(08)72386-7
15. Chen, G.Y., Goh, C.J., Yang, X.Q.: On gap functions for vector variational inequalities. In: F. Giannessi (ed.) Vector Variational Inequalities and Vector Equilibria, pp. 55–72. Springer (2000). https://doi.org/10.1007/978-1-4613-0299-5_4
16. Cornet, B.: Existence of slow solutions for a class of differential inclusions. J. Math. Anal. Appl. **96**(1), 130–147 (1983). https://doi.org/10.1016/0022-247X(83)90032-X
17. Deimling, K.: Multivalued Differential Equations, vol. 1. Walter de Gruyter, Berlin (1992). https://doi.org/10.1515/9783110874228
18. El Moudden, M., El Mouatasim, A.: Accelerated diagonal steepest descent method for unconstrained multiobjective optimization. J. Optim. Theory Appl. **188**(1), 220–242 (2021). https://doi.org/10.1007/s10957-020-01785-9
19. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000). https://doi.org/10.1007/s001860000043
20. Henry, C.: An existence theorem for a class of differential equations with multivalued right-hand side. J. Math. Anal. Appl. **41**(1), 179–186 (1973). https://doi.org/10.1016/0022-247X(73)90192-3
21. Liu, C.G., Ng, K.F., Yang, W.H.: Merit functions in vector optimization. Math. Program. **119**(2), 215–237 (2009). https://doi.org/10.1007/s10107-008-0208-y
22. Luenberger, D.G.: Optimization by Vector Space Methods. John Wiley & Sons, New York (1997)
23. Miettinen, K.: Nonlinear Multiobjective Optimization. Springer, New York (1998). https://doi.org/10.1007/978-1-4615-5563-6
24. Miglierina, E.: Slow solutions of a differential inclusion and vector optimization. Set-Valued Anal. **12**(3), 345–356 (2004). https://doi.org/10.1023/B:SVAN.0000031332.10564.f0
25. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk. SSSR **269**(3), 543–547 (1983)
26. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. **73**(4), 591–597 (1967). https://doi.org/10.1090/S0002-9904-1967-11761-0
27. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. **4**(5), 1–17 (1964). https://doi.org/10.1016/0041-5553(64)90137-5
28. Sener, O., Koltun, V.: Multi-task learning as multi-objective optimization. In: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (eds.) Advances in Neural Information Processing Systems, vol. 31 (2018). https://papers.nips.cc/paper_files/paper/2018/file/432aca3a1e345e339f35a30c8f65edce-Paper.pdf
29. Smale, S.: Global analysis and economics I: Pareto optimum and a generalization of Morse theory. In: M. Peixoto (ed.) Dynamical Systems, pp. 531–544. Elsevier (1973). https://doi.org/10.1016/B978-0-12-550350-1.50044-8
30. Sonntag, K., Peitz, S.: Fast convergence of inertial multiobjective gradient-like systems with asymptotic vanishing damping. (2023) https://doi.org/10.48550/arXiv.2307.00975. arXiv preprint arXiv:2307.00975
31. Su, W., Boyd, S., Candes, E.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. Advances in Neural Information Processing Systems **27** (2014). https://proceedings.neurips.cc/paper_files/paper/2014/file/f09696910bdd874a99cd74c8f05b5c44-Paper.pdf
32. Svaiter, B.F.: The multiobjective steepest descent direction is not Lipschitz continuous, but is Hölder continuous. Oper. Res. Lett. **46**(4), 430–433 (2018). https://doi.org/10.1016/j.orl.2018.05.008
33. Tanabe, H., Fukuda, E.H., Yamashita, N.: Convergence rates analysis of a multiobjective proximal gradient method. Optim. Lett. (2022). https://doi.org/10.1007/s11590-022-01877-7
34. Tanabe, H., Fukuda, E.H., Yamashita, N.: An accelerated proximal gradient method for multiobjective optimization. Comput. Optim. Appl. (2023). https://doi.org/10.1007/s10589-023-00497-w
35. Tanabe, H., Fukuda, E.H., Yamashita, N.: New merit functions for multiobjective optimization and their properties. Optimization (2023). https://doi.org/10.1080/02331934.2023.2232794

36. Witting, K.: Numerical Algorithms for the Treatment of Parametric Multiobjective Optimization Problems and Applications. Ph.D. thesis, Paderborn, Universität Paderborn, Dissertation (2012)
37. Yang, X.Q., Yao, J.C.: Gap functions and existence of solutions to set-valued vector variational inequalities. J. Optim. Theory Appl. **115**(2), 407–417 (2002). https://doi.org/10.1023/A:1020844423345

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.